

CENUR Litoral Norte / Facultad de Química

Universidad de la Republica

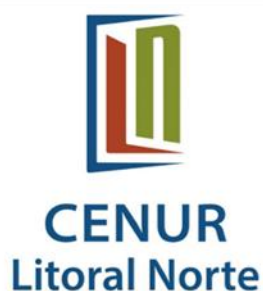
**Desarrollo de Métodos Quimiométricos Basados en
Resonancia Magnética Nuclear (RMN) y su Aplicación al
Estudio de Matrices de Interés Agropecuario**

**Tesis presentada por el Q.F. Andrés López-Radcenco aspirando al título de
Doctor en Química**

Tutores:

Dr. Guillermo Moyna, Ph.D

Dr. Pablo Hoijemberg, Ph.D



Índice	
Agradecimientos	7
Abreviaturas	8
Resumen	9
Capítulo I. Introducción general	11
1.1 Introducción a la Metabolómica basada en RMN.....	15
1.2 Fundamentos Metodológicos.....	20
1.2.1 Bases y modelo vectorial de la resonancia magnética nuclear.....	21
1.2.2 Apantallamiento.....	25
1.3 Esquema general de trabajo en metabolómica.....	27
1.3.1 Análisis de muestras y Procesado de datos.....	28
1.4 Análisis Multivariado de Datos.....	34
1.4.1 Análisis de Componentes Principales (PCA).....	38
1.4.2 Proyecciones de Estructuras Latentes (PLS) y Proyecciones de Estructuras Latentes Ortogonales (OPLS).....	42
1.5 Identificación de variables predictivas.....	47
1.5.1 Gráfico de valores de carga.....	47
1.5.2 Gráfico de Volcán (“Volcano plot”).....	48
1.5.3 Grafico S (“S-Plot”).....	49
1.6 Validación de modelos supervisados	49
1.6.1 Ajuste y Predictibilidad (R^2 y Q^2)	49
1.6.2 Prueba de Permutaciones	50
1.6.3 Curvas de características operativas del receptor (ROC).....	51

1.7 Identificación de metabolitos y búsqueda en bases de datos	53
1.7.1 Estadística de correlación espectroscópica total (STOCSY)	54
1.7.2 Bases de datos	56
1.7.3 Niveles en la identificación de metabolitos	47
Capítulo II. Seguimiento del estado sanitario e identificación de biomarcadores de salud de ganado lechero.....	58
2. 1. Introducción	58
2.2. Materiales y Métodos.....	61
2.2.1. Recolección de muestras y animales	61
2.2.2. Bioquímica Clínica	62
2.2.3. Espectroscopía de RMN	63
2.2.4. Procesado de datos de RMN	63
2.2.5. Determinación de metabolitos	64
2.2.6. Análisis Estadísticos	64
2.3. Resultados y discusión	65
2.3.1. Espectros de ¹ H RMN de muestras de suero	65
2.3.2. Metabolómica sérica por RMN	66
2.4. Conclusiones.....	77
Capítulo III. Identificación de fitomarcadores para la selección y mejoramiento de cultivares.....	79
3.1. Introducción	79
3.2 Materiales y métodos	82
3.2.1 Cría de áfidos	82
3.2.2 Evaluación de respuesta antixenótica	83

3.2.3 Material vegetal	84
3.2.4 Preparación de muestras para RMN	85
3.2.5 Espectroscopía de RMN	85
3.2.6 Procesamiento de datos	86
3.2.7 Análisis estadístico.....	86
3.2.8 Agregado de estándares.....	87
3.3 Resultados.....	87
3.3.1 Evaluación de la respuesta antixenótica.....	87
3.3.2 Espectros de ^1H RMN de extractos de cebada.....	88
3.3.3 Metabolómica basada en RMN.....	90
3.4 Discusión y Conclusiones.....	96
Capítulo IV. Uso de metodologías no tradicionales de procesamiento y análisis de datos de RMN en estudios metabolómicos.....	99
4.1 Introducción.....	99
4.2 Materiales y métodos.....	102
4.2.1 Recolección de muestras orina y sangre.....	102
4.2.2 Preparación de muestras.....	102
4.2.3 Adquisición de datos de RMN.....	102
4.2.4 Software.....	104
4.2.5 Análisis estadísticos.....	104
4.3 Resultados y Discusión.....	105
4.3.1 Desarrollo de JIBA	105
4.3.2 Alineación empleando JIBA.....	106
4.3.3 Estrategias para el uso de espectros HOMO 2D J en MVDA.....	108

4.3.4 Fusión de datos.....	110
4.4 Conclusiones y Perspectivas.....	113
Referencias.....	116
Anexo.....	139

Agradecimientos

Nunca caminé solo.

Quiero empezar por expresar mi mas profundo agradecimiento y dedicar este trabajo a mi madre. Ella a sido y es la persona que me ha apoyado e impulsado incondicionalmente siempre en cada etapa de mi vida.

También quiero agradecer a la familia, los amigos, y compañeros que hacen el lugar que habito y que me acompañan en el camino.

Fundamentalmente agradezco a mis tutores Guille y Pablo por compartir su saber conmigo, por su guía, su apoyo incondicional, y principalmente, por su paciencia. Guille además de ser mi mentor, eres un gran compañero en le camino desde que llegue Paysandú y al DQL.

Además, quiero dar gracias a todos los colegas del DQL, quienes son una segunda familia donde también encontré amigos y compañeros. Gracia a todos por siempre dar una mano cuando se necesita. También quiero agradecer a los compañeros y colegas en Facultad de Química. Es ahí donde me formé, aprendí a maravillarme con la química y donde comencé mi carrera académica. Agradezco especialmente a Laura y Bea del grupo Farmacología-Lea donde di mis primeros pasos dentro de la academia y Facultad. Fueron ellas quienes me apoyaron cuando decidí tomar la oportunidad de continuar en el litoral del país en el CENUR-LN.

Hay otra persona a quien quiero dar gracias, que hace muchos años me ayudo a elegir la química sobre otras ciencias y curiosidades que un adolescente

puede tener. Gracias, Rubén Ferreira, por ser un docente comprometido con su vocación.

A todos mis mentores quienes compartieron su saber y experiencia conmigo, a todos ellos de quienes pude aprender algo nuevo, ¡muchas gracias!

Por último, quiero agradecer a distintas instituciones: la UdelaR, la Facultad de Química, el CENUR-Litoral Norte, al CIBION(Argentina), a PEDECIBA y CSIC por permitirme realizar mi trabajo de posgrado. Mas importante aún, por impulsar y fomentar la ciencia, por crear los lugares que nuclea a aquellos que persiguen el conocimiento y buscan compartirlo. Sin estos espacios y la gente que los habita no me podría haber formado. La ciencia avanza desde la cooperatividad y los espacios nutritivos para que se desarrolle.

Resumen

El análisis metabolómico, o metabolómica, permite estudiar las variaciones en los metabolitos de un sistema biológico en función de factores como la dieta, el entorno, fármacos, o enfermedades. La disciplina es relativamente nueva en Uruguay, pero ya tiene un impacto significativo a nivel mundial en diversas áreas, yendo desde la industria farmacéutica a la agroalimentaria. Esta tesis aborda principalmente la aplicación de la metabolómica a diferentes matrices de interés para el sector agropecuario uruguayo, así como el desarrollo de nuevas estrategias para el análisis de datos. Luego de un primer capítulo introductorio, se presenta un detallado análisis del metaboloma sérico de vacas lecheras durante el parto, identificando alteraciones asociadas con el periodo de transición y producción de leche. El análisis por RMN revela alteraciones en el metabolismo energético, permitiendo además observar cómo se movilizan metabolitos para suplir demandas energéticas y nutricionales relacionadas a la producción de leche. El siguiente bloque del trabajo presenta el uso de metabolómica basada en RMN para estudiar la resistencia de la cebada al pulgón verde. Los estudios identificaron tetrametilaminas, como la colina y la betaína, las cuales están correlacionadas con la preferencia o rechazo del pulgón. En esta sección se realiza además un análisis detallado de todos los pasos involucrados en la confirmación inequívoca de metabolitos en un esquema de trabajo en metabolómica. Esta sección demuestra además un posible uso de esta herramienta como parte de esquemas de manejo integrado de plagas. Finalmente se describe el desarrollo de un algoritmo para el alineamiento de espectros

bidimensionales, paso necesario para emplear este tipo de datos en el análisis metabolómico, y se exploran estrategias de fusión de datos de diferentes biofluidos como estrategia para mejorar los modelos de clasificación y la interpretación de la dinámica metabólica de un sistema biológico.

Capítulo I

Introducción general

Las moléculas involucradas en procesos bioquímicos aportan una enorme cantidad de información sobre el estado y función de un sistema biológico. El proceso de monitorear y evaluar las variaciones en el conjunto de estas moléculas, o metabolitos, se denomina análisis metabolómico o metabolómica. Esta disciplina engloba la comprensión simultánea y ordenada de los perfiles y niveles de múltiples metabolitos y sus cambios sistemáticos. Las variaciones pueden ser causadas por diferentes factores, como ser la dieta, el entorno, o el efecto de un fármaco o enfermedades, y pueden ser estudiados en biofluidos o tejidos empleando distintos métodos quimiométricos. Se combinan técnicas analíticas, como ser la cromatografía gaseosa (GC) o líquida (LC) acoplada a espectrometría de masas (MS), la espectroscopía de resonancia magnética nuclear (RMN), y la espectroscopía infrarroja (IR), entre otras, con herramientas estadísticas [1-3].

La implementación y utilización de métodos quimiométricos y estudios metabolómicos es un área en desarrollo incipiente en el Uruguay. Esto se debe, en parte, a que la metabolómica como disciplina es más reciente que sus pares dentro de las “ómicas” (i. e., genómica, transcriptómica y proteómica). Sin embargo, ha logrado tener un gran impacto a nivel mundial, empleándose en el descubrimiento de nuevos fármacos, en estudios toxicológicos y/o de seguridad alimentaria, y en el diagnóstico de enfermedades. Su aplicación a temas de interés para los sectores de la salud y productivo del país es por ende de sumo interés[1].

El impacto más importante de la aplicación de herramientas quimiométricas en el sector productivo uruguayo se observa en la industria agropecuaria, ya que utilizando desarrollos y enfoques metodológicos comunes es posible estudiar una gran variedad de problemas. Por ejemplo, el análisis metabolómico permitiría identificar quimiomarcadores que se correlacionen con características organolépticas deseadas en diferentes productos agropecuarios, como ser carne, leche, frutas, y hortalizas. Por medio de este enfoque sería posible también seleccionar cultivares que presenten resistencia a plagas en base a la presencia o ausencia de fitomarcadores específicos, así como correlacionar sus quimiotipos y genotipos. Del mismo modo, la posibilidad de identificar y cuantificar quimiomarcadores correlacionados con la salud de animales de producción permitiría hacer un seguimiento y diagnóstico de su estado sanitario y proponer tratamientos específicos. Por lo tanto, mi plan de posgrado tuvo como objetivo afianzar al análisis metabolómico como herramienta en el estudio de algunos de los problemas detallados anteriormente, y para ello se dividió el trabajo en tres bloques [1].

El primero, descrito en el Capítulo II, aborda el análisis de los cambios en los perfiles del metabolóma sérico de ganado lechero a lo largo del parto empleando análisis metabolómico basada en RMN. Como se detalla en ese capítulo, en este periodo se observan alteraciones metabólicas que repercuten negativamente en la salud de los animales y la producción de leche. Se observan variaciones significativas en especies como acetato, betaína y creatina en etapas tempranas después del parto, mientras que otros marcadores de estrés

metabólico, incluyendo acetoacetato, β -hidroxibutirato (BHB) y colina se acumulan significativamente en el punto más alto de la producción de leche. Además, marcadas variaciones en los niveles de lactato, alantoína, alanina y otros aminoácidos revelan la activación de diferentes vías gluconeogénicas después del parto. Concomitante con un retorno a la homeostasis, se produce una normalización gradual del metabolóma sérico ocho semanas después del parto. En general, estos resultados muestran que los métodos quimiométricos basados en RMN son ideales para monitorear las manifestaciones de enfermedades metabólicas durante el período de transición y para evaluar el impacto de los esquemas de manejo nutricional en el metabolismo y salud de las vacas lecheras.

En el Capítulo III, se describe el uso de la quimiometría basada en RMN para la identificación de variedades resistentes de cebada cervecera (*Hordeum vulgare* L). Una de las principales problemáticas que enfrentan los productores cerealeros es el bajo rendimientos de los cultivos causado por plagas, particularmente los áfidos o pulgones (Hemiptera: Aphididae). Al comparar los perfiles metabólicos de extractos vegetales de genotipos de *H. vulgare* con diferente respuesta antixenótica (mecanismo de resistencia vegetal que hace a las plantas poco atractivas para los insectos plaga, evitando así la colonización, alimentación u oviposición) al pulgón verde, se lograron identificar fitomarcadores correlacionados con la misma. El uso conjunto de métodos de dereplicación química espectroscópicos y estadísticos, búsquedas en bases de datos, y finalmente adiciones de estándares sobre muestras representativas, permitieron la identificación de la colina y la betaína como metabolitos

correlacionados, respetivamente, con la preferencia y no preferencia del pulgón por la planta.

El Capítulo IV aborda el desarrollo de estrategias alternativas de procesamiento y análisis de datos de RMN en estudios metabolómicos. Por un lado, se estudió el alineamiento de espectros HOMO-2DJ, paso necesario para el uso de este tipo de datos bidimensionales en análisis metabolómico. Se describe el desarrollo de un algoritmo que adapta estrategias de alineado de datos monodimensionales a espectros HOMO-2DJ. A su vez se explora la fusión de datos provenientes de dos biofluidos, como ser suero y orina, y la utilización de las matrices de datos resultantes en análisis multivariado de rutina en metabolómica. Esta estrategia busca no solo mejorar los modelos discriminantes, sino también lograr una comprensión más exhaustiva del metabolóma.

Previo a estos capítulos se presenta una introducción general a la metabolómica, incluyendo los conceptos básicos de RMN, el procesado de datos y análisis estadístico multivariado.

1.1 Introducción a la Metabolómica basada en RMN

Para situarnos históricamente y comprender los orígenes de la RMN y la metabolómica resumiremos algunos hitos en la siguiente línea del tiempo.

- 1924** Wolfgang Pauli prueba que las características de la estructura hiperfina de los espectros atómicos se pueden explicar a través del movimiento de espín de los núcleos y el momento magnético generado [4].
- 1933** Otto Stern calcula el momento magnético del núcleo de hidrógeno mediante la técnica de haces moleculares en un campo no homogéneo [5].
- 1938** Isaac Rabi utiliza el término “espectroscopía por radiofrecuencia” para referirse a una incipiente RMN [6].
- 1945** Felix Bloch y Edward Purcell retoman la investigación sobre la resonancia magnética de la materia condensada. Ambos grupos desarrollan, de manera independiente, nuevos métodos que les permiten medir por primera vez la señal de RMN [7, 8].
- 1950** E. L. Hahn logra excitar el espín nuclear utilizando un pulso de radiofrecuencia. Esta excitación momentánea es registrada en forma de señal eléctrica inducida en una bobina (i.e., “free induction decay”, o FID) [9].

1950	H. S. Gutowsky y C. J. Hoffman descubren el fenómeno conocido como acoplamiento escalar o acoplamiento espín-espín, que explica el desdoblamiento de señales [10].
1966	Richard R. Ernst publica el trabajo “Application of Fourier Transform Spectroscopy to Magnetic Resonance”. En éste explica cómo la señal emitida por los núcleos en función del tiempo puede ser convertida en una señal en el dominio de las frecuencias mediante la transformada de Fourier (TF) [11].
1971-1976	J. Jeener aborda la idea de expandir en dos dimensiones la información contenida dentro de un experimento monodimensional de RMN [12]. Fue Ernst quien luego expandió el concepto e introdujo las primeras aplicaciones de la espectroscopia bidireccional [13].
1974	Kurt Wütrich estudia la aplicación de la RMN a la determinación estructural de proteínas, desarrollando procedimientos generales para la asignación sistemática de las frecuencias de resonancia de los diferentes núcleos magnéticos de estas biomoléculas. Demuestra que a partir de esta información se podía llegar a determinar la estructura tridimensional de proteínas en solución [14].
1974	David Hoult demuestra que la espectroscopía de RMN se puede utilizar en la detección de metabolitos en muestras biológicas intactas. Mediante RMN <i>in vivo</i> sobre tejido muscular comprueba

	que el 90% del adenosín trifosfato (ATP) celular se encuentra complejado con magnesio [15].
1984	Jeremy K. Nicholson demuestra que los espectros de RMN de suero humano pueden ser utilizados en el diagnóstico de diabetes mellitus y otras patologías [16].
2000	Nicholson propone el término “metabonómica” para hacer referencia a “la medición cuantitativa de la respuesta metabólica de naturaleza dinámica y multiparamétrica de los sistemas vivos ante estímulos patofisiológicos o bien, ante la modificación genética” [17,18].
2021	Mediante análisis metabolómico basado en RMN, Nicholson propone que la infección causada por SARS-CoV-2 no solo produce un síndrome respiratorio agudo sino también daños en múltiples órganos y alteraciones sistémicas en diversas rutas metabólicas [19].

La metabolómica forma parte de lo que se conoce como biología de sistemas, siendo la cuarta y más reciente de las ciencias “ómicas”. Sin embargo, la idea detrás de la metabolómica de integrar información de muchas variables simultáneamente no es nueva. Por ejemplo, con la rueda de la orina medieval, como ser la rueda de la orina de Pinder, se combinaban distintos atributos de este biofluido, incluyendo color, olor, textura, sabor y sedimentos, a fin de determinar la causa de una dolencia (**Figura 1**) [20].



Rueda orina de uroscopia de Epiphane Medicorum. Ullrich Pinder (1506)

Figura 1. “Uroscopía”, Davis Teniers (1610-1690) (a), Rueda de la orina de Pinder (1909) (b)

Desde ese entonces todo ha avanzado y evolucionado. Actualmente esta “ómica” busca contarnos “historias biológicas” al observar el estado metabólico de un sistema determinado [21]. Si comparamos a la metabolómica con la genómica o la proteómica, vemos que el volumen de información aumenta y que la interpretación química y biológica se vuelve cada vez más compleja. De hecho, la genómica y la transcriptómica estudian los arreglos y secuencias que existen entre cuatro bases nitrogenadas y la preoteómica investiga las secuencias, arreglos y estructuras que derivan de un grupo de 20 aminoácidos. Sin embargo, la metabolómica estudia las variaciones miles de metabolitos de bajo peso molecular. Estas variaciones son el resultado de la modulación y respuesta a estímulos externos y/o internos que derivan de la expresión de genes, y de miles de reacciones mediadas por enzimas que dan lugar al metabolóma. A su vez, este presenta una gran diversidad de especies químicas con propiedades disímiles y en un amplio intervalo de concentraciones [22]. Desde su nacimiento, el impacto

de la metabolómica ha crecido considerablemente. Esto refleja el crecimiento de las publicaciones en esta área en los últimos años (**Figura 2**) [23].

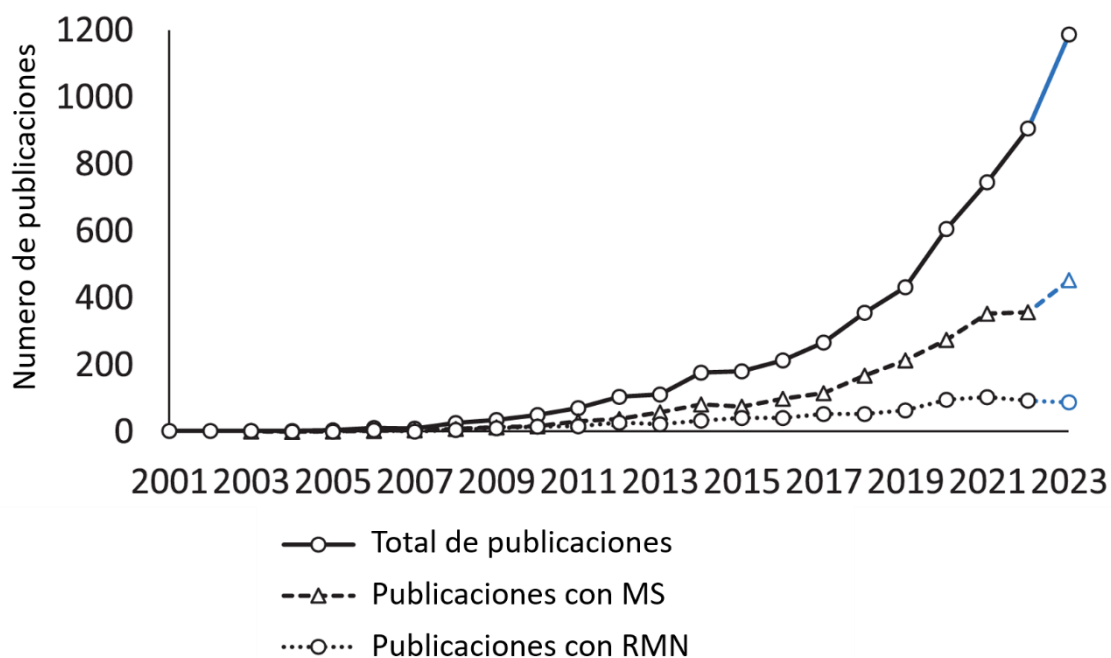


Figura 2. Representación gráfica de la progresión del número de publicaciones por año según plataforma analítica

Podemos clasificar a la metabolómica en las siguientes categorías según el enfoque que se emplee [20]:

Metabolómica no dirigida	Identificación y cuantificación de la mayor cantidad posible de los metabolitos en un sistema biológico. La preparación de la muestra no debe agrupar el mayor número de metabolitos posibles y la técnica analítica a emplear debe ser sensible a especies químicas de naturaleza diversa.
--------------------------	---

Metabolómica dirigida	Análisis cualitativo y cuantitativo de un grupo de metabolitos conocidos o de interés con una ruta metabólica específica. En algunos casos requiere amplia preparación de muestras, orientada a un grupo de metabolitos. La técnica analítica a emplear puede estar orientada a una clase de compuestos.
-----------------------	--

Análisis de huellas metabólicas	Análisis global, rápido y de alto rendimiento de muestras para proporcionar una clasificación de las mismas. Generalmente no se requiere una cuantificación o identificación de metabolitos. Es una herramienta de cribado para discriminar entre muestras de diferente origen o estado biológico.
---------------------------------	--

1.2 Fundamentos Metodológicos

Ninguna técnica analítica permite obtener una imagen completa del estado metabólico de un ser vivo. Teniendo en cuenta que, por ejemplo, el metaboloma de una planta puede tener más de 30.000 compuestos de naturaleza química diversa, esta es una meta ambiciosa [3]. Combinando diferentes métodos de extracción y tratamiento de muestra con diversas técnicas analíticas es posible concentrarse en una parte del perfil metabólico, lo que permite extraer conclusiones de una fracción del metabolismo del organismo en estudio. Todos los estudios presentados en esta tesis involucran análisis no dirigidos y emplearon RMN como técnica analítica. A continuación, se describen por lo tanto los fundamentos básicos de esta herramienta espectroscópica.

1.2.1 Bases y modelo vectorial de la resonancia magnética nuclear

La RMN es una técnica no destructiva ampliamente utilizada en química que aporta información estructural de compuestos puros y composición en mezclas complejas, así como también información sobre sus concentraciones absolutas o relativas. Se basa en las propiedades de núcleos magnéticamente activos, los cuales se comportan como si fueran partículas cargadas girando sobre un eje. Al igual que una corriente eléctrica en movimiento, generan un campo magnético, y al someterlos a un campo magnético externo (B_0) se generan orientaciones preferenciales. Dado que los núcleos poseen momento angular (espín nuclear, ρ), la interacción entre el momento magnético del núcleo y B_0 genera un torque que lleva a un movimiento de precesión (**Figura 3**). La frecuencia de esta precesión es conocida como frecuencia de Larmor (ω_0), y es proporcional a B_0 y, a un parámetro propio de cada núcleo llamado constante giromagnética (γ) (**Figura 4 y 5**).

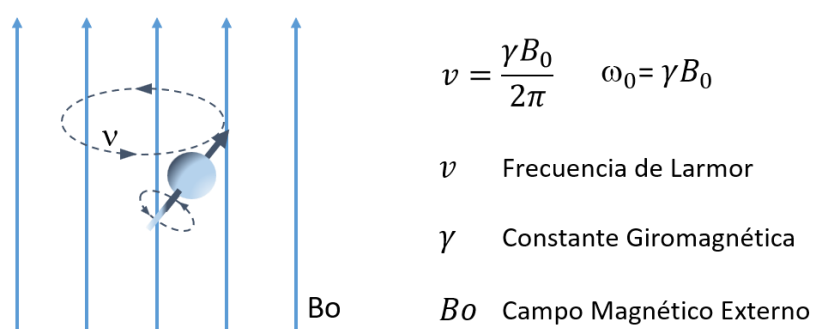


Figura 3. Esquema representativo del espín nuclear de un protón en un campo magnético externo (B_0).

Un caso particular es el de los isótopos con espín nuclear (I) igual a $1/2$ (i.e., ^1H , ^{13}C , ^{31}P , o ^{19}F). Estos núcleos pueden adoptar dos estados cuando son sometidos a B_0 , a menudo referidos como $+1/2$ y $-1/2$ o α y β . La diferencia de

energía (ΔE) entre ambos estados del espín es proporcional a la intensidad del campo magnético externo (efecto Zeeman, **Figura 4b**).

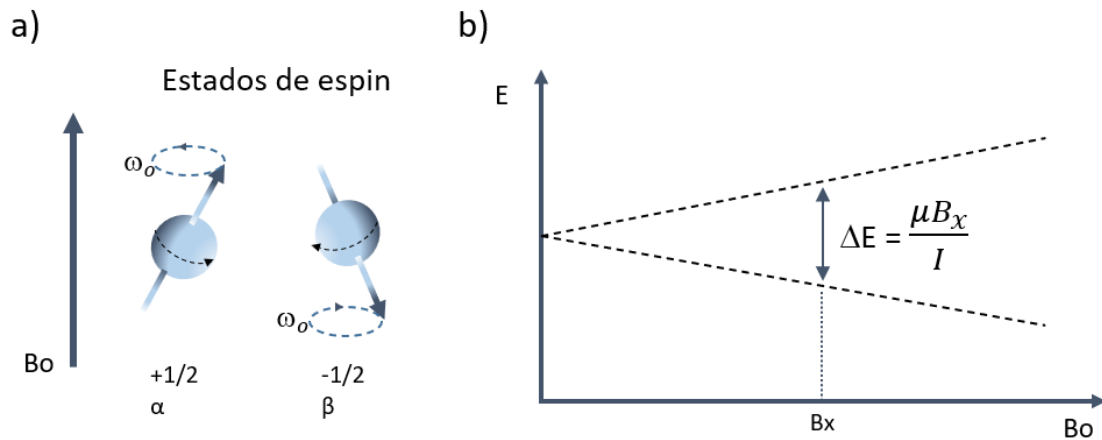


Figura 4. Esquematización de la dependencia en la diferencia energética entre los estados de espín **(a)** y el campo magnético **(b)**.

Si consideramos más de un espín del mismo tipo, y acorde con la ecuación de Boltzmann (**Ec. 1**), este ΔE lleva a un exceso de población de espines en el estado de menor energía. Teniendo también en cuenta el movimiento de precesión alrededor del eje z mencionado anteriormente, los espines se dispondrán como se presenta gráficamente en la **Figura 5**. Dado que las componentes de los momentos magnéticos en el plano $\langle x, y \rangle$ están igualmente distribuidas en el plano, y a que existe un leve exceso a lo largo de $+z$, existirá una magnetización neta (M_0) en este eje (**Figura 5**).

$$\text{Ecuacion de Boltzmann : } \frac{N_{\text{paralelo}}}{N_{\text{antiparalelo}}} = e^{\frac{\Delta E}{kT}} \quad (\text{Ec1})$$

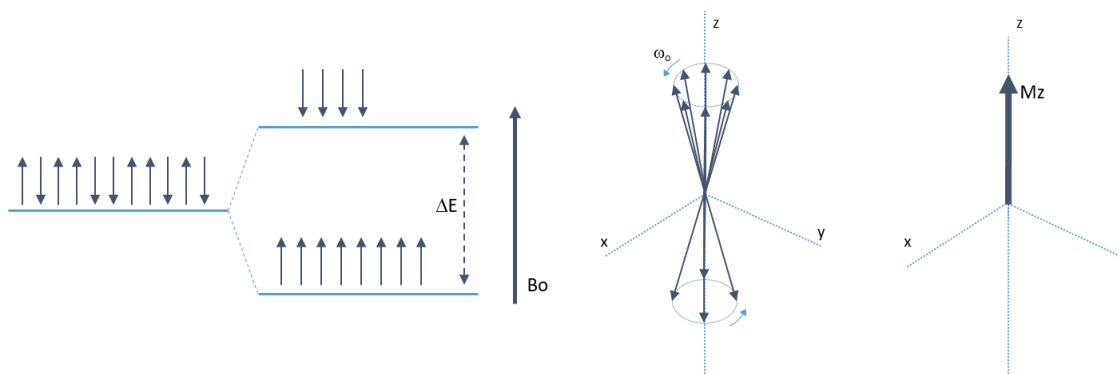


Figura 5. Representación esquemática de la magnetización neta para los espines de una molécula

Si ahora aplicamos un pulso de radiofrecuencias generando un segundo campo magnético B_1 en un plano perpendicular a B_0 , este será capaz de inducir transiciones entre los estados del espín y alterar sus poblaciones. Esto sucederá cuando la energía del pulso coincida con el ΔE que separa los estados α y β , o dicho de otro modo, cuando su frecuencia de resonancia coincida con la frecuencia de Larmor. A nivel macroscópico, durante el tiempo que dura el pulso, B_1 produce un torque sobre los espines nucleares, rotándolos del eje z al plano $\langle x, y \rangle$. Cuando el pulso deja de aplicarse, los espines dejan de rotar y quedan en una dirección que puede no ser la de equilibrio (**Figura 6**).

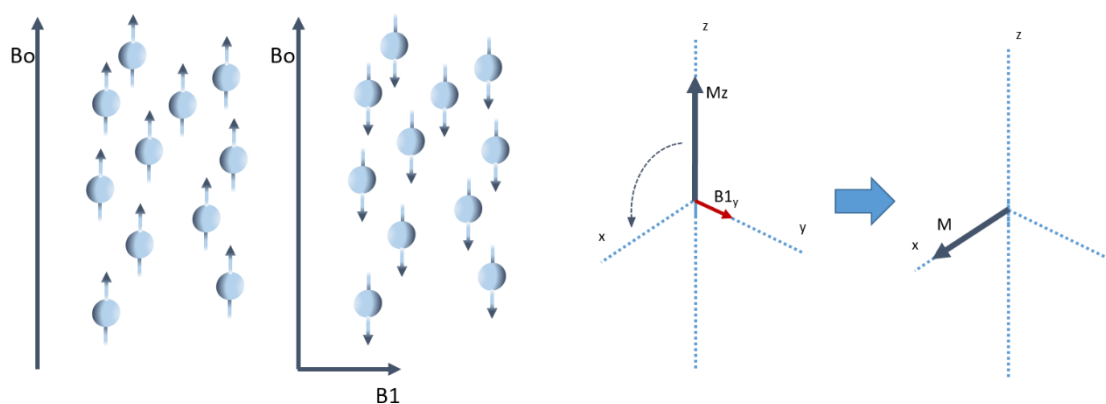


Figura 6. Representación vectorial de la aplicación de un pulso de radiofrecuencia (B_1) sobre nuestro sistema de espines (M_z)

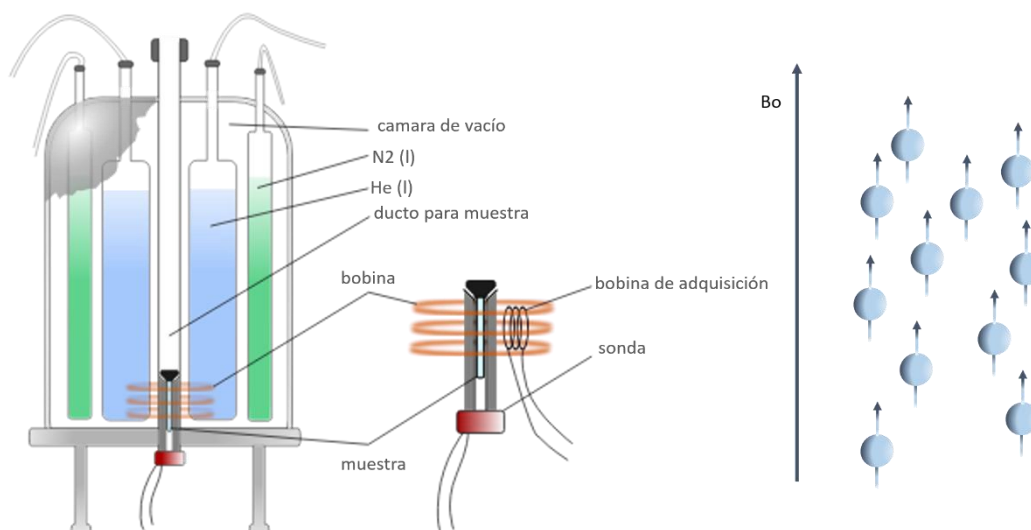


Figura 7. Esquema de un equipo de RMN.

Mediante diferentes procesos de relajación, los espines nucleares recuperan paulatinamente el equilibrio. Si posicionamos una bobina en el plano $\langle x, y \rangle$, los cambios en la magnetización (M) generarán corrientes inducidas que nos permitirán monitorear la evolución de esta en el plano (**Figura 7**). Dado que M precesiona a la frecuencia de Larmor y está afectada por relajación en el plano $\langle x, y \rangle$, la corriente inducida en la bobina variará de forma cosinusoidal y decaerá exponencialmente (**Figura 8**). Ya que los núcleos en una molécula tienen diferente entorno químico (ver sección 1.2.2) también tendrán diferentes frecuencias de Larmor, y la señal resultante que recoge la bobina corresponde a la sumatoria de todas las corrientes inducidas. Esto último es lo que conocemos como FID y es poco práctica para extraer información de los espines en la molécula en forma directa. Es por esto que recurrimos a la transformada de Fourier (TF), una operación matemática que nos permite convertir señales del dominio del tiempo al dominio de las frecuencias (**Figura 9**).

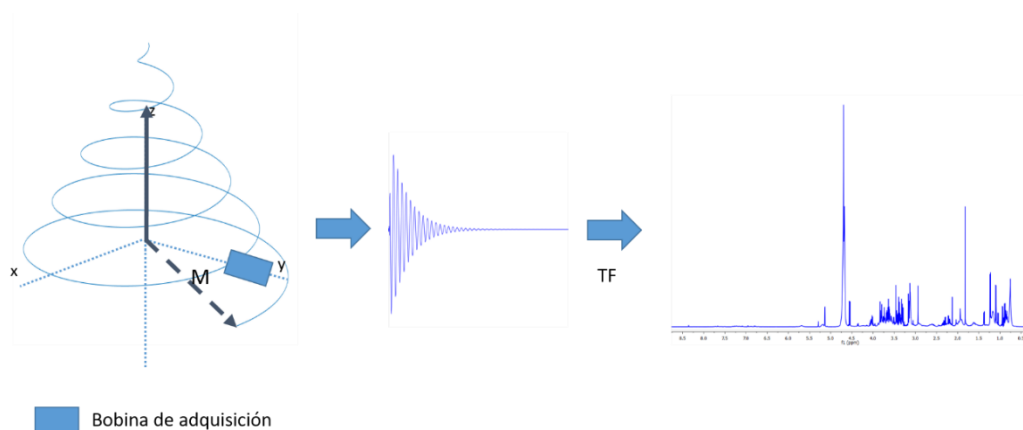


Figura 8. Representación de adquisición de una FID y posterior aplicación de TF sobre la misma para obtener un espectro de RMN.

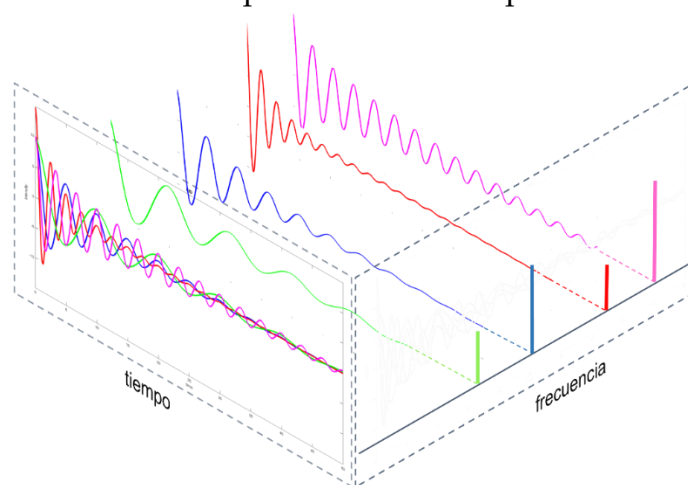


Figura 9. Representación gráfica de la descomposición de una FID (en el dominio del tiempo) a un espectro (en el dominio de las frecuencias) mediante TF.

1.2.2 Apantallamiento

La frecuencia de resonancia de cada núcleo está vinculada al campo magnético efectivo que lo rodea (B_{ef}) y se puede expresar según la siguiente ecuación:

$$\nu = \frac{B_{ef}}{2\pi} \quad (Ec2)$$

El B_{ef} está determinado por B_0 , y un campo magnético B' que depende de los efectos electrónicos locales que generan todos los núcleos enlazados o cercanos al núcleo observado. Esto último es lo que conocemos como entorno químico.

Dado que el campo magnético efectivo va a ser diferente para cada tipo de núcleo no equivalente (i.e. entorno químico diferente), también lo va a ser la diferencia de energía entre sus estados α y β , y en definitiva también su frecuencia de resonancia característica.

Hasta este punto se hizo una introducción de las bases de la RMN y describió de manera simplificada la adquisición de un espectro de ^1H RMN. Sin embargo, los equipos de RMN modernos nos permiten realizar una gran variedad de experimentos mono- o multidimensionales sobre ^1H y un gran número de núcleos magnéticamente activos, incluyendo ^{13}C , ^{15}N , ^{31}P , entre otros. Estos nos permiten, por ejemplo, atenuar las señales de ciertas clases de macromoléculas en los espectros de ^1H (i.e., Carr-Purcell-Meiboom-Gill, o CPMG), obtener correlaciones entre núcleos enlazados a uno o más enlaces (i.e., “Heteronuclear Single Quantum Coherence”, o HSQC; “Homonuclear Correlation Spectroscopy”, o COSY; “Total Correlation Spectroscopy”, o TOCSY; “Heteronuclear Multiple Bond Correlation”, o HMBC), identificar núcleos próximos en el espacio (i.e., “Nuclear Overhauser Enhancement Spectroscopy”, o NOESY), o separar información de corrimientos químicos y acoples escalares (i.e., “Homonuclear 2D J Spectroscopy”, o HOMO-2D). En su conjunto, todos estos experimentos nos brindan información estructural, de un compuesto, ya sea estando este puro o formando parte de una mezcla compleja. Este último es el caso de los fluidos biológicos, como suero y orina, y los extractos de tejidos, muestras típicas en metabolómica.

1.3 Esquema general de trabajo en metabolómica

Los fluidos biológicos y los extractos de tejidos son mezclas complejas de metabolitos, y sus espectros de ^1H RMN son por ende relativamente complejos. A su vez, un único espectro de muestras de este tipo no nos permite interpretar de manera confiable lo que está sucediendo en un sistema dinámico como lo es un organismo. Es por ello que el análisis metabólico involucra un número representativo de muestras y herramientas estadísticas robustas. Los datos crudos que se obtienen a partir de plataformas analíticas como RMN o MS no pueden ser empleados directamente para estos análisis, y requieren un procesamiento previo. A continuación, se brinda una descripción de este proceso, así como de otros pasos necesarios en un esquema de trabajo típico de análisis metabólico (**Figura 10**).

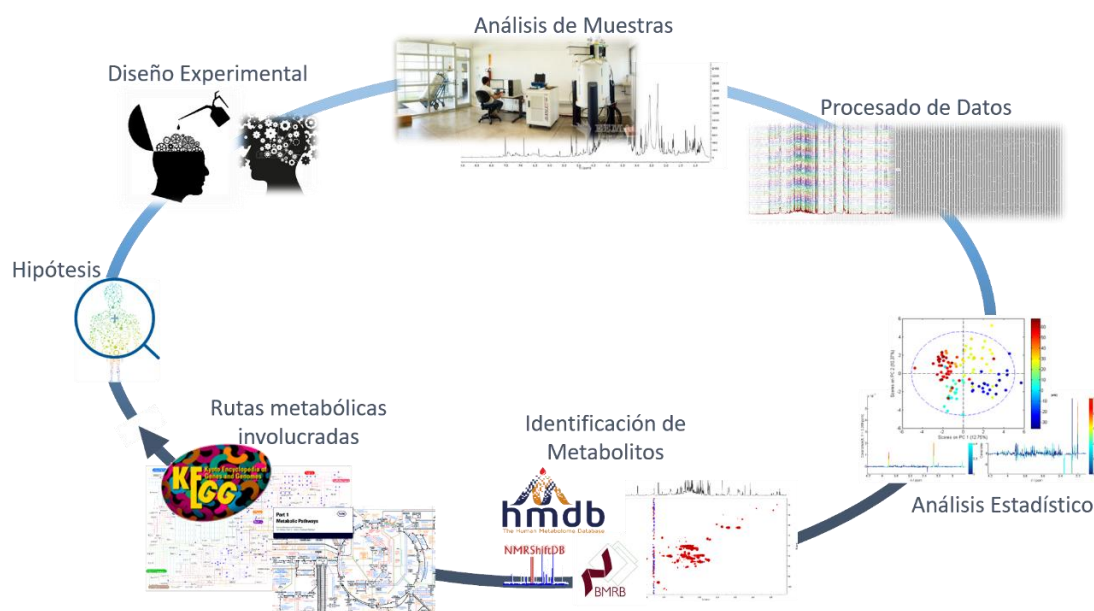


Figura 10. Esquema de trabajo en metabolómica.

1.3.1 Análisis de muestras y procesamiento de datos

En el esquema de trabajo el diseño experimental es clave para que los resultados sean confiables. Algunos aspectos relevantes son, por ejemplo, la determinación del número de réplicas biológicas a emplear, los métodos de recolección y almacenamiento de muestras, y el procesamiento de las mismas. Además, es importante definir la técnica analítica a emplear (RMN, CG-MS, HPLC-MS), así como las condiciones experimentales de la misma. Concentrándonos en RMN, los aspectos importantes a definir son, por ejemplo, la secuencia de pulsos a emplear y los parámetros del experimento, incluyendo el número de barridos, el tiempo de reciclado, el tamaño espectral, etc. Una vez obtenidos los espectros nos enfocaremos en el esquema de preprocesado de datos. Esto involucra desde el procesamiento individual de cada espectro hasta la obtención una matriz de datos con la información de todos los espectros de interés adecuada para los MVDA [21-25].

El esquema de adquisición de datos de RMN descritos en la sección 2.2.3 lleva a la generación de ciertos artefactos en los espectros de ^1H RMN que deben ser corregidos previo a su interpretación. Los más sobresalientes son los desajustes en la fase y de la línea de base como se muestra en la **Figura 11** [24, 25]. Por lo tanto, el primer paso involucra la corrección de la fase de los espectros buscando que todas las señales sean absorptivas, así como la línea de base no presente gradientes.

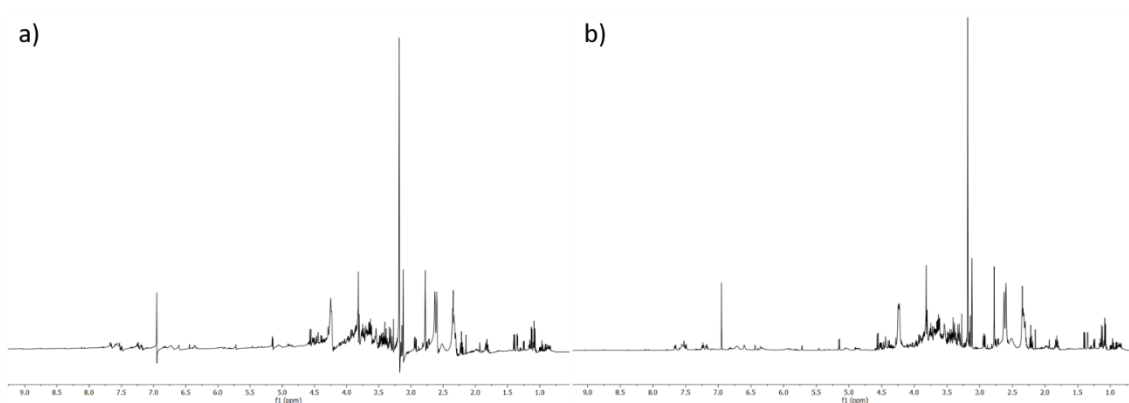


Figura 11. Ejemplo de un espectro de extracto acuoso de hojas de cebada antes de la corrección de fase y línea de base **(a)**, y después de las correcciones **(b)**.

Posteriormente es necesario referenciar los espectros, utilizando la señal de un estándar interno como el TSP o TMS, o la de un metabolito presente en todas las muestras. En este último caso, es deseable que nuestra referencia no presente cambios en el corrimiento químico por variaciones en pH o fuerza iónica. Independientemente del referenciado, tanto el pH como la fuerza iónica pueden variar entre las muestras analizadas, afectando los corrimientos químicos de otros metabolitos. La corrección de estas variaciones o alineado global es uno de los pasos más importantes en el preprocesado de los datos. Cada espectro puede verse como un vector de “n” puntos, y cada punto es una variable en nuestro set de datos. De no ser correctamente alineadas, una misma variable tomaría valores diferentes de corrimiento químico en distintos espectros. A los efectos de los análisis estadísticos, esto último es equivalente a sean consideradas como variables diferentes. Existen varios algoritmos de alineamiento, como ser “Interval correlation optimised shifting algorithm” (ICOShift), “peak alignment fast Fourier transform” (PAFFT), Recursive alignment fast Fourier transform” (RAFFT), “correlation optimized warping” (COW), siendo el primero

uno de los más empleados [26, 27]. Una vez que los datos están alineados, se procede a la eliminación de zonas de no interés, como ser señales residuales de solventes y regiones sin señal. Cada espectro es luego normalizado, ya sea al total del área, al área de una señal de interés, o por la masa de extracto, o, alternativamente, empleando algoritmos de normalización como la normalización por cociente de probabilidad (QPN) [26]. Finalmente se realizan operaciones de centrado, escalado o transformaciones, que a diferencia del normalizado se realiza para cada una de las variables. Estas operaciones están compuestas por parámetros estadísticos que se desprenden de cada variable, como ser la desviación estándar, la media, o la mediana (**Tabla 1** y **Figura 12**) [28,29]. Una vez que la matriz de datos está lista es exportada en un formato adecuado para su lectura por los programas de análisis estadístico.

Tabla 1. Factores de escalados. Adaptado de Ref. 29.

Tipo	Método	Formula	Objetivo	Ventaja	Desventaja
Centrado	Centrado	$\tilde{x}_{ij} = x_{ij} - \bar{x}_i$	Desplaza los datos al origen (centrado). Se centrará en las diferencias y no en las similitudes de los datos.	Elimina el desplazamiento de los datos	Si los datos son heteroscedásticos, el efecto de esto método de pretratamiento no siempre es suficiente
Escalados	Auto-escalado	$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_i}{s_i}$	Comparar metabolitos basados en su correlación	Todas las variables se consideran con igual importancia	Inflación de los errores de medida
	Rango	$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_i}{x_{imax} - x_{imin}}$	Compara los metabolitos relativo al rango de respuesta biológica	Todos los metabolitos son considerados igualmente importantes. El escalado se relaciona con la biología	Inflación de los errores de medida. Baja sensibilidad a los datos fuera de tendencia (outliers)
	Pareto	$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_i}{\sqrt{S_i}}$	Reduce la importancia de las variables más grandes. Mantiene los datos parcialmente intactos	Conserva más similitud con los datos originales que el auto-escalado	Disminuye la sensibilidad a las variaciones relativas cuando estas vienen valores grandes. (grandes “fold changes”)

	Vasto	$\tilde{x}_{ij} = \frac{(x_{ij} - \bar{x}_l)}{s_i} \cdot \frac{x_i}{s_i}$	Enfocado en los metabolitos con variaciones pequeñas	Puede tomarse en cuenta la información de clases en los coeff. de variación (CV)	No adecuada para grandes variaciones en los datos sin estructura de grupos
	Nivel	$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_l}{\bar{x}_l}$	Énfasis en las variaciones relativas	Bueno para identificación de biomarcadores	Inflación de los errores de medida
Transformaciones	Logarítmica	$\tilde{x}_{ij} = \log(\hat{x}_{ij})$ $\hat{x}_{ij} = \tilde{x}_{ij} - \bar{\tilde{x}}_{lj}$	Corregir la heteroscedastisidad, es un pseudoescalado	Reduce la heteroscedastisidad	Problemas con valores con grandes des. estándar y ceros
	Potencia	$\tilde{x}_{ij} = \sqrt{\hat{x}_{ij}}$ $\hat{x}_{ij} = \tilde{x}_{ij} - \bar{\tilde{x}}_{lj}$	Corregir la heteroscedastisidad, es un pseudoescalado	Corregir la heteroscedastisidad, es un pseudoescalado. No presenta problema con valores pequeños	La elección de usar la raíz cuadrada es arbitraria

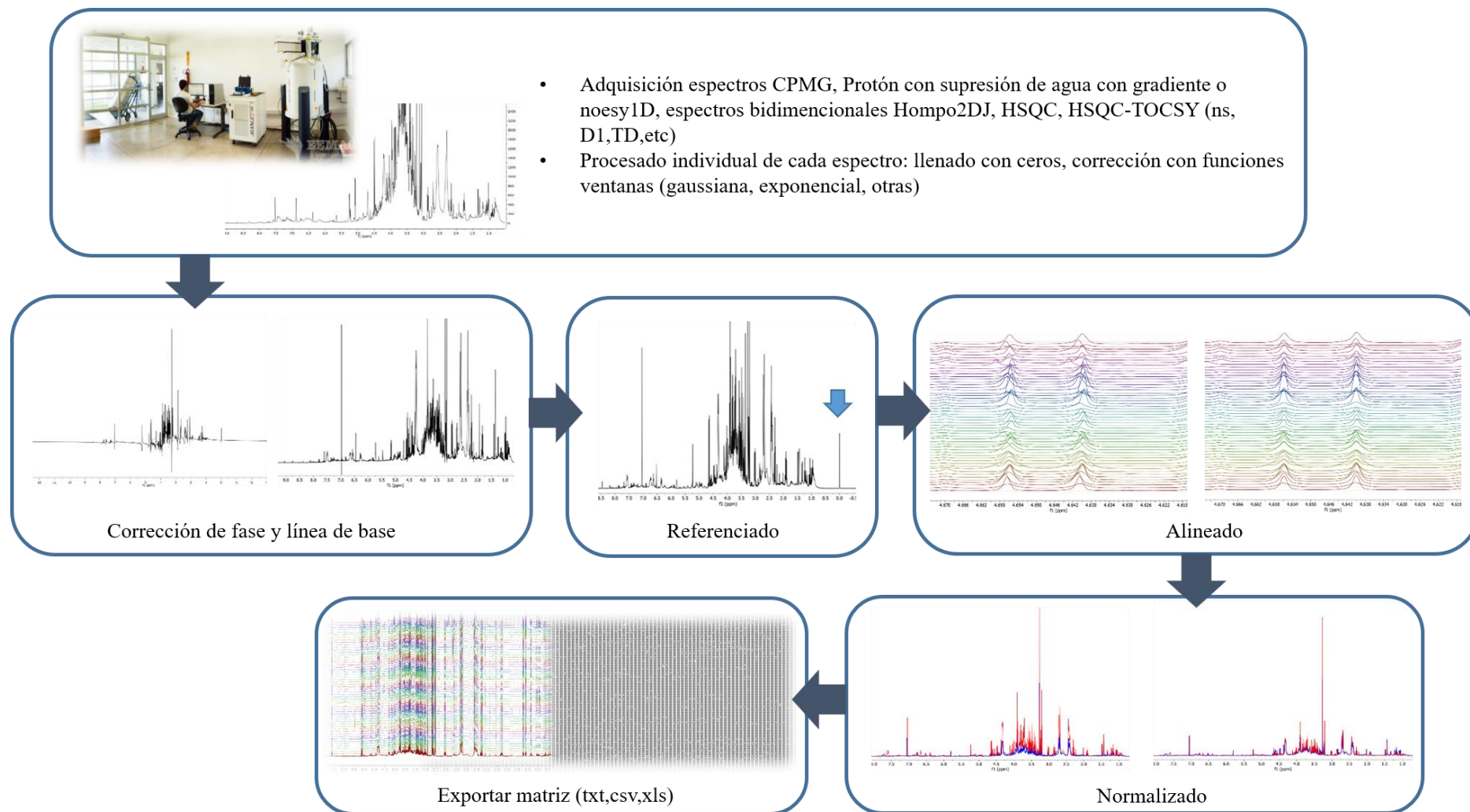


Figura 12. Esquema de pretratamiento y procesamiento de datos.

1.4 Análisis Multivariado de Datos

Para entender la idea general del análisis multivariado de datos, o MVDA por sus siglas en inglés, podemos recurrir a un ejemplo sencillo. Supongamos que queremos distinguir a un grupo de personas entre hombres y mujeres, pero solo podemos medir ciertos atributos, como ser la altura, el peso, el color de ojos, el color y largo del pelo, el ancho de hombros, cintura, cadera, y el tamaño del pie. Algunos de estos nos permiten diferenciar fácilmente a hombres de mujeres, mientras que otros no se asocian con ningún grupo. Por ejemplo, podemos asociar una mayor altura con los hombres y pelo más largo con las mujeres, mientras que el color de ojos no se asociaría con ninguno de los grupos. Sin embargo, clasificar correctamente ponderando un solo atributo como la altura o el largo del cabello claramente resulta insuficiente, y llevará a tener muchas personas mal clasificadas. Una persona pondera toda la información que obtiene en forma conjunta, existiendo atributos que contribuyen más o menos a diferenciar entre hombres y mujeres. Análogamente, los MVDA buscan integrar los valores de diferentes variables y generar modelos que nos permitan clasificar, pudiendo identificar en estos modelos las variables más relevantes para la clasificación [3]. Volviendo a nuestro ejemplo, el color de ojos será un atributo poco relevante para el modelo, tal vez la altura será moderadamente relevante, mientras que la relación entre ancho de hombros y cadera será muy relevante (Figura 13).

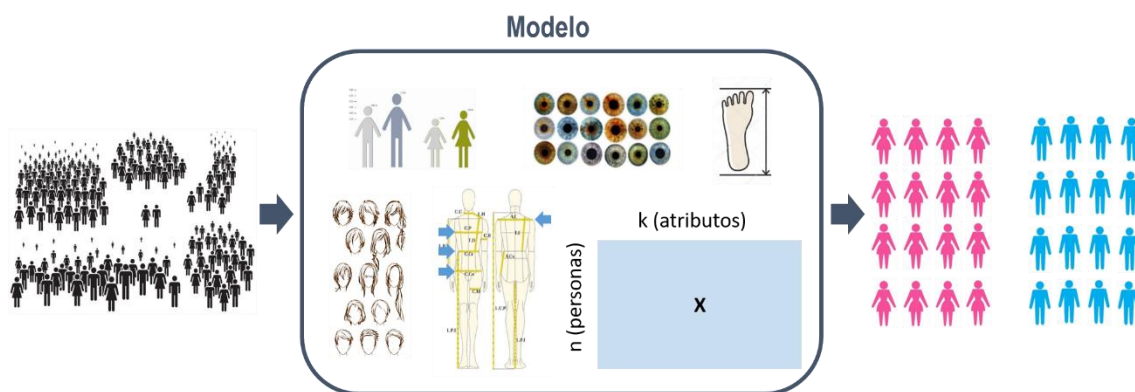


Figura 13. Representación gráfica de un modelo de clasificación

Continuando con nuestro ejemplo, el número de personas y los distintos atributos medidos para estas quedan expresados en forma de una matriz X . En esta se ordenan el número de personas como n observaciones representados con vectores (filas) de k variables o atributos (columnas), respectivamente (**Figura 13**). Esta matriz es luego empleada en los MVDA, algunos de los cuales se resumen y clasifican según el siguiente cuadro.

No supervisados

Agrupamiento Jerárquico (HC)	Forma grupos en orden jerárquico en función de la similitud entre los mismos. Esto es determinado por el cálculo de una métrica entre las observaciones (i.e., la distancia euclidiana, de Manhattan, máxima o Mahalanobis) y comparaciones de dos en dos por un criterio de enlace que determina la disimilitud entre ambas [30].
------------------------------	--

Análisis de Componentes Principal (PCA)	Es un método de reducción de la dimensionalidad de los datos, aumentando la interpretabilidad y al mismo tiempo minimizando la pérdida de información. Este análisis crea nuevos ejes de coordenadas ortogonales a partir de las variables originales que maximizan sucesivamente la varianza de nuestra matriz de datos [31].
---	--

Supervisados

k vecinos más cercanos (knn)	Este método clasifica las observaciones en función de los núcleos (clases) que se definan y las distancias entre las observaciones, número de vecinos y otros atributos [32].
Proyección de Estructuras Latentes (PLS) y Proyección de Estructuras Latentes Ortogonales (OPLS)	Es un método que busca una relación cuantitativa entre dos matrices de datos X, que posee por ejemplo datos espectrales o cromatográficos, e Y, que contiene valores cuantitativos, como ser, las concentraciones de metabolitos. También puede ser utilizado en el análisis discriminante (PLS-DA) donde la matriz Y contiene valores cualitativos (grupos o clases). El OPLS busca separar la variación sistemática en X en dos partes, una que está relacionada linealmente con Y, y otra que es ortogonal a Y [33].

Máquinas de soporte vectorial (SVM)	Se basa en la construcción de un hiperplano que permita la separación óptima entre las clases, es decir que tengan la mayor distancia a este [34].
Redes neuronales (NN)	Una red neuronal se basa en un conjunto de algoritmos que procesan los datos de entrada y los van transfiriendo de un conjunto de algoritmos, o capas, a los siguientes. Se transfieren secuencialmente los datos de entrada a las neuronas en una capa a las de una capa siguiente, repetidamente, miles de veces, donde cada ciclo o iteración se conoce como una época. A su vez, cada neurona puede aceptar entradas de muchas otras neuronas. El “conocimiento” de la NN se captura en “valores de peso”. Los pesos son análogos a los coeficientes o cocientes de probabilidad en un modelo estadístico tradicional. Esto crea redes o arboles de decisión que permiten obtener modelos de clasificación [35].

En este trabajo nos enfocaremos en la aplicación de PCA, PLS y OPLS sobre datos provenientes de espectros de RMN. Nuestra matriz X podrá ser directamente descompuesta usando métodos de reducción de dimensionalidad no supervisados (PCA) o correlacionada con una matriz respuesta Y para las n

observaciones. A continuación, se presenta una descripción de estos métodos de análisis estadísticos.

1.4.1 Análisis de Componentes Principales (PCA)

El PCA es un método de reducción de dimensionalidad en el cual expresamos nuestras observaciones en nuevos ejes de coordenadas llamados componentes principales (CP). Estas son combinaciones lineales de las variables originales y contienen, en términos de varianza, la mayor parte de la información significativa de las variables de nuestra matriz de datos. Para visualizar esto gráficamente, consideremos nuestras observaciones en un espacio k -dimensional, siendo cada dimensión una variable o punto en un espectro y cada espectro una observación (n). Limitemos esto para simplificar aún más, por ejemplo, a un set de datos X_{nk} , de $n = 30$ observaciones y $k = 3$ variables que han sido centrados ($x_i - \bar{x}$). Buscar la dirección de mayor varianza equivale a buscar la CP donde las distancias en las proyecciones de las observaciones (muestras) sobre la componente sea la mayor. Lo que equivale a maximizar la varianza (o minimizar los residuales de las proyecciones sobre la CP). En este espacio tridimensional, la CP va a ser la recta en la dirección de mayor varianza (o menor suma de residuales de la proyección sobre la CP, **Figura 14**). Tomando solamente dos observaciones y rotando sobre el centro la CP que acabamos de trazar, podemos observar que hay un caso donde los residuales son menores (o las distancias de las proyecciones son mayores). Esta será la componente que mejor ajuste en nuestro PCA.

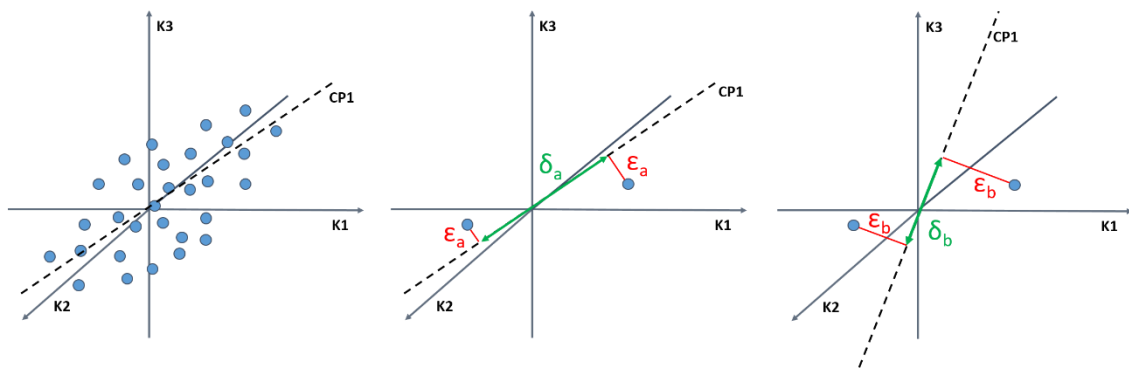


Figura 14 Esquema de dos posibles CP y como ajustan los datos sobre estas.

Como puede verse en el gráfico central los residuales (ϵ) son menores y la distancia entre las proyecciones (δ) las mayores, en comparación con la gráfica a la derecha.

Luego buscaremos otras direcciones ortogonales que expresan mayor varianza, en nuestro caso la siguiente dirección de mayor varianza será la segunda CP (CP₂). Este nuevo eje de coordenadas CP₁ vs CP₂ nos determina un plano sobre el cual proyectaremos nuestras observaciones (**Figura 15**).

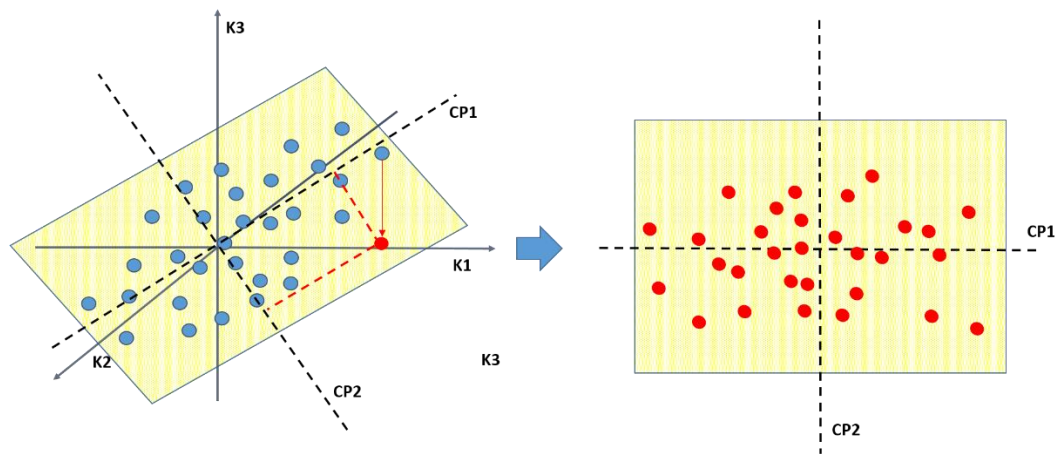


Figura 15. Ejemplo de proyecciones en un PCA.

Cada CP es una composición lineal de las variables originales, donde cada variable tiene un peso equivalente a la proyección de las variables sobre nuestras componentes (Figura 16).

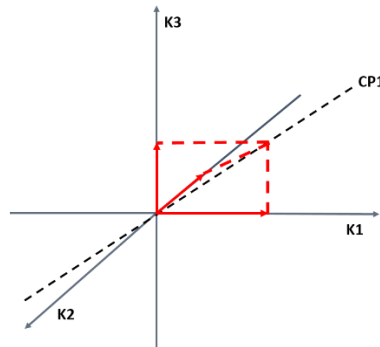


Figura 16. Proyección de las tres variables k_i sobre la CP1.

Luego de este ejemplo sencillo en tres dimensiones, donde explicamos gráficamente la idea general de un PCA y como este se construye, podemos pasar a nuestro caso k -dimensional. No podemos dibujar más de tres dimensiones, pero la idea y fundamentos para un caso con k dimensiones sigue siendo la misma. En este ultimo la matemática para desarrollar un PCA es algebraica. Como resumimos a continuación, el método descrito se basa en la descomposición de valores singulares. En este método, para nuestra matriz original X_{nk} , calcularemos primero la matriz de covarianza C_{kk} . Luego para nuestra matriz C , debemos encontrar los vectores propios q , y la matriz diagonal de valores propios Λ que satisfagan la siguiente ecuación (Ec. 3) [36].

$$C \cdot q_k = \lambda_k q_k \text{ (Ec. 3)}$$

Donde q_i es uno de los vectores propios de C y λ_i es el valor propio para ese vector. Podemos expresar lo anterior también de la siguiente manera, (Ec. 4):

$$C = Q\Lambda Q^T \text{ (Ec. 4)}$$

Luego se ordenan los vectores y valores propios en orden decreciente según el valor propio. Nuestro componente principal será el vector que, a través del espacio k -dimensional, maximice la varianza de todas las observaciones proyectadas sobre él. Este es el vector propio de la matriz de covarianza asociada con el valor propio más grande. Seleccionaremos como nuestro nuevo eje de coordenadas a las dos direcciones que nos determinan los vectores propios con los valores propios más grandes. Finalmente, en nuestro nuevo eje de coordenadas debemos encontrar las proyecciones de nuestras observaciones (n) en el mismo. En este nuevo sistema de coordenadas la contribución proporcional de cada variable en cada CP está dada en relación al total de la varianza capturada por la misma. La relación proporcional la podemos obtener a partir de vector propio de esa componente y la varianza que captura, la cual se corresponde con la raíz cuadrada del valor propio (Ec. 5) [37]. A esto último es a lo que llamamos de valores de cargas (p) o vector de valores de carga (P)

$$P_k = Q_k \cdot \sqrt{\lambda_k} \text{ (Ec. 5)}$$

Finalmente, podemos entonces representar a un PCA como lo describe la

Ec. 6:

$$X = TQ^T + E \text{ (Ec. 6)}$$

En la **Ec. 6** X es nuestra matriz de datos, T los vectores de puntuación, Q los coeficientes, y E la matriz de residuales (**Figura 17**).

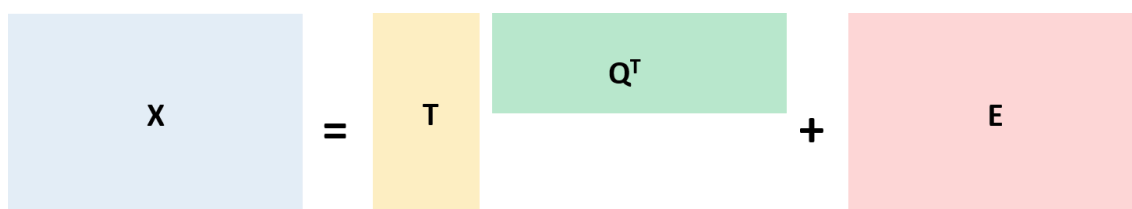


Figura 17. Esquematización canónica de un PCA.

1.4.2 Proyecciones de Estructuras Latentes (PLS) y Proyecciones de Estructuras Latentes Ortogonales (OPLS)

El análisis de PLS es un método que busca una relación cuantitativa entre una matriz de datos (X , datos espectrales) y una matriz respuesta o clase (Y). Esta última puede contener valores cuantitativos, como por ejemplo concentraciones de metabolitos. Pero en caso de realizarse un análisis discriminante (PLS-DA), la matriz respuesta contendrá valores cualitativos de pertenencia a una clase, como ser sexo, tratamiento de las muestras, o estado sanitario [38]. A diferencia del PCA, este método busca las direcciones de mayor varianza que mejor explican o ajustan a la matriz respuesta. Para poder entender esto de forma gráfica podemos pensar en la matriz del caso anterior, X_{nk} , donde ahora a las observaciones les

asignaremos dos clases. El PLS en este caso mueve las componentes principales, a las que ahora llamaremos variables latentes (VL), para que estas ajusten a la matriz Y. Si proyectamos nuestras observaciones sobre la VL podemos observar cómo mejora la separación entre las clases (**Figura 18**).

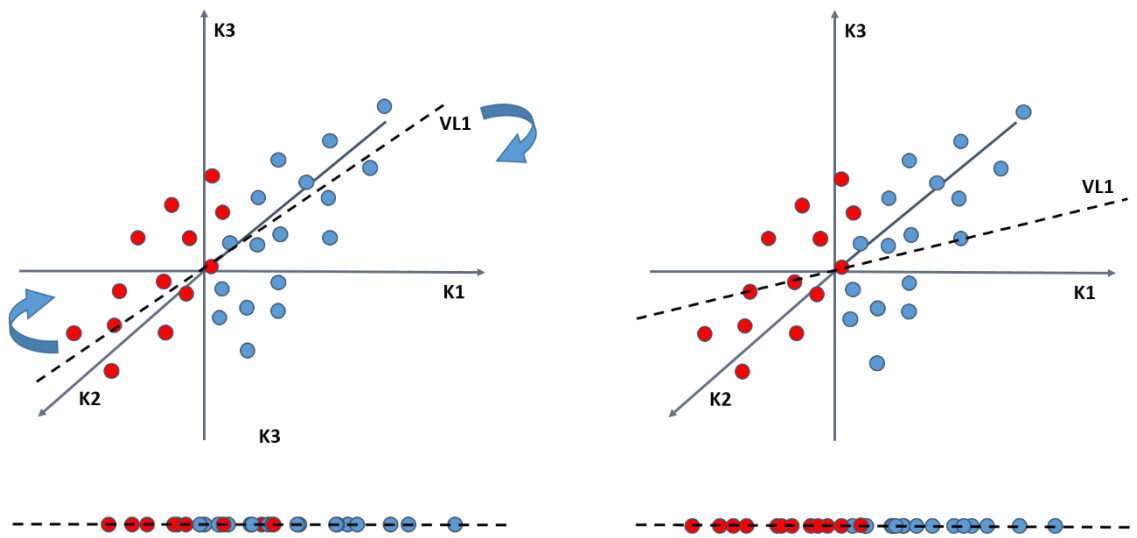


Figura 18. Proyección de las observaciones sobre la VL₁ en dos direcciones diferentes. Como vemos la proyección de la derecha separa mejor las observaciones siendo esta la dirección de nuestra VL en el PLS.

Los modelos de PLS los podemos describir a partir del algoritmo de NIPALS (Nonlinear estimation by Iterative Partial Least Squares) [38, 39]. Este consiste en ajustar un modelo de regresión entre los vectores de puntuación de la matriz X con Y. En el análisis por PLS se descomponen y representan las matrices X e Y de la siguiente forma (**Ec. 7**, **Ec. 8** y **Figura 18**):

$$X = TQ^T + E \text{ (Ec. 7)}$$

$$Y = UG^T + F \text{ (Ec. 8)}$$

En este modelo las matrices X e Y representan nuestra matriz de datos y la matriz respuesta, respectivamente. A modo de ejemplo, X sería nuestra matriz de datos espectrales, e Y nuestra matriz de clases. Las matrices T y U representan a las matrices de vectores de puntuación en nuestro modelo. Por último, las matrices Q y G corresponden a los coeficientes, y las matrices E y F a las matrices de residuales, (**Figura 19**) [38, 39].

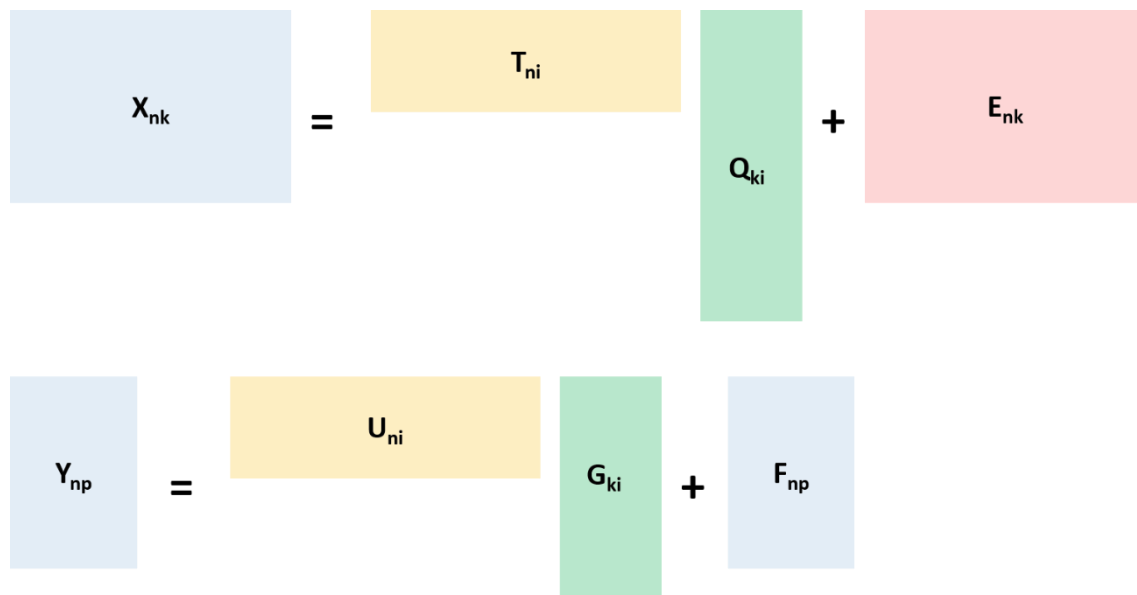


Figura 19. Esquematzación canónica de un modelo de PLS.

El análisis de PLS no descompone X e Y en forma separada como si se realizara un PCA para cada matriz, sino que cruza información entre ambas para descomponerlos buscando explicar Y en X con F y E siendo lo menor posible.

Finalmente para estimar cuál es el número de VL_s óptimo para nuestro modelo podemos basarnos en los parámetros estadísticos R^2 y Q^2 , los cuales se definen a continuación.

$R^2 = 1 - \frac{Var(\widehat{X})}{Var(X)}$	Podemos determinar R^2 para una componente en función de la varianza de $\widehat{X}$ y X, y así para cada componente. Nos permite estimar el ajuste de nuestro modelo a los datos en función de la varianza explicada por cada componente. Se puede expresar también como el valor acumulado del total de componentes empleadas en el modelo (R^2_{cum}). Podemos tomar como bueno un modelo con valores de R^2 superiores a 0,8 [40].
$Q^2 = \frac{1 - PRESS}{RSS}$	El Q^2 nos permite estimar la predictibilidad del modelo generado. Para esto debemos calcular la suma cuadrada de los errores residuales de predicción (PRESS) a partir de la gráfica de u vs t, y la suma cuadrada de los errores residuales (RSS) a partir de proyectar X en nuestra VL, y el acumulado del total de las componentes del modelo (Q^2_{cum}). Es deseable que el Q^2_{cum} sea superior a 0,5 [40].

La interpretación de los valores de carga en un PLS puede ser compleja. Al igual que en un PCA, la varianza o información asociada a Y puede estar distribuida en más de una componente. Por lo cual para su interpretación se deberían analizar todas las VL del modelo. Además, se debería considerar que luego de ajustar la primer VL en un PLS estamos trabajando sobre las matrices

deflactadas de X [33,37-39]. Por lo antes mencionado, para poder interpretar que variables son relevantes para explicar Y solemos usar el análisis por OPLS. Este análisis busca descomponer la varianza de X en dos partes, una linealmente relacionada a Y , y otra no correlacionada (ortogonal). Esto facilita la interpretación ya que toda la información relevante queda contenida en la VL_1 (componente predictiva). Podemos describir el modelo de OPLS con una componente predictiva ($T_p P_p^T$) y otra componente no predictivo u ortogonal ($T_o P_o^T$). Como podemos ver en la **Figura 19**, el PLS tiene información correlacionada a Y en VL_1 y VL_2 . En cambio, el OPLS “rota” las observaciones dejando la varianza correlacionada con Y solamente sobre la VL_1 , facilitando así su interpretación (**Figura 20**) [38]. Podemos describir el modelo con las Ec. 8 y Ec. 9:

Modelo X: $X = T_p P_p^T + T_o P_o^T + E$ (Ec 8)

Modelo Y: $Y = U_p G_p^T + F$ (Ec 9)

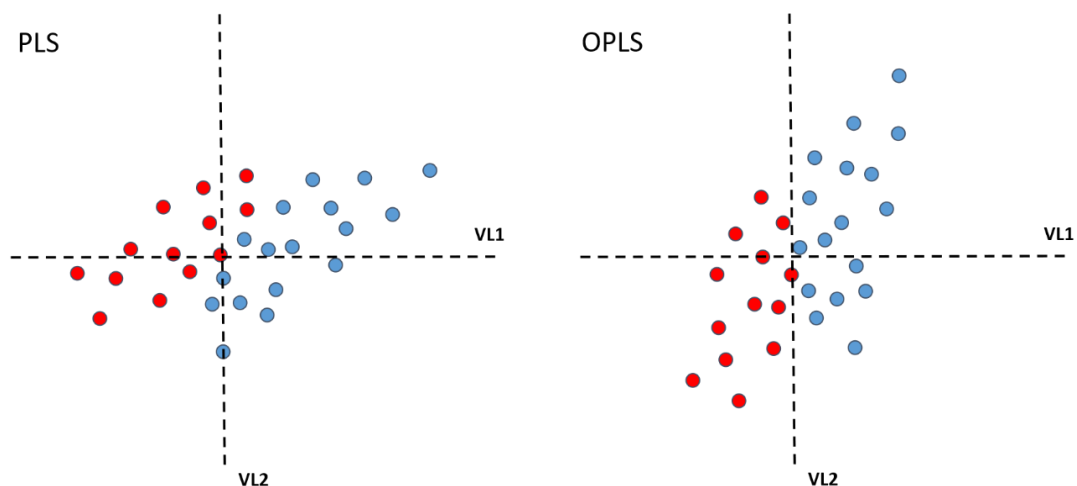


Figura 20. Representación gráfica de un PLS vs OPLS. Como se puede apreciar la variable discriminante en el OPLS es la VL₁.

1.5 Identificación de variables predictivas

Una vez que tenemos nuestros modelos de clasificación, muchas veces es necesario identificar cuáles son las variables más relevantes para explicar Y. A continuación, se describen distintas formas de visualización de los resultados.

1.5.1 Gráfico de valores de carga

Se construye a partir del vector de pesos del OPLS, al cual le aplicamos el inverso del escalado que se haya realizado en X al generar los modelos. Luego se grafica este vector contra las variables, que en el caso de RMN son corrimientos químicos, y donde el color del gráfico está dado por el vector de pesos. Este gráfico es de fácil interpretación dado que es similar a un espectro. Las señales positivas en el gráfico correlacionan con los valores positivos de la VL₁, y los valores negativos correlacionan con sus valores negativos (**Figura 21**).

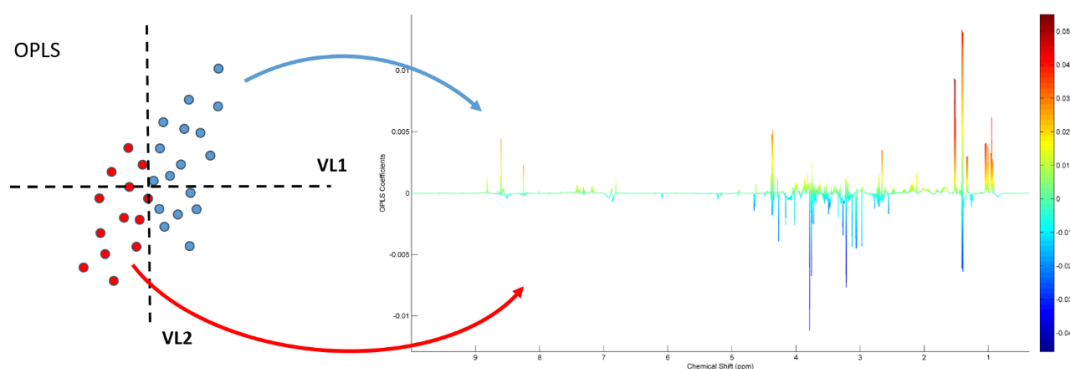


Figura 21. Gráfico de OPLS con su correspondiente gráfico de valores de carga.

1.5.2 Gráfico de Volcán ("Volcano plot")

Se construye una gráfica del logaritmo del cociente entre variables ("fold change") contra el p -valor para la comparación entre clases de cada variable. Cada punto del gráfico representa una variable y serán relevantes las que se encuentren arriba y en ambos extremos del gráfico. Es decir, aquellas con "fold change" elevado y con mayor p -valor (**Figura 22**).

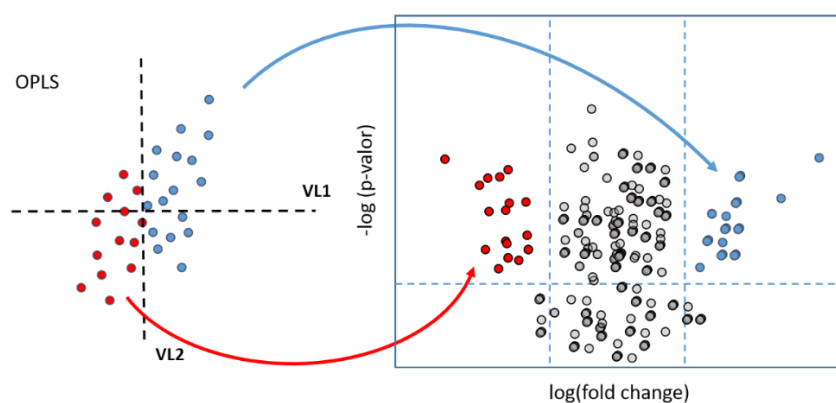


Figura 22. Gráfico de OPLS con su correspondiente gráfico volcán de la VL₁.

1.5.3 Gráfico S ("S-Plot")

Para construir un S-plot debemos graficar el vector de valores de carga contra el vector de coeficientes de correlación para cada variable. De este modo las variables que son de mayor relevancia se encuentran arriba y abajo en ambos extremos del gráfico. Estas presentan valores de carga y correlación con Y altos (**Figura 23**).

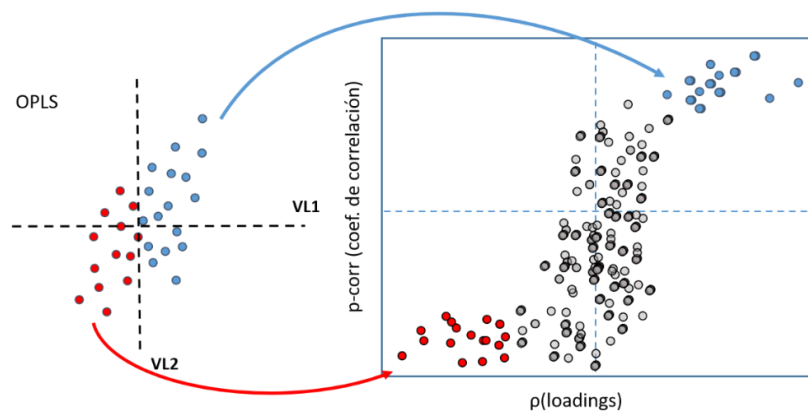


Figura 23. Gráfico de OPLS con su correspondiente S-plot de la VL₁.

1.6 Validación de modelos supervisados

El número óptimo de variables latentes y el desempeño predictivo de los modelos debe ser validado para minimizar el sobreajuste y disminuir el número de falsos positivos y/o negativos, evitando así clasificaciones incorrectas. Para esto es necesario estimar diferentes parámetros estadísticos y realizar diversas pruebas de validación de los modelos.

1.6.1 Ajuste y Predictibilidad (R^2 y Q^2)

Como se mencionó anteriormente (ver 1.4.2), los valores de R^2 y Q^2 nos permiten no solo estimar el número de componentes óptimo de nuestro modelo, sino también inferir que tan bien ajustan a nuestros datos y que tan robusto es el modelo a la hora de predecir sobre un nuevo conjunto de datos. En la práctica es deseable tener un valor de R^2 superior a 0.8 y que no difiera en más de 0.3 con el Q^2 [42,43].

1.6.2 Prueba de Permutaciones

Esta prueba tiene por objetivo evaluar la performance de nuestro modelo original frente a otros modelos generados a partir de los datos originales, pero permutados. Para esto se permuta la matriz Y en forma aleatoria un número de veces determinado. Con los datos así permutados se generan nuevos modelos para los cuales se calculan el Q^2 , R^2 , y la correlación que existe con el modelo original [44]. Para poder afirmar que nuestro modelo es robusto ninguno de los valores de Q^2 y R^2 de los modelos permutados pueden superar los valores del modelo original. A su vez, la correlación entre el modelo original y los permutados debe ser baja (**Figura 24**). Es importante tener en cuenta que los resultados de la prueba son afectados por el número de componentes, la estructura de los datos (variabilidad intra- e inter-clases), número de muestras y número de iteraciones [45].

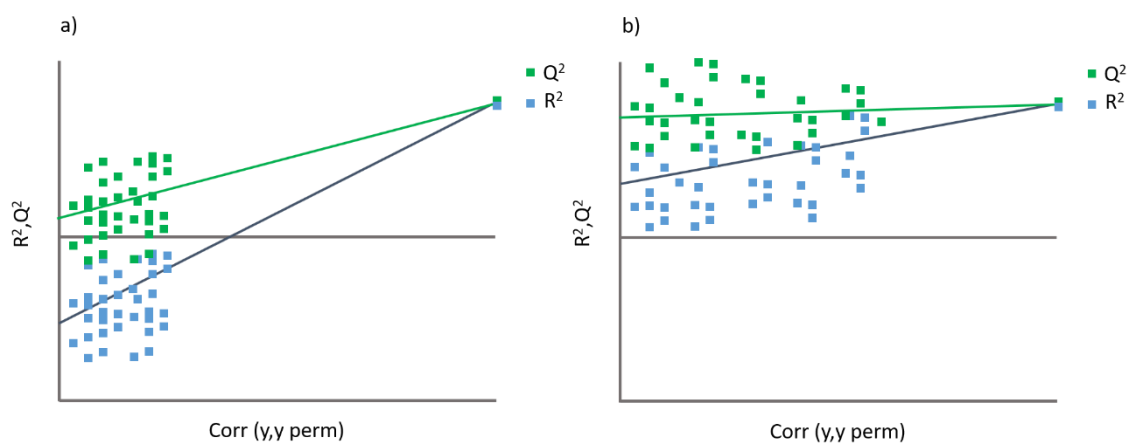


Figura 24. Gráfico prueba de permutaciones esperado **(a)**, gráfico prueba de permutación con sobreajuste o no deseado **(b)**.

1.6.3 Curvas de características operativas del receptor (ROC)

Para lograr comprender las curvas de ROC nos basaremos en el siguiente ejemplo gráfico. Para esto tomaremos la VL_1 de un modelo de OPLS-DA donde están las proyecciones de nuestras observaciones sobre esta, y dado que queremos clasificar entre dos clases podemos asumir estas como clases 0 y 1. Luego construiremos un gráfico de clases vs VL_1 y ajustaremos una curva sigmoidea a la que llamaremos curva de probabilidad. Para evaluar nuestro modelo definiremos un punto de corte en nuestra curva de probabilidad sobre el cual clasificaremos nuestras observaciones como clase 0 o clase 1. Este punto de corte lo conoceremos como clasificador. Como se observa en la **Figura 25** tendremos observaciones bien y mal clasificadas con las cuales construiremos una matriz de confusión (**Tabla 2**). Al mover la línea del clasificador se pueden construir diferentes matrices de confusión para cada clasificador.

Tabla 2. Matriz de confusión.

Valores de predicción	Valores reales	
	Verdaderos Positivos (VP)	Falsos Positivos (FP)
	Falsos Negativos (FN)	Verdaderos negativos (VN)

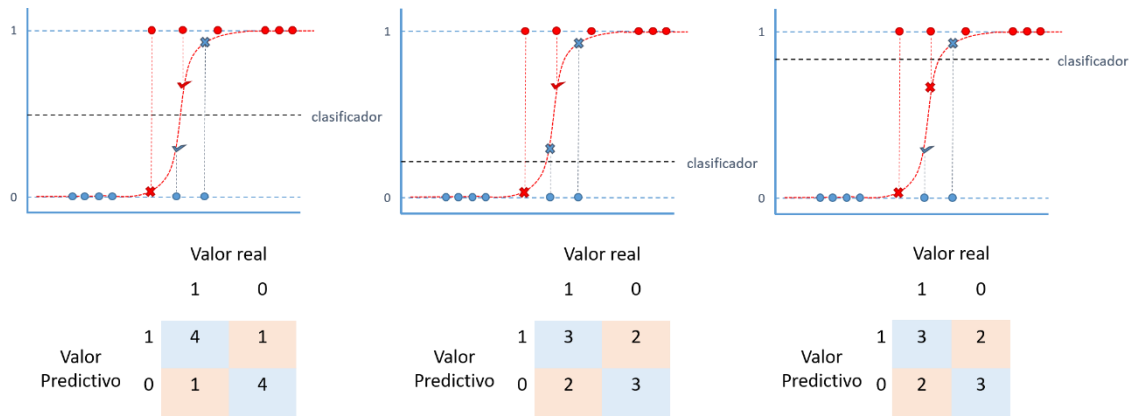


Figura 25. Curvas de probabilidad vs VL1 con tres puntos de corte para el clasificador. Para cada clasificador se construyen las tablas de confusión.

Luego debemos calcular la selectividad (Ec. 10) y especificidad (Ec. 11), y variando el límite del clasificador construimos las curvas de ROC. El área debajo de la curva (AUC) nos indica que tan bien clasifica nuestro modelo, nos permite determinar el mejor límite del clasificador, y comparar entre modelos (Figura 26) [46].

$$Selectividad = \frac{VP}{VP + FN} \text{ (Ec. 10)}$$

$$Especificidad = \frac{VN}{FP + VN} \text{ (Ec. 11)}$$

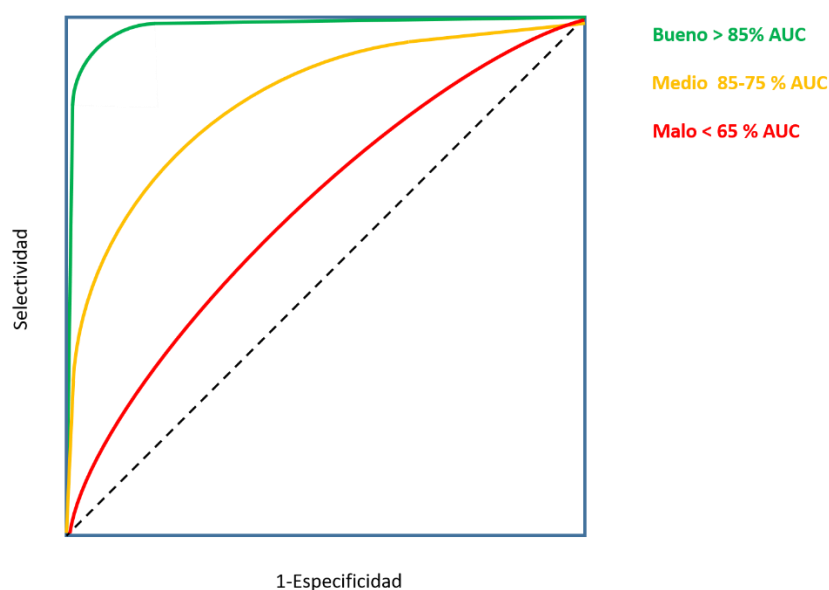


Figura 26. Gráfico de ROC para tres modelos de clasificación con tres niveles de performance diferentes.

1.7 Identificación de metabolitos y búsqueda en bases de datos

Una vez que encontramos las variables que son importantes para explicar nuestra matriz respuesta Y (clases o grupos) debemos identificar a que compuestos químicos corresponden. Esto no es una tarea sencilla dado que, como mencionamos anteriormente, trabajamos con mezclas complejas de compuestos que pueden abarcar una gran diversidad química. Primeramente, debemos recopilar toda la información estructural que podamos. Para esto nos valemos de experimentos como ser el HSQC, el HOMO2D-J, el HMBC, y el 1D-TOCSY, e información de otros núcleos (i.e., ^{13}C). También se pueden empelar herramientas estadísticas como el STOCYSY, el cual se describe a continuación.

1.7.1 Estadística de correlación espectroscópica total (STOCSY)

El STOCSY se basa en el análisis de la matriz de correlación (Ec. 12):

$$C = \frac{1}{n-1}XX' \text{ (Ec. 12)}$$

En esta ecuación, X , dimensiones $n \times k$, con n y k como las observaciones y variables, respectivamente, representa nuestra matriz de datos. X' corresponde a la matriz X transpuesta. Por último, cada elemento de la matriz C , de dimensiones $k \times k$, corresponderá al coeficiente de correlación entre las variables de X con X' .

Como sabemos, un metabolito puede presentar más de una señal en el espectro, y la relación de proporcionalidad entre sus intensidades siempre va a ser la misma. Si además se mantienen inalteradas las condiciones y parámetros experimentales al momento de adquirir un espectro, esto será válido para todas las muestras. Al aumentar o disminuir de concentración un metabolito en un grupo de muestras, también lo hacen en igual proporción sus señales de resonancia. Para construir los gráficos de STOCSY se considera primero una señal directriz, la cual es una de las variables relevantes que determinamos previamente al analizar el gráfico de valores de cargas del OPLS-DA. Para esa señal vamos a tomar la traza correspondiente de la matriz de covarianza, la vamos a graficar en función del corrimiento químico, y colorear por la misma traza de la matriz de autocorrelación (**Tabla 3**). Este gráfico es similar a un espectro de nuestras muestras con las señales coloreadas según qué tan alta sea

su correlación con la señal directriz (**Figura 27**). Esta correlación puede ser estructural, es decir señales que corresponden a una misma molécula, o biológica, cuando existe una alta correlación entre metabolitos de una misma ruta biológica [47].

Tabla 3 Formulas para calcular cada traza de covarianza (cov) y correlación (corr) de nuestra matriz X.

$cov = \sum \frac{(x_i - \bar{x})^2}{n - 1}$	x_i cada elemento del vector traza seleccionado, $\bar{x}$ la media del vector y n el número de elementos
$corr = \frac{s_{xy}}{s_x s_y}$	s_{xy} la desviación estándar de x con y, s_x la desviación estándar de x, s_y la desviación estándar de y

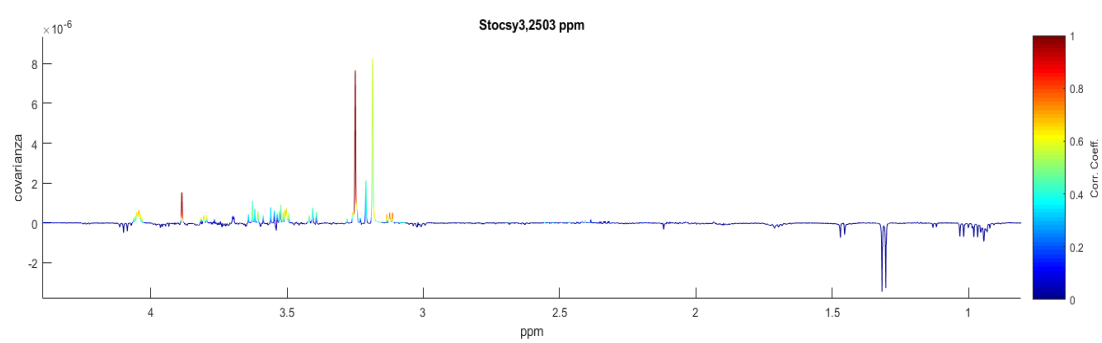


Figura 27. Ejemplo de gráfico de STOCSY para la betaína en una muestra de suero bovino. Este se construyó a partir de las trazas de correlación y covarianza para el pico directriz de la betaína (3.2503 ppm) en una matriz X de muestras de suero.

1.7.2 Bases de datos

Con toda la información estructural de nuestros metabolitos discriminantes realizamos búsquedas en bases de datos y bibliografía a fin de poder asignar las señales. A continuación, se listan algunas de las bases de datos libres disponibles con una breve descripción de estas (**Tabla 4**) [48,49].

Tabla 4. Bases de datos libres. Adaptada de Johnson, S. & Lange, B.M. (2015).

BML-NMR http://www.bml-nmr.org/	Se enfoca principalmente en el metabolóma humano. Con más de 208 compuestos y 3328 espectros de referencia.
BMRB https://bmr.io/	Metabolismo principalmente de plantas. Con 1249 compuestos y más de 8500 espectros
HMDB https://hmdb.ca/	Enfocado al metabolóma humano. Con 41800 compuestos y más de 2240 espectros de RMN.
NAPROC-13 https://c13.materia-medica.net/	Productos naturales, terpenoides, esteroides, alcaloides, flavonoides, ligninanos, otros. Con más de 20000 compuestos y espectros de RMN
NMR ShiftDB https://nmrshiftdb.nmr.uni-koeln.de/	Enfocan en productos naturales. Posee en su base de datos más de 45000 compuestos y espectros de RMN.

MMCD

Con más de 20000 compuestos y más de 5000

<http://mmcd.nmrfa.m.wisc.edu> espectros de referencia

1.7.3 Niveles en la identificación de metabolitos

Luego de obtener toda la información estructural de nuestros metabolitos debemos identificarlos. Para esto podemos definir cuatro niveles de confiabilidad (Tabla 5) [48].

Tabla 5. Niveles de identificación de metabolitos

Nivel 1	Además de identificación contra un estándar, se debe identificar al compuesto mediante técnicas o experimentos ortogonales, i.e. ^1H RMN y experimentos bidimensionales como HSQC.
Nivel 2	Anotación putativa del compuesto, identificado por análisis de datos y búsqueda en bases de datos
Nivel 3	No identificado, pero la información permite caracterizarlo y asignarle una clase
Nivel 4	No identificado y no clasificado. Solo caracterizado espectroscópicamente

Capítulo II

Seguimiento del estado sanitario e identificación de biomarcadores de salud de ganado lechero.

2. 1. Introducción

Durante el período de transición desde la conclusión de la gestación hasta el final de la lactancia, las vacas lecheras sufren cambios drásticos en su metabolismo. Esto es debido, principalmente, a la disminución de la ingesta de materia seca (“dry matter intake”, DMI), que ocurre incluso antes del parto, y al aumento de las demandas energéticas asociadas con la producción de leche (**Figura 28**) [50]. Estos cambios pueden conducir a lo que se conoce como un estado de balance energético negativo (“negative energy balance”, NEB), el cual desencadena la activación del catabolismo de lípidos y proteínas [50,51].

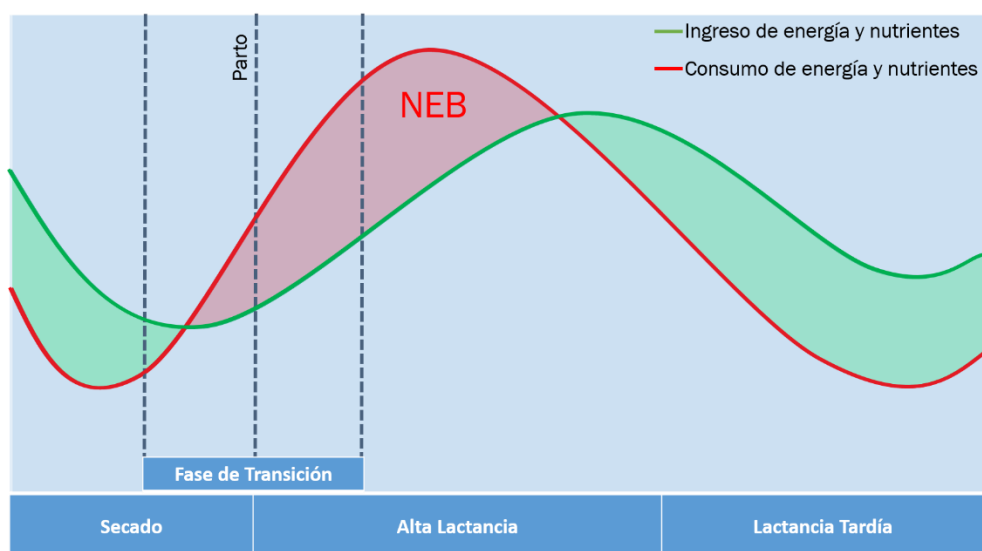


Figura 28. Esquema del balance de energía y nutrientes según etapas en el periparto y lactancia.

Si el NEB es grave o prolongado puede afectar la producción de leche, y, lo que es más importante, puede tener efectos en los sistemas inmunitario y endocrino del animal perjudicando la salud y el rendimiento reproductivo [52-55]. De hecho, el período de transición se caracteriza por un aumento en la incidencia de enfermedades productivas como la mastitis, metritis, cetosis, y lipidosis hepática [53,54,56,57].

El proceso de selección y mejoramiento genético continuo que se ha empleado en las últimas décadas para mejorar los rendimientos de producción de leche ha exacerbado estos problemas asociados con el estrés metabólico [58,59]. Todo esto se traduce en importantes pérdidas económicas, y por ende se han dedicado esfuerzos considerables para comprender mejor los cambios en el metabolismo de las vacas lecheras en el periparto. Varios estudios se han centrado en monitorear las variaciones en los niveles de metabolitos asociados con el NEB, particularmente la de los ácidos grasos no esterificados (“Non-Esterified Fatty Acids”, NEFA) en sangre [60,61], que generalmente son seguidos por aumentos en la concentración de β -hidroxibutirato (BHB) y acetoacetato [62]. Los cambios en la concentración de glucosa, insulina y otras hormonas, y las proteínas de la sangre también se correlacionan con los desafíos metabólicos experimentados por los animales alrededor del parto [54,60,61]. Sin embargo, la problemática es más compleja, dado que las enfermedades de producción están interrelacionadas etiológicamente y no se pueden considerar de forma aislada [57].

Por estos motivos, los enfoques analíticos integrados en los que se monitoreen simultáneamente grupos de metabolitos resultan la mejor opción para la evaluación y seguimiento del estado sanitario de un animal o rodeo. El análisis metabolómico ofrece esta posibilidad y, en consecuencia, se ha empleado con éxito para investigar la salud del ganado lechero [64]. Por ejemplo, la metabolómica dirigida basada en MS se ha utilizado para investigar mastitis, metritis, cetosis y lipidosis hepática en vacas periparturientas, y también se aplicaron enfoques no dirigidos basados en RMN para el estudio de varias de estas condiciones [65-73]. También se han aplicado métodos dirigidos y no dirigidos basados en MS para seguir los niveles de distintos metabolitos en vacas durante el periodo de transición con el fin de identificar biomarcadores que podrían evidenciar etapas tempranas de enfermedades [74-78]. La mayoría de estos estudios se realizaron en sistemas de producción lechera confinados. Sin embargo, maximizar la proporción de pastoreo a cielo abierto en la dieta es un factor fundamental para minimizar los costos de producción [79]. Es bien sabido que el DMI en vacas lecheras en pastoreo es insuficiente para sostener la alta producción de leche que podría lograrse por su potencial genético, por lo que se administran suplementos y concentrados para alcanzar los niveles de producción deseados y minimizar la magnitud y la duración del NEB [80]. En contraste con los esquemas de dieta de raciones totalmente mezcladas (“total mixed ration”, TMR), la predicción de la disponibilidad de nutrientes en sistemas mixtos donde el forraje es el componente principal de la dieta es compleja, porque incluye incertidumbres asociadas con el comportamiento del animal en pastoreo. En este

escenario, la disponibilidad de propionato y otros metabolitos glucogénicos para suplir las demandas hepáticas puede verse limitada, afectando el suministro de glucosa para la glándula mamaria en vacas lecheras de alta producción. Por lo tanto, la caracterización de precursores y vías metabólicas alternativas involucradas en la síntesis de glucosa *de novo* y la identificación de adaptaciones del tejido hepático y periférico a la lactancia podrían contribuir a la optimización de las estrategias de manejo nutricional en un rodeo. Además, este conocimiento podría ser fundamental en el diseño de programas de medicina preventiva destinados a disminuir la incidencia de enfermedades de producción y los costos asociados a estas. Con esto en mente, investigamos la idoneidad de la metabolómica no dirigida basada en RMN como una herramienta para seguir los cambios en el metaboloma sérico de un grupo de vacas lecheras sin condiciones clínicas iniciales y acceso a pastos después del parto.

2.2. Materiales y Métodos

2.2.1. Recolección de muestras y animales

El protocolo experimental fue aprobado por la Comisión Honoraria de Experimentación Animal de la UdelaR (Exp. 111130-002261-13). El estudio fue realizado sobre un grupo de 26 vacas Holstein, Jersey, cruce Holstein × Jersey manejadas en conjunto en el tambo de Facultad de Veterinaria, UdelaR (San José, Uruguay). Incluyó 17 vacas primíparas y 9 vacas multíparas que tenían entre tres y siete partos. Su dieta consistía en TMR a base de concentrado y ensilado antes del parto, a lo que se le agregó pastoreo como una opción después del parto

(**Tabla S1**). Se llevó a cabo un seguimiento por veterinarios capacitados desde ~30 días antes del parto hasta ~60 días después del parto. Animales con manifestaciones clínicas de hipocalcemia, placenta retenida, metritis moderada o mastitis severa (es decir, grados 2 o 3) fueron excluidos del estudio [81-84], mientras que los que desarrollaron mastitis clínica leve (es decir, grado 1) o cetosis subclínica que no mostró la afectación sistémica no se excluyeron [84,85]. El sangrado se realizó cada 15 días previo al parto, y una vez por semana inmediatamente después del ordeño y antes del pastoreo después del parto. Se extrajeron muestras de sangre de la vena coccígea en tubos sin anticoagulante, se centrifugaron a 3000 g durante 10 min, y el suero resultante se almacenó a -20 °C hasta su análisis.

2.2.2. Bioquímica Clínica

Los análisis bioquímicos se realizaron en el Laboratorio de Endocrinología y Metabolismo Animal (LEMA) de Facultad de Veterinaria. Se determinaron colesterol, proteína total, urea, y calcio empleando kits colorimétricos comerciales (Weiner Laboratorios S.A.I.C., Rosario, Argentina) por medio de un analizador bioquímico automático Vitalab Selectra 2. Se estimó la globulina por la diferencia entre la proteína total y la albumina. La concentración de NEFA se determinó empleando los métodos de la Acil-CoA sintasa - acil-CoA oxidasa (ACS ACOD, NEFA-C kit, Wako Chemicals USA Inc., Richmond, VA, EUA). Para determinar las concentraciones de BHB se emplearon métodos enzimáticos (Ranbut kit, Randox Laboratories Ltd., Crumlin, R.U.).

2.2.3. Espectroscopía de RMN

Las muestras de suero se dejaron equilibrar a temperatura ambiente. Luego se alicuotaron 400 μL a los que se les agregó 200 μL de una solución 0,9% NaCl en D_2O , y el resultante fue transferido a un tubo de RMN de 5 mm (NE HL5 7, New Era Enterprises Inc., Vineland, NJ, EUA). Todos los espectros se adquirieron a 25 °C en un equipo Bruker AVANCE III 500 NMR operando a frecuencias de ^1H y ^{13}C de 500,13 y 125,76 MHz, respectivamente, equipado con una sonda TXI con gradiente-z. Para eliminar las señales de proteínas se empleó la secuencia de pulsos de Carr Purcell Meiboom-Gill con supresión de agua, un τ de 0,4 ms, y un η de 80 ms [86,87]. Se empleó un ancho espectral de 10 kHz, un tamaño de datos de 32K, y 128 barridos con un tiempo de reciclado de 4 s entre barridos. Cuando fue necesario se realizaron otros experimentos de dereplicación química como HSQC y 1D-TOCSY empleando las secuencias de pulso estándar del espectrómetro.

2.2.4. Procesado de datos de RMN

Todas las FID fueron llenadas con ceros a 64K y apodizadas con una función de ventana exponencial de 0,3 Hz previo a la transformada de Fourier. La fase y línea de base se corrigieron manualmente empleando el software MNova (version 10.0, MestreLab Research, S.L., Santiago de Compostela, España). Todos los espectros fueron referenciados a la señal del lactato a 1,25 ppm. Se seleccionaron manualmente las regiones del espectro para su alineación por medio del algoritmo icoshift [27], y luego se normalizó al total del área previa

exclusión de la señal residual del agua (4,06-5,00 ppm). La matriz de datos resultante fue exportada como archivo de texto para su uso en los análisis estadísticos.

2.2.5. Determinación de metabolitos

Las concentraciones absolutas de los metabolitos previamente identificados se calcularon por integración de sus señales en el espectro ^1H RMN tomando como referencia la concentración determinada para el BHB por métodos enzimáticos.

2.2.6. Análisis Estadísticos

Los análisis de varianza de una vía (ANOVA) se realizaron empleando el software GraphPad Prism (versión 7.0, GraphPad Software Inc., San Diego, CA, EUA). Los MVDA, incluyendo análisis de PCA y OPLS-DA, se llevaron a cabo empleando el software PLS_Toolbox (versión 8.5, Eigenvector Research Inc., Manson, WA, EUA) implementado para MATLAB (revisión 2014a, The MathWorks Inc., Natick, MA, EUA). Para todos los modelos multivariados los datos fueron centrados en la media y escalados empleando el factor de Pareto [24]. Primeramente, se realizaron análisis de PCA en busca de la identificación de agrupamientos o tendencias en los datos e identificar la presencia de observaciones que estén fuertemente fuera del modelo (“outliers”). Los análisis de OPLS-DA se realizaron sobre la matriz completa de datos y sobre la región aromática (6,0 a 9,0 ppm). Los modelos fueron validados empleando el método de subgrupos aleatorios, o test de Montecarlo, que toma iterativamente particiones de un grupo de datos de los datos originales [88,89]. También se

realizaron pruebas de permutaciones con 200 iteraciones para determinar el grado de sobreajuste y la robustez de los modelos [44,90].

2.3. Resultados y discusión

2.3.1. Espectros de ^1H RMN de muestras de suero

En la **Figura 29** y **Tabla 5** se muestra un espectro representativo de ^1H RMN de una muestra de suero utilizada en el estudio y las anotaciones correspondientes para los metabolitos identificados. Además de las señales correspondientes a lipoproteínas, azúcares, y aminoácidos, se pueden asignar putativamente por comparación con datos de literatura las señales de metabolitos como acetato (1,91 ppm), acetoacetato (2,22 ppm), alantoína (5,37 ppm), betaína (3,25 y 3,90 ppm), BHB (1,19, 2,29, 2,40 y 4,13 ppm), colina (3,15, 3,43 y 3,97 ppm), citrato (2,52 y 2,68 ppm), citrulina (1,50, 1,80, 3,07 y 3,70 ppm), creatina (2,31 y 3,93 ppm), creatinina (3,04 y 4,05 ppm), formiato (8,43 ppm), glicerolfosfocolina (3,22, 3,41, 3,67, 3,90 y 4,30 ppm), isobutirato (1,11 y 2,62 ppm), lactato (1,33 y 4,10 ppm), fosfocolina (3,20, 3,65 y 4,17 ppm), trimetilamina (TMA) (2,91 ppm), trimetilamino *N*-óxido (TMAO) (3,34 ppm), y tirosina (6,92 y 7,12 ppm) [71,86,91]. Estas asignaciones fueron, en algunos casos, corroboradas con la ayuda de espectros HSQC y 1D-TOCSY. Como se discutió anteriormente, varias de estas especies están relacionadas con el estrés metabólico asociado con el NEB en el período de transición de vacas lactantes.

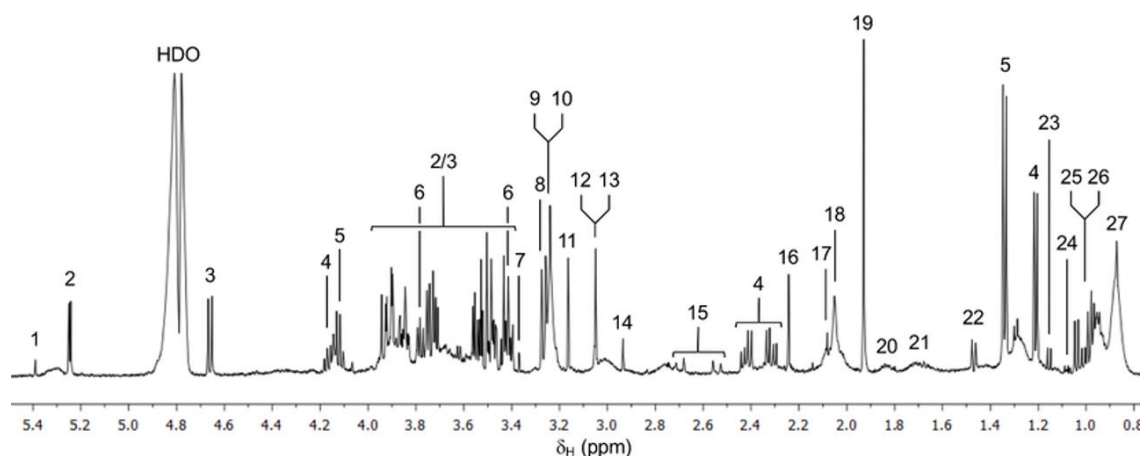


Figura 29. Espectro de ^1H RMN representativo de una muestra de suero de vaca lechera. Las señales corresponden a la alantoína (1), α -glucosa (2), β -glucosa (3), BHB (4), lactato (5), mioinositol (6), TMAO (7), betaína (8), fosfocolina (9), glicerolfosfocolina (10), colina (11), creatinina (12), creatina (13), TMA (14), citrato (15), acetoacetato (16), O-acetilglicoproteína (17), N-acetilglicoproteína (18), acetato (19), citrulina (20), lisina (21), alanina (22), valina (23), isobutirato (24), isoleucina (25), leucina (26) y lipoproteína L1 (27).

Anotación	Metabolito	Corrimiento Químico (ppm)	Multiplicidad	Group
1	Alantoína	5,36	s	CH
2	α -Glucosa	5,22	d	H1
3	β -Glucosa	4,63	d	H1
4	Betahidroxibutirato (BHB)	2,34/4,13	m/m	$\alpha\text{CH}_2/\beta\text{CH}_2$
5	Lactato	1,33/4,1	d/q	CH_3/CH
6	Mioinositol	3,84	dd	CH
7	Trimetilamino N-oxido (TMAO)	3,34	s	CH_3
8	Betaina	3,25	s	$\text{N}(\text{CH}_3)_3$
9	Fosfocolina	3,2	s	$\text{N}(\text{CH}_3)_3$
10	Glicerofosfocolina	3,2	s	$\text{N}(\text{CH}_3)_3$
11	Colina	3,18	s	$\text{N}(\text{CH}_3)_3$
12	Creatinina	3,02	s	CH_3
13	Creatina	3,02	s	CH_3
14	Trimetilamina (TMA)	2,91	s	$\text{N}(\text{CH}_3)_3$
15	Citrato	5,59	dd	CH_2
16	Acetoacetato	2,22	s	CH_3
17	O-acetilglicoproteína	2,01	s	NHCOCH_3

18	N-acetilglicoproteína	2,02	s	NHCNCH ₃
19	Acetato	1,9	s	CH ₃
20	Citrulina	1,83	m	βCH ₂
21	Lisina	1,68	m	dCH ₂
22	Alanina	1,45	d	CH ₃
23	Valina	1,02/0,96	d/d	CH ₃ /CH ₃
24	Isobutirato	1,18	d	CH ₃
25	Isoleucina	0,98/0,93	d/t	βCH ₃ /dCH ₃
26	Leucina	0,91	dd	dCH ₃
27	Lipoproteínas	0,84	m(b)	CH ₂

Tabla 5. Anotaciones para los metabolitos identificados en suero bovino en un espectro de ¹H RMN. Singulete (s), doblete (d), triplete (t), multiplete (m), señal ancha (broad, b)

2.3.2. Metabolómica sérica por RMN

El paso inicial en el estudio involucró la identificación de grupos entre los animales a través del periparto. Para ello, se llevó a cabo un PCA considerando los datos derivados de los espectros de ¹H RMN para todas las muestras de suero. El gráfico resultante no mostró diferenciación según raza o paridad (**Figura 30a** y **b**, respectivamente). Se obtuvieron resultados similares cuando se consideraron las componentes principales que capturan menos varianza. Sin embargo, cuando las observaciones se colorearon por un vector continuo según los días en torno al parto, se identificaron cuatro grupos. Estos involucraron muestras de animales cuatro semanas antes del parto (C - 4), y una, cuatro y ocho semanas después del parto (C + 1, C + 4 y C + 8, respectivamente), coloreadas en azul, celeste, amarillo, y rojo en la **Figura 30c**, respectivamente. Los datos de bioquímica sanguínea e información sobre los animales involucrados en cada una de estas etapas se resumen en la **Tabla 6**.

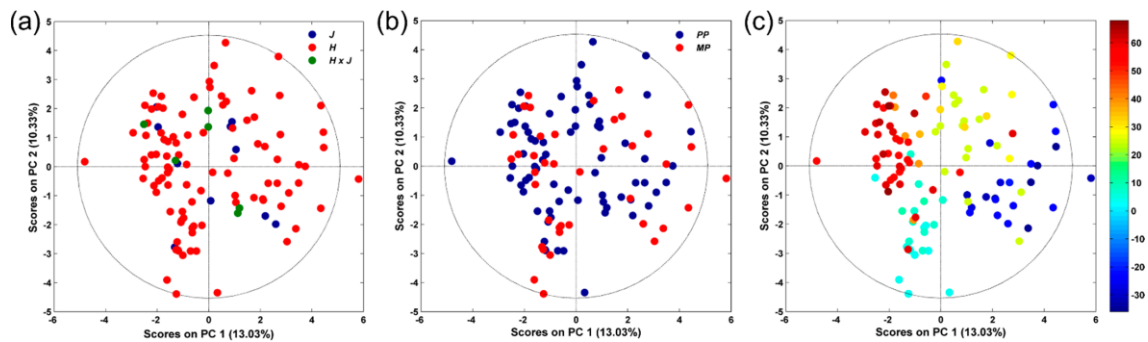


Figura 30. Gráfico de PCA para todas las muestras de suero. Los gráficos se colorearon según la raza (a), paridad (b), o utilizando un vector de color continuo que varía en función del día del período de transición en relación con el parto (c).

	Estadio			
Datos de los animales	C-4	C+1	C+4	C+8
Días por etapa	-36 to -22	3 to 11	23 to 34	39 to 68
N (H, J, H × J)	23 (19, 2, 2)	20 (18, 0, 2)	24 (20, 2, 2)	20 (17, 1, 2)
Paridad (p, m)	15, 8	12, 8	15, 9	13, 7
Enfermedad (ssc,mcl)	6, 1	7, 1	7, 2	6, 0
Bioquímica				
Albumina (g/L)	31.00 ± 2.19	29.64 ± 2.91	29.44 ± 5.67	31.42 ± 2.78
Urea (mM)	2.40 ± 1.31	2.82 ± 1.10	2.65 ± 1.11	3.43 ± 1.46
Colesterol (mM)	2.71 ± 0.43	2.29 ± 0.50	3.81 ± 0.75	5.09 ± 0.93
Proteína Total (g/L)	79.61 ± 7.00	74.23 ± 7.19	78.13 ± 7.72	80.52 ± 6.64
Globulina (g/L)	48.90 ± 6.50	44.59 ± 5.17	48.69 ± 5.12	49.10 ± 4.60
BHB (mM)	0.38 ± 0.21	0.56 ± 0.32	0.78 ± 0.63	0.65 ± 0.35
NEFA (mM)	0.29 ± 0.18	0.63 ± 0.22	0.68 ± 0.36	0.47 ± 0.20

Tabla 6. Datos de bioquímica e información sobre los animales dentro de cada etapa (a) H, J y H × J representan las razas Holstein, Jersey y cruce Holstein × Jersey, respectivamente. (b) PP y MP indican primíparas y multíparas, respectivamente. (c) ssc y mcl indican cetosis subclínica y mastitis clínica leve (grado 1), respectivamente.

Habiendo identificado cuatro etapas en el periparto se investigaron las variaciones en el metabolóma sérico tomando como referencia los niveles de metabolitos en preparto. Esto se logró a través una comparación por pares de C - 4 contra los grupos C + 1, C + 4 y C + 8, respectivamente, empleando OPLS-DA y ANOVA (**Figura 31, Figura 32, Tabla 7**). En todos los casos, el pico positivo a 1,25 ppm observado en los gráficos de valores de carga del OPLS-DA corresponde a una fuerte caída en la concentración de lactato después del parto. Si bien hay varias fuentes de este metabolito, incluida la deshidrogenación del piruvato a través del sistema de la glioxalasa en el metabolismo de grasas, carbohidratos, y proteínas [92], en este caso la fermentación bacteriana anaeróbica en rumen y colon probablemente conduce a los altos niveles de lactato observados antes del parto [93,94]. Como se concluyó recientemente en un estudio a gran escala del metabolóma sérico de ganado lechero clínicamente sano [78], el lactato puede estar asociado con un aumento de la ingesta de carbohidratos fermentables como los presentes en la dieta a base de TMR [95,96]. Al mismo tiempo, la marcada disminución de los niveles de lactato después del parto sugiere que este metabolito se emplea para apoyar las demandas de glucosa a través de la gluconeogénesis por activación del ciclo de Cori [97,98]. Esto, junto con un aumento en el metabolismo de los lípidos a través de la oxidación y la activación del ciclo de la alanina en el punto álgido de la producción de leche, también es consistente con el aumento de dos a tres veces observado en los niveles de glucosa en sangre después del parto (**Tabla 7**). Estos cambios podrían ayudar a explicar por qué ninguno de los animales tuvo manifestaciones de

cetosis durante el estudio. Una semana después del parto se observaron valores negativos en el gráfico de valores de carga del OPLS-DA para acetato, betaína, y creatina (**Figura 31a**), lo que indica un aumento en los niveles de estos marcadores de estrés metabólico [70,76]. En el análisis bioquímico la tendencia al aumento de las concentraciones de BHB y NEFA también está en acuerdo con este hallazgo (**Tabla 6**). En general, estos datos concuerdan con el aumento en la movilización de lípidos que tiene lugar al principio de la lactancia [64,99]. Como se ve en la **Tabla 7**, también hay aumentos en los niveles de asparagina, citrato, citrulina, glucosa, y lisina. Esto indica alteraciones en los ciclos del ácido cítrico y de la urea, y en las vías glicolíticas y gluconeogénicas inmediatamente después del parto [63,76], y confirma que los animales entran en una condición de NEB. El aumento de los niveles de alantoína circulante (**Tabla 7**), un metabolito ligado al catabolismo de las purinas, es consistente con mayores rendimientos microbianos ruminales asociado con el aumento de DMI, así como la inclusión de forraje en la dieta después parto [100].

Además de acetato y BHB, el gráfico de valores de carga para el OPLS-DA que contrasta los grupos C - 4 y C + 4 muestra que los niveles de otros marcadores de estrés metabólico, como acetoacetato, colina, y formiato, aumentan significativamente cuatro semanas después del parto (**Figura 30b** y **Figura 31b**). De manera similar, el análisis de datos univariado indica aumentos de dos a cuatro veces para las especies identificadas en la etapa C + 1, incluidos el acetato, el BHB, la betaína, y la glucosa, así como la creatina y la creatinina (**Tabla 7**). También se observa una mayor acumulación de alantoína, que en esta etapa

puede asociarse no solo con una mayor producción de proteínas bacterianas en el rumen causado por el aumento de DMI, sino también con un aumento en el daño tisular [101]. De hecho, la acumulación de alanina apunta a una activación del ciclo de alanina, en el cual la proteína muscular se cataboliza para alimentar la gluconeogénesis hepática [51,102]. Los niveles más altos de acetato y citrato en sangre también revelan un mayor aumento en la lipomovilización a través de la lipólisis y la oxidación. Esto, a su vez, activa la vía cetogénica, lo que conduce al marcado aumento observado en las concentraciones de BHB, lisina y acetoacetato [76,101]. No es sorprendente que estos cambios pronunciados ocurran cuando los animales alcanzan el pico de producción de leche (**Figura 30 y Tabla 6**) y concordante con los grandes cambios en el metabolismo que se requieren para satisfacer las demandas de energía durante este período. De hecho, basándose exclusivamente en los niveles de BHB, casi un cuarto de los animales desarrolló cetosis subclínica en esta etapa [80].

Tabla 7. Concentraciones promedio de los metabolitos durante el peripato determinadas por ^1H NMR (media $\pm$ des.est), y cambios respecto a los niveles en el parto

Metabolito	Concentración (mM)						Fold change				
	C-4	C+1	C+4	C+8	C+1 / C-4	P	C+4 / C-4	P	C+8 / C-4	p	
Acetato	0.83 ± 0.80	1.36 ± 1.22	1.99 ± 2.17	1.86 ± 1.13	1.63	0.0494	2.40	0.0091	2.25	0.0002	
Acetoacetato	0.11 ± 0.15	0.12 ± 0.08	0.17 ± 0.16	0.12 ± 0.09	1.14	0.3519	1.62	0.0724	1.10	0.3734	
Alanina	0.18 ± 0.17	0.26 ± 0.22	0.43 ± 0.46	0.51 ± 0.39	1.50	0.0788	2.48	0.0071	2.88	0.0002	
Alantoina	0.07 ± 0.07	0.16 ± 0.16	0.24 ± 0.27	0.20 ± 0.12	2.16	0.0167	3.41	0.0021	2.81	< 0.0001	
Aspargina	0.61 ± 0.42	1.71 ± 2.20	2.36 ± 2.90	1.92 ± 1.27	2.82	0.0116	3.88	0.0074	3.16	< 0.0001	
Betaina	0.18 ± 0.15	0.34 ± 0.38	0.56 ± 0.63	0.43 ± 0.26	1.92	0.0329	3.15	0.0035	2.44	0.0001	
BHB	0.38 ± 0.20	0.56 ± 0.32	0.78 ± 0.62	0.65 ± 0.35	1.46	0.0175	2.04	0.0028	1.71	0.0009	
Citrato	0.10 ± 0.09	0.26 ± 0.25	0.39 ± 0.50	0.22 ± 0.18	2.59	0.0097	4.01	0.0085	2.29	0.0035	
Citrulina	0.13 ± 0.12	0.26 ± 0.31	0.48 ± 0.65	0.28 ± 0.23	2.01	0.0442	3.75	0.0069	2.18	0.0039	
Creatina	0.04 ± 0.04	0.10 ± 0.14	0.15 ± 0.22	0.10 ± 0.09	2.51	0.0201	3.70	0.0105	2.33	0.0029	
Creatinina	0.18 ± 0.11	0.28 ± 0.22	0.39 ± 0.58	0.31 ± 0.27	1.54	0.0325	2.13	0.0484	1.68	0.0229	
α-Glucosa	1.00 ± 0.88	2.27 ± 2.54	3.50 ± 4.15	2.91 ± 2.04	2.28	0.0145	3.51	0.0033	2.29	< 0.0001	
β-Glucosa	0.92 ± 0.81	2.19 ± 2.61	3.49 ± 4.74	2.81 ± 2.08	2.38	0.0415	3.77	0.0068	3.04	< 0.0001	
Lactato	3.60 ± 3.10	1.16 ± 0.63	1.67 ± 1.13	1.60 ± 0.77	-2.23	0.0038	-2.16	0.0015	-2.26	0.0006	
Leucin	0.12 ± 0.10	0.06 ± 0.03	0.06 ± 0.03	0.03 ± 0.03	-1.83	0.0166	-1.93	0.0036	-3.80	< 0.0001	
Lisina	0.08 ± 0.07	0.23 ± 0.31	0.37 ± 0.48	0.26 ± 0.20	2.63	0.0169	4.16	0.0037	2.97	0.0001	
Valina	0.23 ± 0.17	0.22 ± 0.07	0.24 ± 0.08	0.26 ± 0.09	-1.02	0.4583	1.04	0.3938	1.14	0.1901	

Los datos del metaboloma en C + 8 son de particular interés. Aunque las vacas todavía están en su pico de producción de leche, el DMI también aumenta en esta etapa. Como resultado, los animales tienden a normalizar su balance energético, lo que se evidencia por una disminución en los niveles de NEFA, **Tabla 6**. Esto también se refleja en el gráfico valores de carga del OPLS-DA presentado en la **Figura 30c**, que revela que las vacas en el parto y el grupo C + 8 se diferencian por metabolitos como acetato y glucosa, pero no por marcadores de estrés metabólico severo presentes en el grupo C + 4. De manera similar, el análisis univariado muestra que los metabolitos con grandes cambios para la etapa C + 4 ahora presentan menores variaciones en comparación con sus niveles antes del parto. Esto indica una tendencia hacia la normalización de las rutas metabólicas identificadas como alteradas en los análisis de los grupos C + 1 y C + 4, y es consistente con la salida de los animales de la condición de NEB y en camino a recuperar la homeostasis.

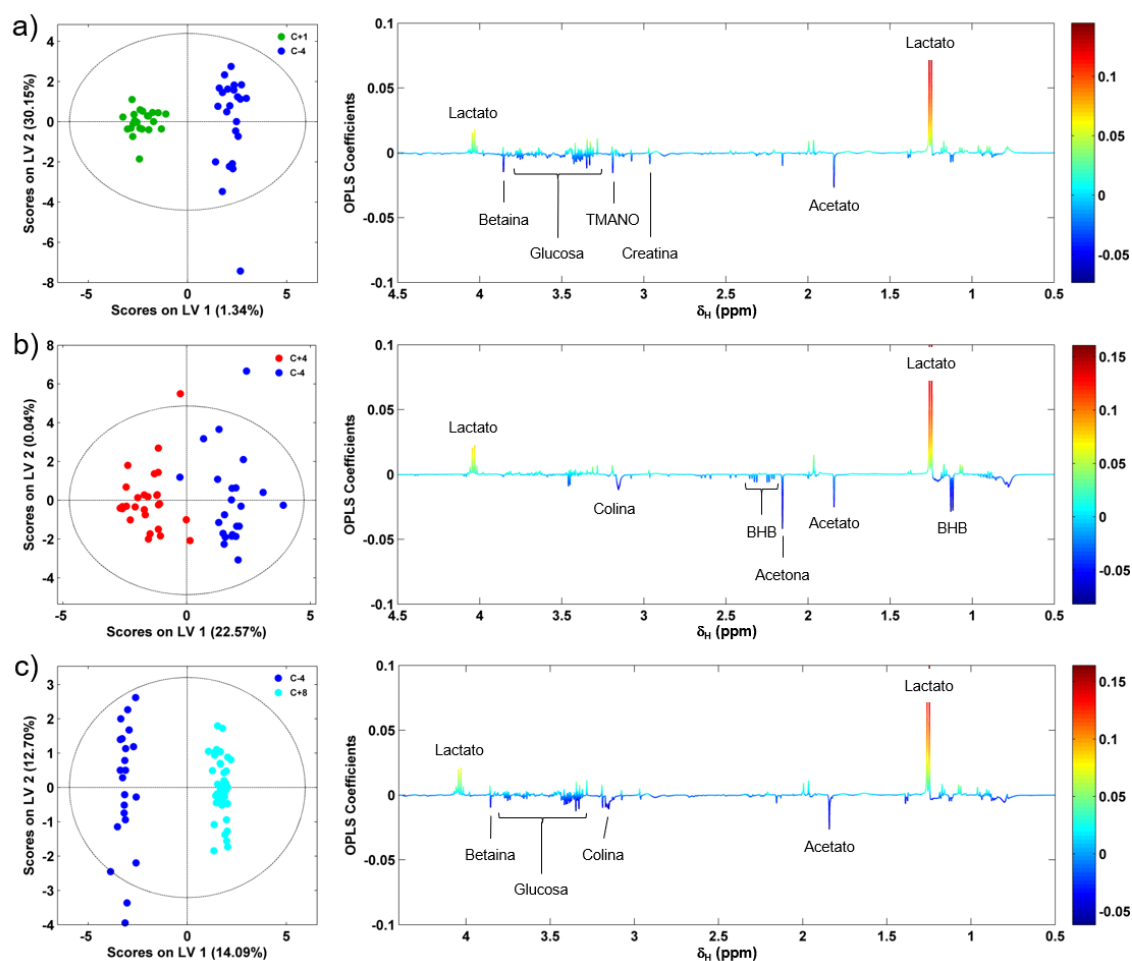


Figura 31. Gráficos de OPLS y sus respectivos gráficos de valores de carga obtenidos entre preparto (C- 4) y primero (C + 1) **(a)**, segundo (C + 4) **(b)**, y tercero (C + 8) **(c)** posparto. Los coeficientes R^2Y y Q^2Y para cada modelo fueron 0,80 y 0,72, 0,89 y 0,37 y 0,94 y 0,87, respectivamente, y sus curvas ROC obtuvieron valores de AUC de 1.00, 0,85 y 0,99. Las figuras de las curvas ROC y test de permutaciones se encuentran en el **Anexo1, Figura S1-S8**.

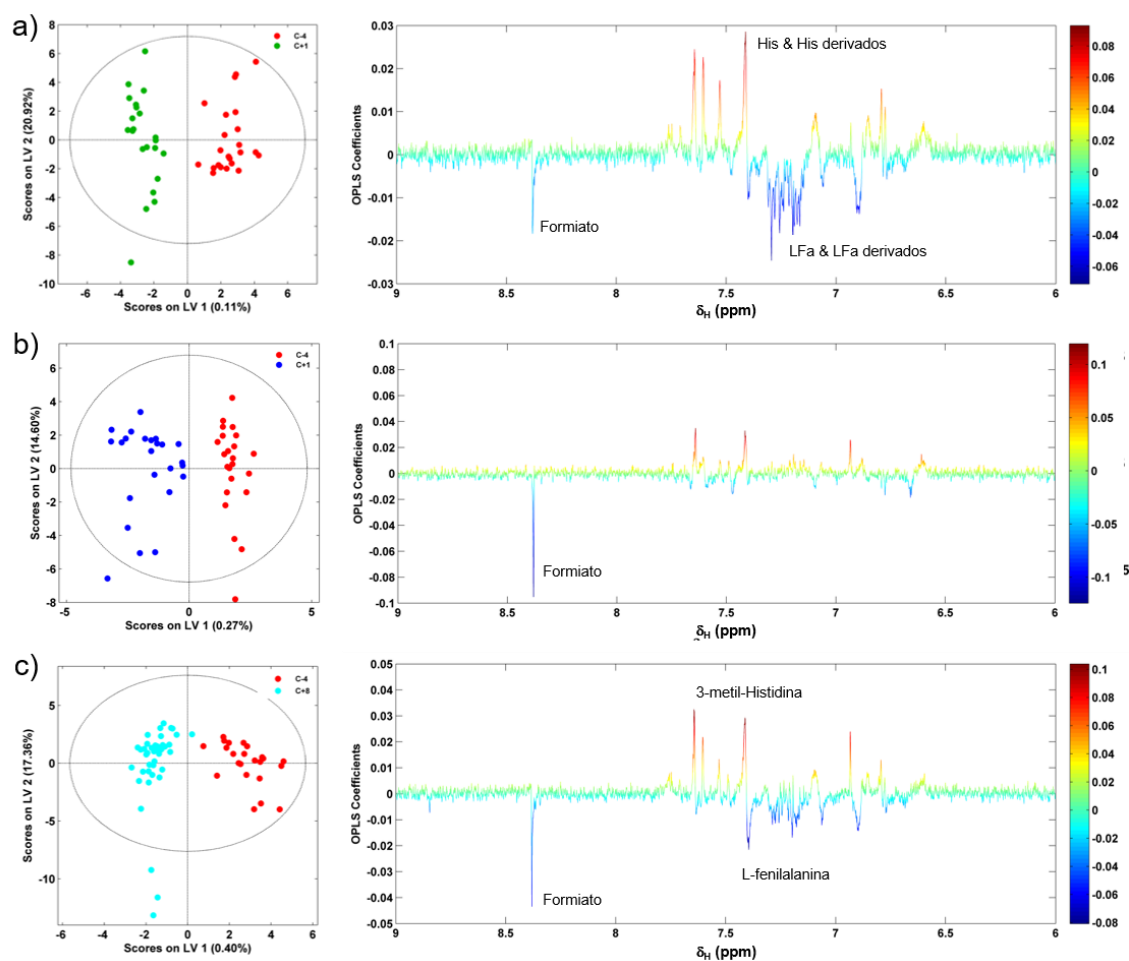


Figura 32. Gráficos de OPLS y sus respectivos gráficos de valores de carga obtenidos con datos de la región de 6,0 a 9,0 ppm. En estos se comparó el preparto (C-4) y el primero (C+1) **(a)**, el segundo (C+4) **(b)** y el tercero (C+8) **(c)** grupos posparto. Los coeficientes R^2Y y Q^2Y para cada modelo fueron 0,84 y 0,66, 0,66 y 0,20 y 0,78 y 0,66, respectivamente, y sus curvas ROC obtuvieron valores de AUC de 0,99, 0,77 y 0,99.

Como se indicó anteriormente, no se observaron diferencias basadas en la paridad de los animales cuando todos los datos se analizaron en conjuntos. Sin embargo, se investigaron las variaciones relacionadas con esta condición dentro de cada una de las cuatro etapas. Para esto se llevó a cabo un PCA para cada uno de los estadios definidos previamente. Si bien no se observaron efectos significativos después del parto, una diferencia entre vacas primíparas y

multíparas fue evidente en el grupo C - 4, **Figura 33a**. El OPLS-DA comparando paridad en el parto revela que las concentraciones de lactato, acetato, y BHB son mayores en vacas multíparas, mientras que la acumulación de creatina, TMAO, colina, glucosa y lipoproteínas se correlaciona con vacas primíparas, **Figura 33 b y c**. Estos datos probablemente reflejan el comportamiento y hábitos alimenticios de las vacas multíparas en el rodeo. Como es sabido, estas se adaptan mejor a las dietas TMR parto y dominan a los animales primíparos al competir por alimento, y por lo tanto tienen un mayor DMI [103,104]. La mayor ingesta de nutrientes fermentables en estos animales conduce, a su vez, a niveles más altos de ácidos grasos volátiles en su torrente sanguíneo. Además, las vaquillas continúan creciendo durante el embarazo, y por lo tanto la concentración de hormonas anabólicas, como el IGF1, suele ser mayor. Esto impacta en la disponibilidad de nutrientes en la glándula mamaria y podría contribuir a reducir los niveles de lactato y cuerpos cetónicos circulantes en estos animales [105]. Pocos estudios han analizado el efecto de la paridad en todo el metaboloma sérico de la vaca lactante [77], hasta donde sabemos uno de los primeros trabajos que vinculan las variaciones observadas con una combinación de adaptaciones nutricionales y metabólicas.

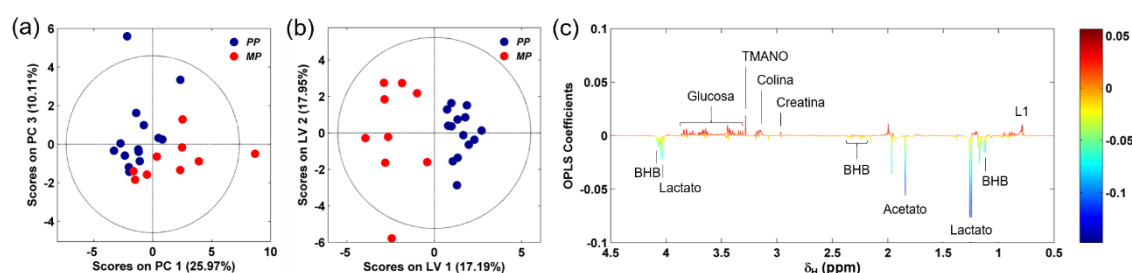


Figura 33. Gráfico de PCA muestras de suero parto (C - 4) coloreadas de acuerdo con la paridad animal **(a)**, y gráficas de OPLS-DA y su correspondiente grafico de valores de carga para la comparación entre vacas primíparas y múltiparas dentro del grupo de animales en el parto **(b,c)**. Los coeficientes R²_Y y Q²_Y del modelo fueron 0,82 y 0,56, y su curva ROC tuvo un valor AUC de 0,93.

2.4. Conclusiones

Como se detalla en las secciones anteriores, la metabolómica basada en RMN nos permitió seguir los cambios de 18 metabolitos séricos durante el período de transición de un rodeo de vacas lecheras. Basándonos en la composición del metaboloma se identificaron claramente cuatro etapas a través del periparto, evidenciando además alteraciones importantes en vías metabólicas centrales. En particular, los datos muestran la activación secuencial de varias rutas gluconeogénicas después del parto, incluyendo los ciclos de Cori y de la alanina. El análisis de los perfiles metabólicos también mostró que la contribución de las vías gluconeogénicas se ve afectada por la paridad del animal antes del parto. Estudios recientes han reportado resultados comparables en dos momentos bien definidos en el periodo de transición [76]. Nuestros hallazgos indican que los marcadores de enfermedades metabólicas subyacentes, como la cetosis y la lipidosis hepática, están presentes durante la mayor parte del período de lactancia, pero alcanzan su punto máximo cuatro semanas después del parto.

Además, los resultados presentados aquí son consistentes con los datos obtenidos empleando metabolómica dirigida por MS [75]. En general, este trabajo demuestra que los métodos basados en RMN son ideales para evaluar y seguir el impacto del manejo nutricional en la salud de rodeos lecheros durante todo el período de transición. A futuro se espera continuar con esta línea de investigación y utilizar este enfoque para evaluar los efectos de diferentes esquemas nutricionales y su impacto sobre los animales en el posparto. Además, como se describe en el capítulo IV tomando en cuenta los resultados obtenidos se empleará un nuevo set de muestras de suero y orina de vacas lecheras para el desarrollo de estrategias alternativas de procesamiento y análisis de datos. Particularmente se describe el desarrollo de un algoritmo para el uso de datos bidimensionales en el análisis metabolómico, y la posibilidad de fusionar datos de diferentes biofluidos para alcanzar una mayor especificidad en los análisis. [81,106]

Capítulo III

Identificación de fitomarcadores para la selección y mejoramiento de cultivares

3.1. Introducción

La cebada (*Hordeum vulgare* L.) es uno de los principales cultivos de invierno en importancia a nivel mundial según la Organización de Naciones Unidas para la Alimentación y Agricultura (ONU-FAO) [107]. Aproximadamente el 60% de su grano se utiliza para la alimentación animal, el 10% se emplea en la industria alimentaria y para el suministro de semillas, y el 30% restante se destina a la maltería [108]. La producción de malta con fines cerveceros impone estrictos requisitos sobre las características del grano, particularmente en lo que respecta a factores relacionados con la fermentabilidad, como el nitrógeno libre (“free amino nitrogen”, FAN) y el contenido de azúcar [109]. Por ello se necesita un desarrollo continuo de germoplasmas agronómicamente adaptados para hacer que el cultivo de cebada sea sostenible [110,111].

Entre los factores que impactan negativamente en la salud y rendimiento de los cultivos de cebada se destacan las infestaciones por áfidos (Hemiptera: Aphididae). Estos insectos fitófagos provocan daños en el tejido vegetal cuando introducen un estilete en el floema para alimentarse de la savia, y al mismo tiempo excretan saliva tóxica que inhibe los mecanismos de sellado del mismo. La extracción de grandes cantidades de fotoasimilados y otros nutrientes provoca falta de vigor, manchas cloróticas, necrosis, desecación y, en última instancia,

muerte del tejido vegetal [112]. La extracción de savia lleva además a desequilibrios de hormonas de crecimiento que debilitan aún más a la planta y atrofian su desarrollo [113]. Los pulgones también son vectores de un gran número de virus vegetales, entre ellos el virus del enanismo amarillo de la cebada ("barley yellow dwarf virus", BYDV) y el virus del enanismo amarillo de los cereales ("cereal yellow dwarf virus", CYDV) [110]. Otro daño indirecto causado por estas plagas está asociado con la producción de melaza durante la alimentación. Esta mezcla rica en azúcares se deposita en las hojas y promueve el crecimiento de diferentes hongos de la familia *Ascomycete*. que conducen a la fumagina [114,115]. A pesar de que los pesticidas se emplean ampliamente contra las infestaciones de áfidos en la cebada y otros cultivos, estos métodos de control basados en agentes químicos tienen varios inconvenientes importantes. Para empezar, varias especies de áfidos han desarrollado resistencia contra una amplia gama de insecticidas [116,117]. Además, estos insectos no se alimentan de la superficie de las hojas exclusivamente, sino que también pueden ocupar el envés de las mismas evitando así la ingestión o el contacto con los insecticidas aplicados. Dada su rápida tasa de reproducción, un pequeño número de sobrevivientes puede además conducir a una pronta recuperación de la población de áfidos. También es sabido que el uso excesivo de pesticidas afecta a los insectos beneficiosos, incluidos los que depredan a los pulgones, y puede a su vez llevar a la acumulación de contaminantes ambientales nocivos para la salud animal y humana [118].

Como respuesta a estas problemáticas han surgido los programas de manejo integrado de plagas (“integrated pest management”, IPM). Estos combinan el monitoreo de poblaciones de plagas, controles mecánicos y biológicos, y el uso de variedades de cultivos resistentes para reducir las cantidades de insecticidas necesarios en la agricultura intensiva. Esta última estrategia requiere de un entendimiento detallado de los mecanismos de resistencia constitutivos e inducidos en plantas, los cuales incluyen antibiosis, tolerancia, y antixenosis. El último refiere al grupo de caracteres de las plantas y de respuestas de los insectos que llevan a preferir a una planta en particular para la oviposición, la alimentación o refugio, o una combinación de estos [119].

La selección de cultivares resistentes se puede lograr empleando diversas herramientas, entre las cuales se ha comenzado a destacar el análisis metabolómico. De hecho, la metodología ha sido ampliamente utilizada para estudiar la tolerancia al estrés, la resistencia a patógenos, y el desarrollo de ecotipos robustos [120]. Por ejemplo, Lessi y Sánchez describen estrategias dirigidas que permitieron estudiar los mecanismos de adaptación a diferentes niveles de estrés por salinidad, revelando mecanismos conservados entre especies y otros especie-específicos [121,122]. Estos autores combinaron datos de ^1H RMN provenientes del metabolóma foliar de variedades de soja con datos provenientes de transcriptómica sobre la expresión de genes responsables de la regulación del metabolismo primario y la tolerancia al estrés oxidativo. Otros autores describen el uso de los perfiles metabólicos obtenidos por ^1H RMN para monitorear eficientemente las respuestas de las plantas al estrés biótico y abiótico

en diferentes etapas de desarrollo de cultivos arroceros [123]. En particular, Sánchez y colaboradores identifican diferencias en los niveles de terpenos, flavonoides, y saponinas como determinantes para la preferencia y crecimiento de áfidos en distintas leguminosas [124]. Otros autores identifican y determinan la distribución espacial *in vivo* de diferentes hormonas vegetales generadas en respuesta a infestaciones por estas plagas. Para esto destacan los avances en la utilización de técnicas como SPME-HPLC-MS [125]. Mas recientemente, Seong y colaboradores reportaron el uso de métodos quimiométricos basados en RMN en la correlación de los perfiles metabólicos de distintos órganos de tres cultivares de *Capsicum annuum* L. con su resistencia a diferentes enfermedades virales [126].

Respaldados por estos y otros antecedentes [127-130], en este capítulo se investiga la utilidad de la metabolómica basada en ^1H RMN en la identificación de fitomarcadores asociados con la respuesta antixenótica de *Hordeum vulgare* L. al pulgón verde (*Schizaphis graminum*).

3.2 Materiales y métodos

3.2.1 Cría de áfidos (con Silvana Abbate y Cintia Tagliani)

La cría de áfidos se realizó en el Departamento de Entomología de la Estación Experimental Mario A. Casinonni (EEMAC), UdelaR (Paysandú, Uruguay). La cría de estos insectos fue iniciada a partir de ejemplares colectados de parcelas de avena situados en la EEMAC durante el otoño. Las poblaciones fueron mantenidas sobre cebada susceptible (cultivar Danielle) en condiciones

controladas de luz y temperatura (24 ± 2 °C y 14 h de fotoperiodo) hasta lograr aproximadamente las 1000 ninfas necesarias para los ensayos de antixenosis.

3.2.2 Evaluación de respuesta antixenótica (con Silvana Abbate y Cintia Tagliani)

Se emplearon 84 genotipos de cebada seleccionados a partir de líneas avanzadas de los programas nacionales de mejoramiento de la UdelaR y el Instituto Nacional de Investigaciones Agropecuarias (INIA). Cada genotipo se sembró, por triplicado, en tubos Falcon de 15 mL (**Figura 34a**). Los tubos sembrados fueron mantenidos en condiciones de luz y temperatura controladas (18 ± 5 °C y 11 h de fotoperíodo), y una vez emergidas las plántulas se fertilizaron con una solución de KNO₃ (0,67 g/L), (NH₄)₂SO₄ (0,17 g/L), MgSO₄ (0,06 g/L) y Ca(NO₃)₂ (0,1 g/L). Se utilizó un diseño de bloques completos con 6 repeticiones en el tiempo [131]. Los ensayos se realizaron sobre plántulas de cebada en el estadio de dos hojas totalmente expandidas. Los tubos con plántulas de los diferentes genotipos se distribuyeron aleatoriamente, y fueron totalmente cubiertos con arena estéril sobre el borde de un recipiente de plástico circular de 60 cm de diámetro (**Figura 34b**). Las ninfas de pulgones ápteros se situaron en el centro de la circunferencia de arena al comienzo del ensayo, y el recipiente fue cubierto con otro igual de forma de realizar el experimento en condiciones de oscuridad. El número de pulgones sobre cada uno de los genotipos de cebada se contabilizó a las 48 y 72 h. Se clasificó como resistentes o no preferidas a las plantas que en promedio contenían menos de 10 pulgones una vez finalizado el ensayo, y susceptibles o preferidas a las que contenían más de 10.

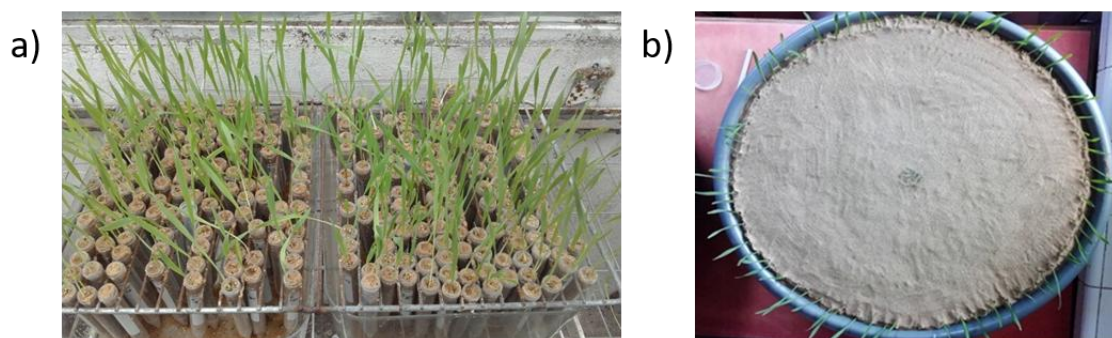


Figura 34. Plántulas en estadio de dos hojas extendidas **(a)**, ensayo de preferencia **(b)**.

3.2.3 Material vegetal (con Silvana Abatte, Ivana Cardozo, y Lucia Meneses)

A partir del grado de respuesta antixenótica frente a *S. graminum* se seleccionaron 5 genotipos de cebada primaverales. Las líneas resistentes fueron Karl (cultivar de origen norteamericano liberado en 1974) y C9614 (línea avanzada del programa de UdelaR). Las líneas susceptibles fueron INIA Ceibo (cultivar de origen uruguayo liberado en 2001), CB9215 (línea del programa de mejoramiento de UdelaR), y M104/TOCTE (línea del programa de mejoramiento ICARDA-CIMMYT [132]). Estas fueron sembradas en tubos Falcon de 50 mL con un sustrato de arena y vermiculita, se fertilizaron análogamente a lo mencionado en la Sección 3.2.2, y se regaron de dos a tres veces por semana hasta que las plántulas alcanzaron el estado de dos hojas expandidas (~15 días post siembra). Luego se cortaron unos 300 mg de las láminas foliares en trozos de $\sim 2 \times 2$ mm, y se conservaron a -80 °C para su posterior procesamiento y análisis.

3.2.4 Preparación de muestras para RMN

El material vegetal se procesó siguiendo la metodología reportada por Choi y colaboradores con modificaciones menores [133]. Una vez equilibrado a temperatura ambiente, el material vegetal triturado y homogeneizado fue extraído con 20 mL de una mezcla de agua-cloroformo-metanol (8:8:1) mediante agitación mecánica. La suspensión resultante se centrifugó a 1300 g durante 20 min, y se separaron las capas orgánica y acuosa. La primera se concentró en un evaporador rotatorio, y la segunda se liofilizó después de remover el metanol a vacío.

3.2.5 Espectroscopía de RMN

Los extractos acuosos secos se disolvieron en 0,6 mL de D₂O con 0,05% de ácido 3-(trimetilsilil)-propiónico-2,2,3,3-*d*₄ (TSP), mientras que los extractos orgánicos se solubilizaron en 0,6 mL de CDCl₃ con 0,03% de tetrametilsilano (TMS), y las soluciones resultantes se filtraron y transfirieron a tubos NMR de 5 mm (NE-HL5-7, New Era Enterprises Inc., Vineland, NJ, EUA). Todos los experimentos de RMN se llevaron a cabo a 25 °C en un espectrómetro de RMN Bruker AVANCE III 500 equipado con una sonda TXI con gradiente-z, y operando a frecuencias de ¹H y ¹³C de 500,13 y 125,76 MHz, respectivamente. En las muestras acuosas se eliminó la señal de agua residual por medio de presaturación. Todos los espectros 1D ¹H se registraron utilizando un ancho espectral de 10 kHz, un tamaño de datos de 32 K y un total de 64 barridos con un tiempo de reciclado entre barridos de 10 s. Cuando fue necesario se adquirieron espectros HSQC, HMBC, HOMO-2DJ, y

1D-TOCSY utilizando secuencias de pulso estándar proporcionadas con el espectrómetro.

3.2.6 Procesamiento de datos

Todos las FID se llenaron con ceros a 64 K y se apodizaron con una función de ventana exponencial de 0,3 Hz antes de la realizar la transformada de Fourier. A todos los espectros se les corrigió manualmente la fase y la línea base utilizando MNova (versión 10.0, MestreLab Research, S.L., Santiago de Compostela, España), y los espectros corregidos fueron luego referenciados a 0,00 ppm a la señal de TSP (muestras acuosas) o TMS (muestras orgánicas). Posteriormente se seleccionaron regiones que se alinearon por medio del algoritmo icoshift implementado en Mnova [27], y estos datos fueron finalmente normalizados al total del área del espectro después de excluir las señales residuales de los solventes (4,60-4,80 y 7,15-7,35 ppm para muestras acuosas y clorofórmicas, respectivamente). Las matrices de datos así obtenidas se exportaron como archivos de texto para su uso en los análisis estadísticos.

3.2.7 Análisis estadístico

Los análisis de PCA y OPLS-DA se llevaron a cabo con el paquete PLS_Toolbox (versión 8.5, Eigenvector Research Inc., Manson, WA, EUA) implementado para MATLAB (revisión 2014a, The MathWorks Inc., Natick, MA, EUA). Para todos los modelos, los datos se centraron en la media y se escalaron utilizando un factor de Pareto [29]. Como se mencionó en la Sección 2.3, se realizó un PCA para facilitar la identificación de tendencias o valores atípicos en los datos [38,134,135].

La validación cruzada de los modelos de OPLS-DA se logró mediante el método de subconjunto aleatorio, que involucró 20 iteraciones sobre datos divididos en 10 partes de igual tamaño. Se trazaron las curvas de ROC y se calcularon las áreas bajo las curvas para asegurar un buen ajuste de los modelos obtenidos [88,89]. También se realizaron pruebas de permutaciones con 200 iteraciones para determinar el grado de sobreajuste y validar los análisis discriminantes [45,136]. Los análisis de STOCYSY se realizaron con un programa desarrollado en MATLAB basado en el algoritmo descrito por Cloarec y colaboradores [47].

3.2.8 Agregado de estándares

Sobre una muestra de extracto acuoso representativa se adicionaron sucesivamente alícuotas de 10 μ L de soluciones estándar de 0,05 M de colina (Sigma Aldrich, pureza 98%) y betaína (Acros Organics, pureza 98%). Luego de cada adición se adquirieron nuevos espectros de ^1H RMN en las mismas condiciones descritas en la Sección 3.2.5.

3.3 Resultados

3.3.1 Evaluación de la respuesta antixenótica

Como se mencionó previamente, se consideraron 84 genotipos de cebada en los ensayos de respuesta antixenótica. De estos se seleccionaron los cinco que presentaron los resultados más contrastantes, los cuales se presentan en la **Tabla 8**.

Tabla 8. Resultados respuesta antixenótica de los diferentes genotipos de cebada.

GENOTIPO	NRO PULGONES (n = 6)	CLASIFICACIÓN
C 9614	5,8 ± 1,5	Resistente
Karl	8.0 ± 1,6	Resistente
INIA Ceibo	11.5 ± 1.1	Susceptible
CB9215	12 ± 1,4	Susceptible
M104/TOCTE	13.7 ± 2,3	Susceptible

3.3.2 Espectros de ^1H RMN de extractos de cebada

En la **Figuras 35 y 36**, y **Tablas 9 y 10** se presentan espectros de ^1H RMN representativos de extractos acuosos y clorofórmicos de cebada empleados en el estudio y las anotaciones correspondientes para los metabolitos identificados. En el extracto acuoso se identifican isoleucina (0,96 ppm), valina (1,08 ppm), leucina (1,05 ppm), lactato (1,33 ppm), alanina (1,43 ppm), arginina (1,91 ppm), metionina (2,14 ppm), acetoacetato (2,22 ppm), γ -aminobutírate (GABA, 2,31 y 3,04 ppm), citrato (2,82 ppm), etanolamina (3,14 ppm), colina (3,20 ppm), betaína (3,26 ppm), scyllo-inositol (3,36 ppm), fructosa (4,12 ppm), malato (4,37 ppm), glucosa (5,22 y 4,65 ppm), *p*-cumarato (6,38 ppm), fumarato (6,51 ppm), tirosina (7,19 ppm), fenilalanina (7,38 ppm), triptófano (7,87 ppm), formiato (8,46 ppm), y trigonelina (9,13, 8,83, y 8,06 ppm). En la fracción orgánica se pueden asignar, entre otras, las resonancias correspondientes a lípidos (señales anchas entre 0,5 y 3,3 ppm), glicerol y triglicéridos (4,18 ppm), ácido ursolínico (5,14 ppm), olefinas (5,38 ppm) y ácidos fenilpropanóicos y derivados (7,10–7,50 ppm). La asignación putativa de estos metabolitos se hizo por comparación con datos reportados en

bibliografía y búsquedas en bases de datos [49]. En algunos casos, estas fueron sustentadas con la ayuda de espectros HSQC y 1D-TOCSY.

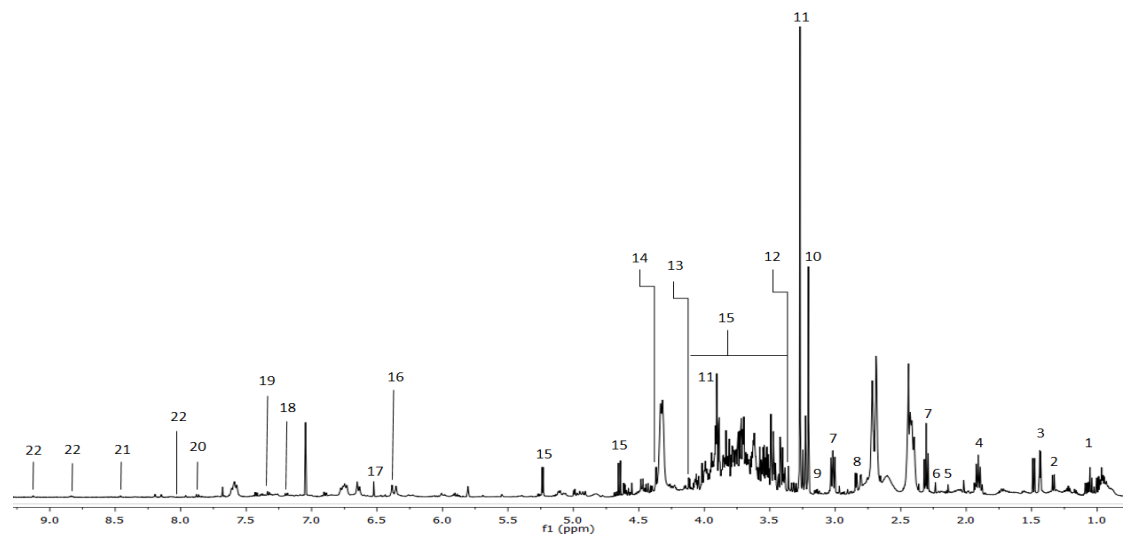


Figura 35. Espectro de ¹H RMN representativo de un extracto acuoso de cebada. Las señales correspondientes a valina, leucina e isoleucina (1), lactato (2), alanina (3), arginina (4), metionina (5), acetoacetato (6), GABA (7), citrato (8), etanolamina (9), colina (10), betaína (11), schyllo-inositol (12), fructosa (13), malato (14), glucosa (15), *p*-cumarato (16), fumarato (17), tirosina (18), fenilalanina (19), triptófano (20), formiato (21), y trigonelina (22).

Anotación	Metabolito	Corrimiento Químico (ppm)	Multiplicidad	Group
1	Valina	1,08/0,93	d/d	CH ₃ /CH ₃
1	Isoleucina	0,98/0,96	d/t	βCH ₃ /δCH ₃
1	Leucina	1,05	dd	δCH ₃
2	Lactato	1,33	d	CH ₃
3	Alanina	1,43	d	CH ₃
4	Arginina	1,91	m	βCH ₂
5	Metionina	2,14	s	S-CH ₃
6	Acetoacetato	2,22	s	CH ₃
7	GABA	2,31/3.04	dd/dd	αCH ₂ /γCH ₂
8	Citrato	5,59	dd	CH ₂
9	Etanolamina	3,14	t	CH ₂
10	Colina	3,20	s	N(CH ₃) ₃
11	Betaina	3,26/3,89	s/s	N(CH ₃) ₃ /CH ₂
12	Scyllo-inositol	3,36	s	CH
13	Fructosa	4,12	m(b)	5β-H4, H3

14	Malato	4,37	dd	CH
15	Glucosa	5,22/4,65/ 4,1-3,19	d/d/ otras señales	$\alpha\text{CH}_2/\beta\text{CH}_2/\text{CHyCH}_2$
16	p-cumarato	6,38	d	(Z)=CH
17	Fumarato	6,51	s	(Z)=CH
18	Tirosine	7,19	m	H2,H6
19	Fenilalanina	7,38	m	H2-H6
20	Triptofano	7,87	d	H4
21	Formiato	8,46	s	CH
22	Trigoleina	8,06/8,83/9,13	m/m/s	CH/CH/CH ₃

Tabla 9. Anotaciones para los metabolitos identificados en extracto acuoso de hojas de cebada en un espectro de ^1H RMN. Singulete (s), doblete (d), triplete (t), multiplete (m), señal ancha (broad, b)

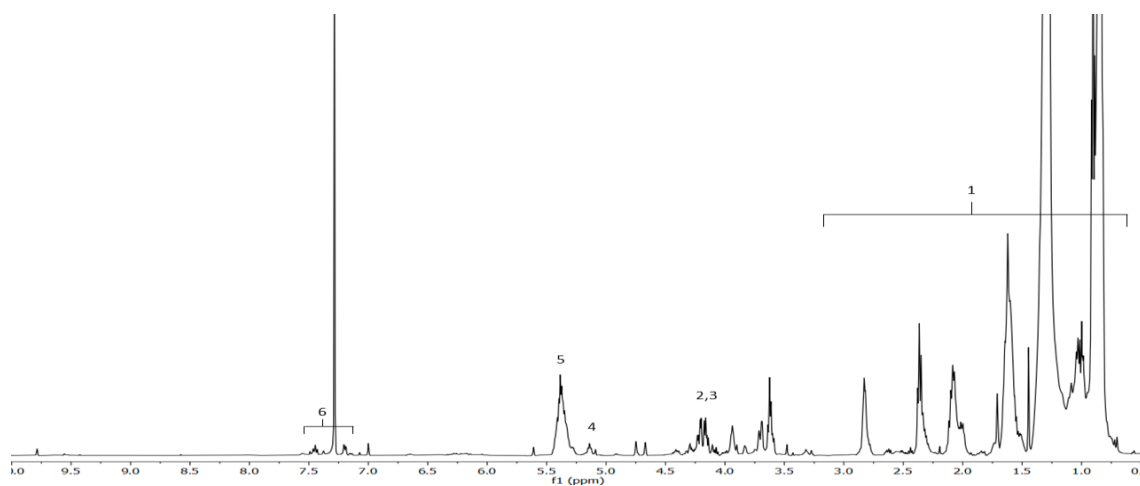


Figura 36. Espectro de ^1H RMN representativo de un extracto orgánico de cebada. Las señales correspondientes a ceras y otros lípidos (1), glicerol (2), trigliceridos (3), ácido ursolínico (4), olefinas (5), y ácidos fenilpropanoicos y derivados (6).

Anotación	Metabolito	Corrimiento Químico (ppm)	Multiplicidad	Grupo
1	Lípidos	0,5-3,3	m(b)	Protones alílicos, bisalílicos, α -carbonilo, CH ₂ y CH ₃
2	Glicerol	4,18	dd/dd	β CH ₂ / β' CH ₂
3	Triglicéridos	4,18	m(b)	-
4	Ácido Ursolínico	5,15	m(b)	CH
5	Glicerol y olefinas (lípidos)	5,38	m(b)	α CH/=CH
6	Ácido fenil propanóico y derivados	7,10-7,50	m	-

Tabla 10. Anotaciones para los metabolitos identificados en extracto orgánico de hojas de cebada en un espectro de ¹H RMN. Singulete (s), doblete (d), triplete (t), multiplete (m), señal ancha (broad, b)

3.3.3 Metabolómica basada en RMN

El paso inicial del estudio involucró la identificación de grupos entre las muestras de cebada. Para ello se llevaron a cabo PCAs considerando los datos derivados de los espectros de ¹H RMN para los extractos acuosos y orgánicos, y se colorearon los gráficos de puntuación respectivos en base a la calificación de los genotipos en los ensayos de antixenosis. Como se evidencia en la **Figura 37**, el PCA de los extractos acuosos muestra una tendencia a separar entre los genotipos según su respuesta antixenótica (**Figura 37b**), mientras que no se observan agrupamientos de los datos derivados de los extractos orgánicos (**Figura 37a**). Por lo tanto, se continuó trabajando únicamente con los datos provenientes de los extractos acuosos.

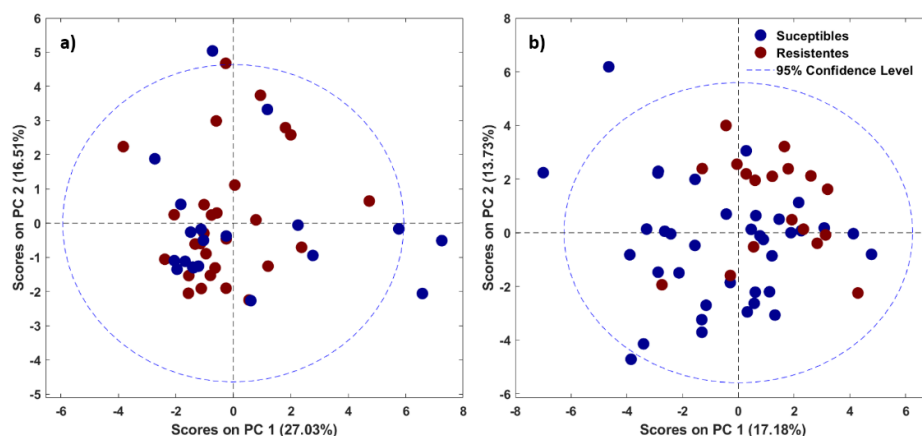


Figura 37. Gráfico de PCA de todos los extractos orgánicos (a) y acuosos (b) de cebada coloreados según respuesta antixenótica.

Basados en la tendencia evidenciada por el PCA, se llevó a cabo la comparación de ambas clases mediante un OPLS-DA. El gráfico de valores de carga de este análisis, el cual fue validado exitosamente (**Figuras S9**), nos permitió identificar señales a 3,20 ppm y 3,26 ppm como las principales discriminantes de los genotipos resistentes y susceptibles, respectivamente (**Figura 38**).

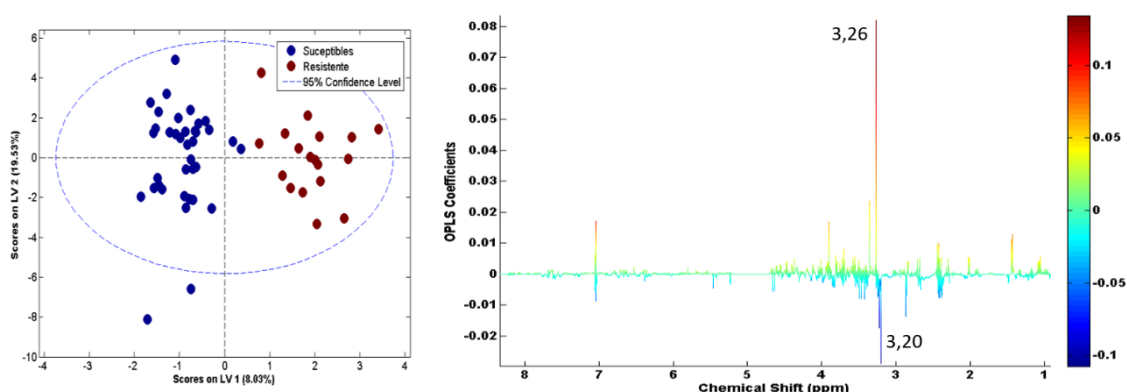


Figura 38. Gráfico de OPLS y su respectivo gráfico de valores de carga obtenidos de comparar el grupo de genotipos resistentes vs susceptibles. Los coeficientes R^2Y y Q^2Y para el modelo fueron 0,85 y 0,58, y su curva ROC obtuvo valores de AUC de 0,94 (S9).

Para confirmar la identidad de las señales identificadas se realizaron inicialmente pseudoespectros de STOCSY a fin de poder extraer más información estructural correlacionada con estos metabolitos (**Figura 38**). Se utilizaron las señales a 3,20 y 3,26 ppm, respectivamente, como directrices del análisis. Como se observa en la **Figura 39a**, la señal a 3,20 ppm presenta correlaciones con otra a 4,07 ppm, así como una correlación más débil con una a 3.61 ppm. Para la señal a 3,26 ppm se observó únicamente una correlación con una a 3.91 ppm (**Figura 39b**).

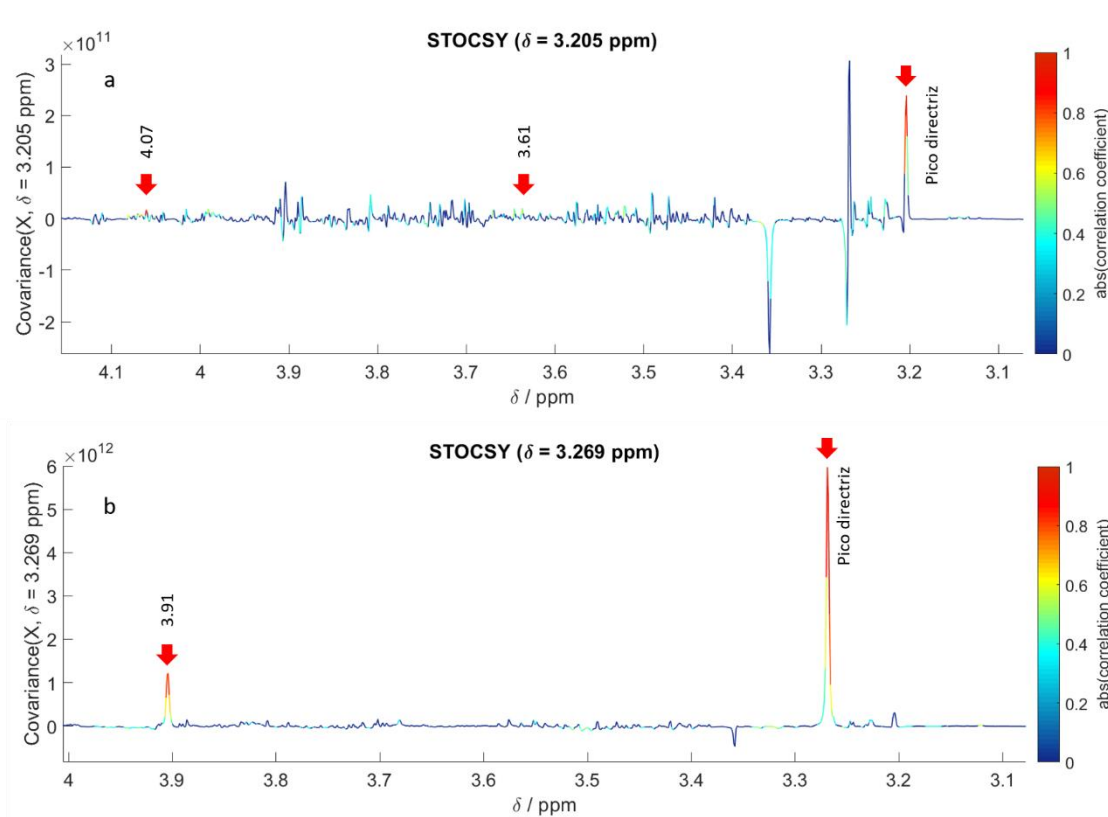


Figura 39. Trazas de pseudoespectros STOCSY obtenidos utilizando como directrices las señales a 3,20 (a) y 3,26 ppm (b).

Los análisis de STOCSY evidencian que el metabolito asociado con la señal a 3,20 ppm contiene un sistema de tres espines, mientras que el correspondiente a la señal en 3,26 ppm contiene uno de dos. Para obtener más información estructural

de ambos metabolitos se investigó la multiplicidad de estas resonancias utilizando espectroscopía HOMO-2DJ. Como se observa en la **Figura 40**, las señales correlacionadas a 3,20, 3,61 y 4,07 ppm corresponden, respectivamente, a un singlete y dos multipletes (i.e., un grupo CH y dos grupos CH_n). Por otro lado, las resonancias a 3,26 y 3,91 ppm son ambas singletes (i.e., dos grupos CH).

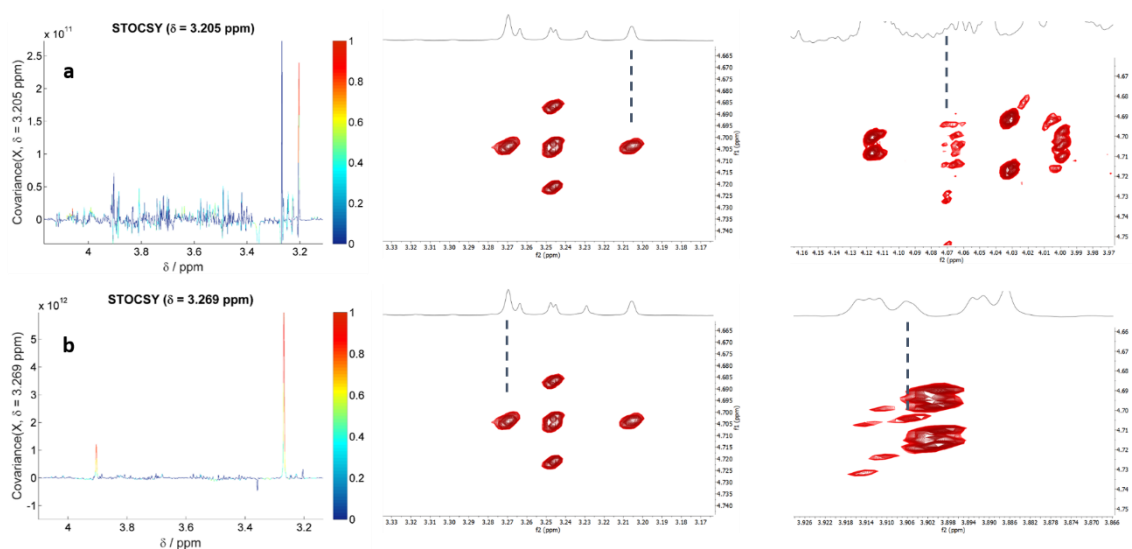


Figura 40. Ampliaciones de un espectro HOMO-2DJ de una muestra de extracto acuoso de cebada indicando las multiplicidades de las señales a 3,20, 3,61 y 4,07 ppm (a), y 3,27 y 3,91 ppm (b).

Luego se estudió el espectro de HSQC, el cual nos permite encontrar los pares ^1H - ^{13}C para las señales correlacionadas en los gráficos de STOCSY (**Figura 41**). Para el par de señales de ^1H a 3,21 y 4,07 ppm se encontraron correlaciones de ^{13}C a 53,8 y 55,6 ppm, respectivamente, mientras que el solapamiento impidió identificar la correlación en ^{13}C para la señal a 3,61 ppm. Para el par de señales de ^1H a 3,28 y 3,91 ppm se encontraron correlaciones con señales de ^{13}C a 53,3 y 66,3 ppm, respectivamente.

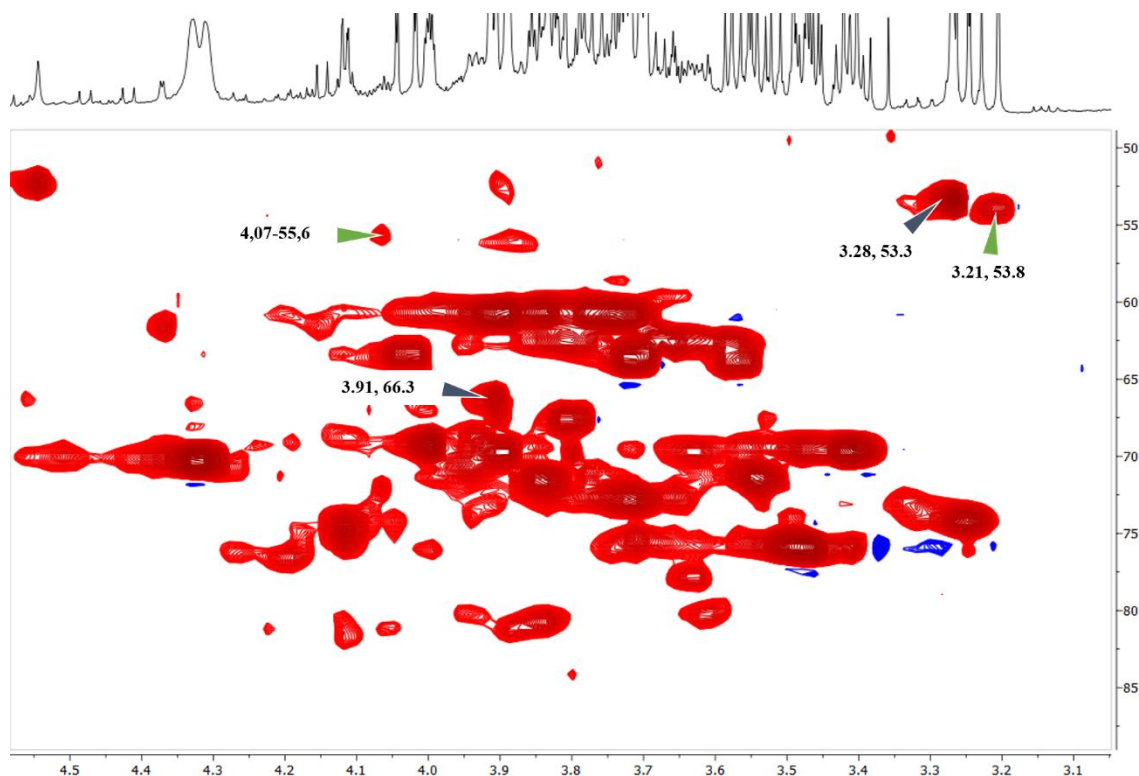


Figura 41. Ampliación de un espectro HSQC de un extracto acuoso de cebada donde se indican los pares ^1H - ^{13}C para las señales a 3,21 y 4,07 ppm (verde), y 3,28 y 3,91 ppm (azul).

Estos datos espectroscópicos, y en particular los pares ^1H - ^{13}C , se utilizaron como entrada para búsquedas en la base de datos HMDB [48,49]. Luego de descartar algunos candidatos por no ajustarse a toda la información estructural, se identificó a la colina y la betaína como las especies en cuestión. La primera presenta tres sistemas de espines que coinciden con la información estructural experimental obtenida en corrimientos químicos de ^1H y ^{13}C , así como en multiplicidad. Por otro lado, la betaina presenta dos sistemas de espines independientes que en ambos casos son singuletes. Los corrimientos químicos de ^1H y ^{13}C también coinciden con los reportados en bases de datos para esta tetralquilamina.

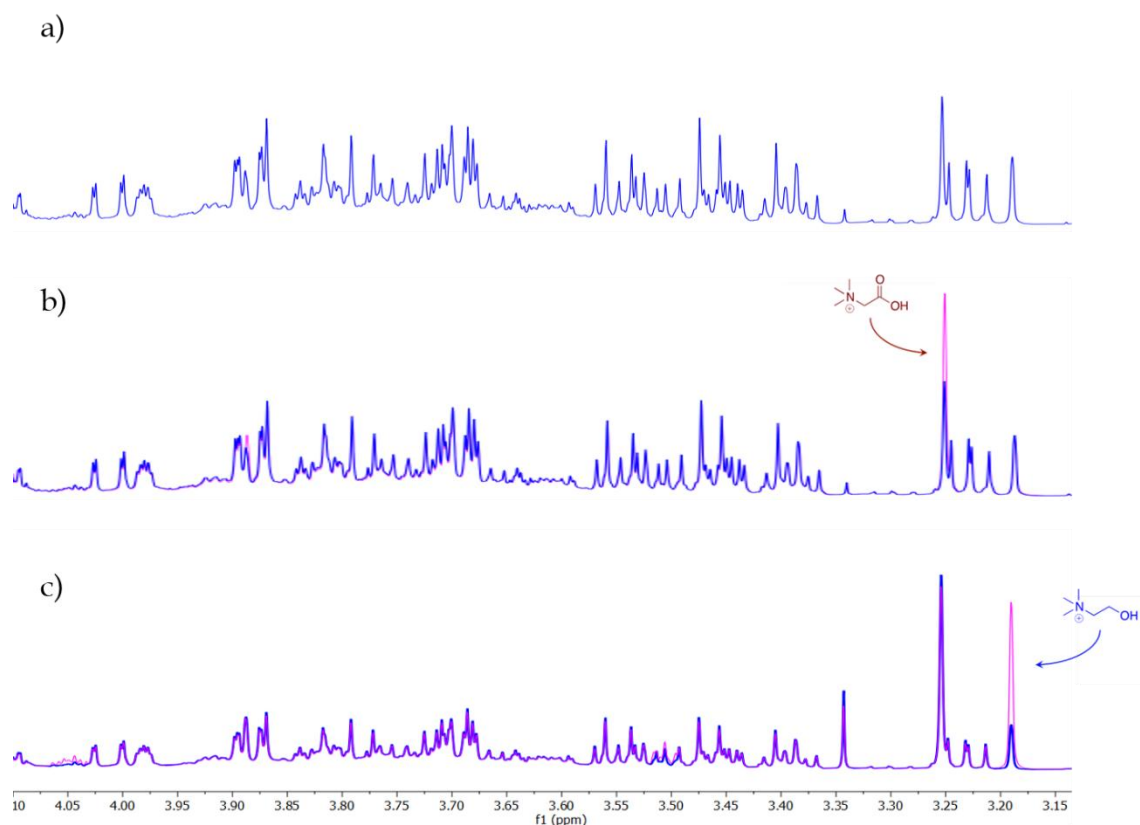


Figura 42. Agregado secuencial de estándar sobre una muestra de extracto acuoso de cebada, espectro base **(a)**, agregado de colina **(b)**, y agregado de betaína **(c)**.

El paso final en la anotación de ambos metabolitos involucró la adición de estándares. El agregado de colina y betaína se realizó en forma secuencial sobre un extracto acuoso de cebada. Podemos observar que al superponer los espectros base y el espectro con el agregado de betaína (**Figuras 42a y 42b**), las señales para este metabolito aumentan. Análogamente, al superponer el espectro con agregado de betaína con el obtenido luego del agregado de colina (**Figuras 42b y 42c**), observamos que las señales para este último metabolito aumentan. De este modo podemos confirmar con un nivel de confiabilidad 4 (ver Sección 1.7.3) a los metabolitos como colina y betaína.

3.4 Discusión y Conclusiones

Nuestros resultados permitieron identificar y correlacionar a la betaína y la colina con los genotipos no preferidos y preferidos por los pulgones, respectivamente. Ambos metabolitos son bloques de construcción ampliamente distribuidos, y están vinculados por dos transformaciones mediadas por la colin-monooxigenasa (CMO) y la betainaldehído-deshidrogenasa (BADH). A su vez, la biosíntesis de la colina en plantas se da por metilación de la etanolamina en varios pasos, los cuales están mediados por la fosfoetanolamina-*N*-metil-transferasa (PEAMT, **Figura 43**) [137-139].

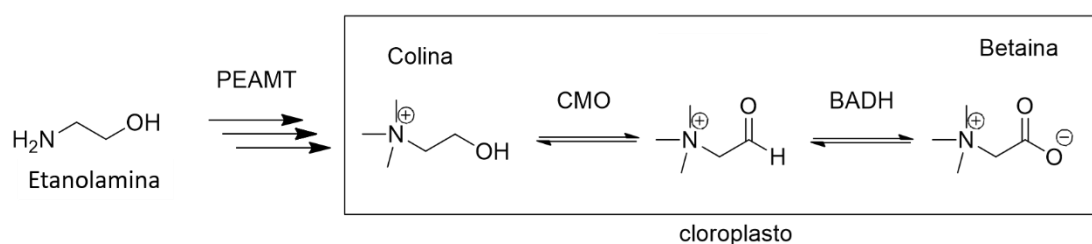


Figura 43. Esquema biosintético de la colina y betaína en plantas.

La bioacumulación o depleción de estos metabolitos depende del flujo, es decir de qué tan rápido o eficientemente se produzcan o consuman. Por lo antes dicho, se puede plantear la hipótesis de que las diferencias están dadas a nivel de la expresión de alguna de estas enzimas, o a la aparición de isoformas de las mismas con diferentes eficiencias [140]. Para confirmar esta propuesta se requieren de estudios genómicos que evalúen la expresión de estas enzimas y sus alteraciones.

Por otro lado, los áfidos elijen una planta por medio de un proceso complejo que no está dictado por un único factor. Mediante mecanismos

fisicoquímicos estos insectos reciben información sobre la textura, olor, color, y gusto de la planta. Estos procesos les permiten evaluar, además, la composición química, siendo nutricionalmente relevante el contenido de amino ácidos [141-144]. Algunos autores indican que la presencia de altas concentraciones de betaína también favorece la reproducción y la ovoposición [145]. Esto último contradice, en principio, nuestros hallazgos. Sin embargo, hay que tener en cuenta que los resultados reportados por Zuñiga y colaboradores fueron obtenidos sobre plantas bajo estrés hídrico, lo que provoca la acumulación de metabolitos osmoprotectores como la betaína [139,145]. Las plantas mejor adaptadas al estrés hídrico podrían tener mejores aptitudes nutricionales que favorecerían la selección por parte del pulgón [145]. Por otra parte, es importante mencionar que la selección de cultivares está orientada a variedades con genéticas multirresistentes [144,146-148]. La expresión conjunta de varios genes puede conferirle a la planta ventajas frente a diferentes factores de estrés en forma simultánea, explicando porque una planta agronómicamente adaptada al estrés hídrico posea también resistencia al pulgón [149].

Por otra parte, la colina presente en la planta puede estar involucrada en los mecanismos de adaptación de los áfidos a las defensas vegetales. Como se mencionó previamente, al insertar el estilete los pulgones secretan saliva que contiene varias proteínas encargadas de inhibir las defensas de las plantas y sostener su alimentación. Entre estas encontramos al grupo de las glucosam-etanol-oxidoreductasa (GCM), las cuales son parte de una gran familia de flavoproteínas (FAD). Formando parte de este grupo está la colina

deshidrogenasa (CDH), la cual tiene como sustrato principal colina exógena al pulgón. Dado que este complejo enzimático es responsable de la gelificación de proteínas de la planta e interviene en la desintoxicación de aleloquímicos vegetales [150-152], mayores concentraciones de colina podrían explicar las preferencias observadas.

Se requieren estudios adicionales para esclarecer los mecanismos que le confieren resistencia a estos cultivos y su interacción con los pulgones. El diseño experimental empleado en nuestro caso lleva a resultados relacionados con defensas constitutivas. Por lo tanto, es necesario realizar nuevos ensayos que evalúen mecanismos de defensa inducidas. Por ejemplo, enfrentar a las variedades de cebada consideradas en este estudio al pulgón podría revelar cambios en el metaboloma de la planta inducidos por infestaciones con estos insectos. Todos estos resultados contribuirán a un mejor entendimiento de los mecanismos involucrados en la interacción planta-áfido. Además, conducirían a la identificación de fitomarcadores de resistencia robustos, los cuales contribuirían a los procesos de selección y mejora de este y otros cultivares.

Capítulo IV

Uso de metodologías no tradicionales de procesamiento y análisis de datos de RMN en estudios metabolómicos

4.1 Introducción

Como se aprecia en los capítulos anteriores, el análisis metabolómico basado en RMN se realiza casi exclusivamente sobre espectros ^1H monodimensionales. Sin embargo, hay ejemplos del uso de experimentos de RMN multidimensional y multinuclear en estos estudios. Algunos han empleado espectros J-resueltos (J-res) o HOMO 2DJ y sus proyecciones como datos de entrada, ya sea sustituyendo los espectros ^1H RMN o usándolos como apoyo para resolver problemas de superposición de señales en espectros ^1H 1D. Esto ha sido particularmente útil en la identificación de metabolitos y en la cuantificación a través de la deconvolución de señales [153,154]. Fonville y colaboradores demostraron el potencial de la espectroscopía HOMO 2DJ en métodos de reconocimiento de patrones. En particular, trabajaron con las proyecciones de los espectros bidimensionales [155]. Ludwig y Viant desarrollaron herramientas para emplear la espectroscopía HOMO 2DJ en MVDA y proponen estrategias para visualizar e interpretar los resultados. Además, discuten ventajas y desventajas de emplear estos espectros en el análisis metabolómico. En su trabajo mencionan como un problema relevante en la cuantificación absoluta la presencia de acoplamiento fuerte, particularmente a campos bajos [156]. Recientemente, Hoijemberg y colaboradores desarrollaron algoritmos que emplean datos derivados de

espectros HOMO 2DJ para la asignación de señales e identificación de metabolitos [157-159].

A pesar de su mayor costo en términos de tiempo de equipo y la complejidad en el procesamiento de los datos, los espectros HOMO 2DJ representan una alternativa para abordar algunas de las limitaciones que surgen al trabajar con espectros ^1H 1D. Estos últimos suelen presentar una gran superposición de señales, lo que dificulta tanto su procesamiento como la identificación inequívoca de compuestos y su cuantificación. Al separar la información de acople escalar (J) y corrimiento químico (δ) en dos dimensiones, los espectros HOMO 2DJ disminuyen en gran medida este problema. Sin embargo, al igual que con los espectros ^1H 1D, uno de los principales desafíos en la metabolómica basada en espectros HOMO 2DJ es el procesamiento de datos. Como se menciona en la sección introductoria de esta tesis, una etapa particularmente relevante es el alineamiento de las señales. Tal cual reportan Fonville y Ludwig [156, 157], esto también es crucial para emplear los espectros HOMO 2DJ.

Para resolver este problema en los espectros 1D, la literatura reporta varios métodos de alineamiento presentados previamente en esta tesis. Por ejemplo, el PAFFT está basado en alineados parciales de segmentos en los espectros y posteriores ajustes empleando corrimientos rígidos, sumado a análisis de componentes principales para detectar regiones no alineadas en combinación con FFT para reducir el costo computacional. Por otro lado, el BAA emplea modelos de estadística bayesiana para alinear pares de espectros [26]. Finalmente, el

algoritmo icoshift emplea corrimientos por intervalos de los espectros en combinación con FFT, permitiéndole trabajar rápidamente y en simultáneo con grandes sets de espectros [27, 160-163]. A grandes rasgos, icoshift busca maximizar la correlación cruzada entre un espectro o conjunto de espectros al desplazarlos respecto a una referencia. Esta última puede ser un espectro en concreto o el promedio ponderado de todo el conjunto [27,160-162]. Sin embargo, existe un solo método de alineamiento reportado para espectros HOMO 2DJ desarrollado por Hoijemberg y colaboradores. Este toma las proyecciones de los espectros inclinados como datos de entrada para la selección de picos, y luego compara los multipletes en esos picos frente a los del espectro de referencia, generando una tabla de correspondencias por emparejamientos de multipletes. Finalmente, el algoritmo utiliza las correspondencias establecidas para los diferentes multipletes para regenerar espectros alineados, o bien permitiendo el análisis estadístico (ya sea STOCSY o bien MVDA) sobre el vector de multipletes concatenados, para luego reconstruir pseudoespectros o vectores de carga [159]. Como alternativa, en esta sección se describe el desarrollo de un algoritmo que adapta icoshift para su aplicación a espectros HOMO 2DJ y posterior uso en MVDA. A este algoritmo lo nombramos alineamiento de espectros J-res por icoshift en Bloques, o JIBA por su nombre en inglés (“J-res icoshift block alignment”).

Gracias a los avances en las capacidades analíticas y el procesamiento de datos, otro enfoque que ha cobrado relevancia en los últimos años es la fusión de datos [163-167]. Esta estrategia busca integrar información de múltiples matrices

biológicas, como suero y orina, o de distintas técnicas analíticas, como RMN y MS. Sin embargo, la literatura relacionada con este tipo de enfoques es aún limitada, dado que trabajar con varias plataformas analíticas o matrices biológicas asociadas a un mismo estudio puede no estar al alcance de muchos grupos de investigación. A pesar de esto, algunos estudios han evaluado los beneficios de la fusión de datos. Por ejemplo, se ha utilizado la combinación de espectros de ^1H RMN y SNIF-RMN para determinar el origen y la edad de vinos [164]. También se han comparado modelos de PLS con datos de RMN y NIR por separado y concatenados para predecir propiedades de alimentos [165]. Otros reportes describen diferentes metodologías para combinar datos y abordan los aspectos del preprocesamiento según la naturaleza y origen de los mismos [166,167]. En muchos de estos estudios, la concatenación de matrices ha mostrado mejoras en los modelos de clasificación. Recientemente, se ha propuesto una clasificación de estas metodologías en tres niveles, dependiendo de si se realiza o no una etapa de selección de variables o curado previo a la obtención de los modelos [167]. Como última sección de este manuscrito detallamos el desarrollo de algunas herramientas para la fusión de datos y su uso en metabolómica.

4.2 Materiales y métodos

4.2.1 Recolección de muestras orina y sangre

Los datos empleados en esta sección fueron obtenidos a partir de muestras de suero y orina de un rodeo lechero en el periparto. El protocolo experimental para la obtención de estas muestras fue aprobado por la Comisión Honoraria de

Experimentación Animal de la Universidad de la República. Se utilizó un grupo de 10 vacas Holstein multíparas, manejadas en conjunto en el tambo de la EEMAC. Las muestras se recolectaron dos veces, siete días antes y siete días después del parto. La dieta de los animales consistía en TMR a base de concentrado y ensilaje antes del parto, y se ofreció el pastoreo como una opción después de eso. Las muestras de orina y sangre se recolectaron en tubos estériles y se almacenaron a -20 °C hasta su análisis.

4.2.2 Preparación de muestras

Las muestras de orina y suero se dejaron descongelar a temperatura ambiente y se mezclaron alícuotas de 400 µl con 200 µl de solución tampón fosfato 1.5 mM pH 7.4 en D₂O [44]. Luego, las soluciones de muestra se centrifugaron a 13000 g durante 10 min y se transfirieron los sobrenadantes a tubos de RMN de 5 mm (NE HL5 7, New Era Enterprises Inc., Vineland, NJ, EUA) [10].

4.2.3 Adquisición de datos de RMN

Todos los espectros de RMN se registraron a 25 °C en el equipo mencionado en la sección 2.2.3. Los espectros de suero se adquirieron como fue descrito en la sección 2.2.3. Para las muestras de orina se adquirieron espectros ¹H 1D utilizando una secuencia de NOESY-1D con supresión de agua, acumulando 128 escaneos, 64K puntos de dominio de tiempo, un tiempo de relajación de 4 s entre escaneos, y una ventana espectral de 20 ppm. Las FID se multiplicaron por una función de ponderación exponencial con un LB de 0,3 Hz. Para los espectros HOMO 2DJ se empleó la secuencia jresgpprqf disponible con el equipo. Los

espectros se adquirieron utilizando un ancho espectral de 16,5 ppm y 8K puntos para f2, y 40 incrementos con una ventana espectral de 75 Hz para f1, acumulando 32 barridos por incremento. Los espectros fueron preprocesados con un llenado con ceros a 64K en f2 y 128 en f1. Las FID se multiplicaron en f2 por una función de ponderación exponencial ($LB = 0,3$ Hz).

Todos los datos se referenciaron a la señal del protón anomérico de la glucosa a 5.22 ppm [71,86].

4.2.4 Software

El código de JIBA se desarrolló en Matlab. El mismo alinea por bloques delimitados en f2 una matriz de datos de espectros HOMO 2DJ empleando icoshift. Los algoritmos de icoshift para el entorno Matlab están disponibles en www.models.life.ku.dk. Además, se desarrollaron otros algoritmos que permiten visualizar los resultados de emplear datos alineados con JIBA en análisis estadísticos.

Los algoritmos de concatenación de matrices también fueron desarrollados en Matlab. Estos unen los datos de dos matrices para una misma observación en forma automática.

Todos los códigos de estos algoritmos se encuentran en el Anexo 1.

4.2.5 Análisis estadísticos

El análisis de datos se realizó en Matlab R2017b (Mathworks, Inc, EUA.) utilizando el paquete PLS_Toolbox (versión 8.5, Eigenvector Research Inc., Manson, WA, EE. UU.) de manera análoga a lo descrito en la sección 2.2.6.

4.3 Resultados y Discusión

4.3.1 Desarrollo de JIBA

Para evaluar el desempeño de JIBA, se trabajó con los espectros HOMO 2DJ de suero bovino sin inclinar, inclinados, e inclinados/simetrizados. Tal como se esperaba el desalineamiento en las señales se observa principalmente en f2 (Figura 46). El primer paso consiste en obtener la matriz 3D de todos los espectros HOMO 2DJ. Para alinear con JIBA, se delimitan las regiones en f2, considerando las proyecciones superpuestas de todos los espectros HOMO 2DJ. De esta forma, definimos un bloque de dimensiones $n \times i \times j$, donde n es el número de espectros, i es el número de puntos delimitados en f2 para cada región, y j es el número de puntos en f1. Luego, el algoritmo toma cada vector ij dentro de cada espectro y lo alinea con el vector ij de un espectro seleccionado como referencia por el usuario. El proceso continúa alineando cada vector ij hasta completar todos los espectros dentro del bloque. Posteriormente, el mismo procedimiento se aplica a cada bloque delimitado en f2 hasta que se haya alineado toda la matriz 3D (Figura 44).

Se debe tener en cuenta que tanto la elección de los bloques como el espectro de referencia afectan los resultados obtenidos con JIBA. En particular, la elección del espectro de referencia afecta al algoritmo del mismo modo que lo hace icoshift para espectros 1D.

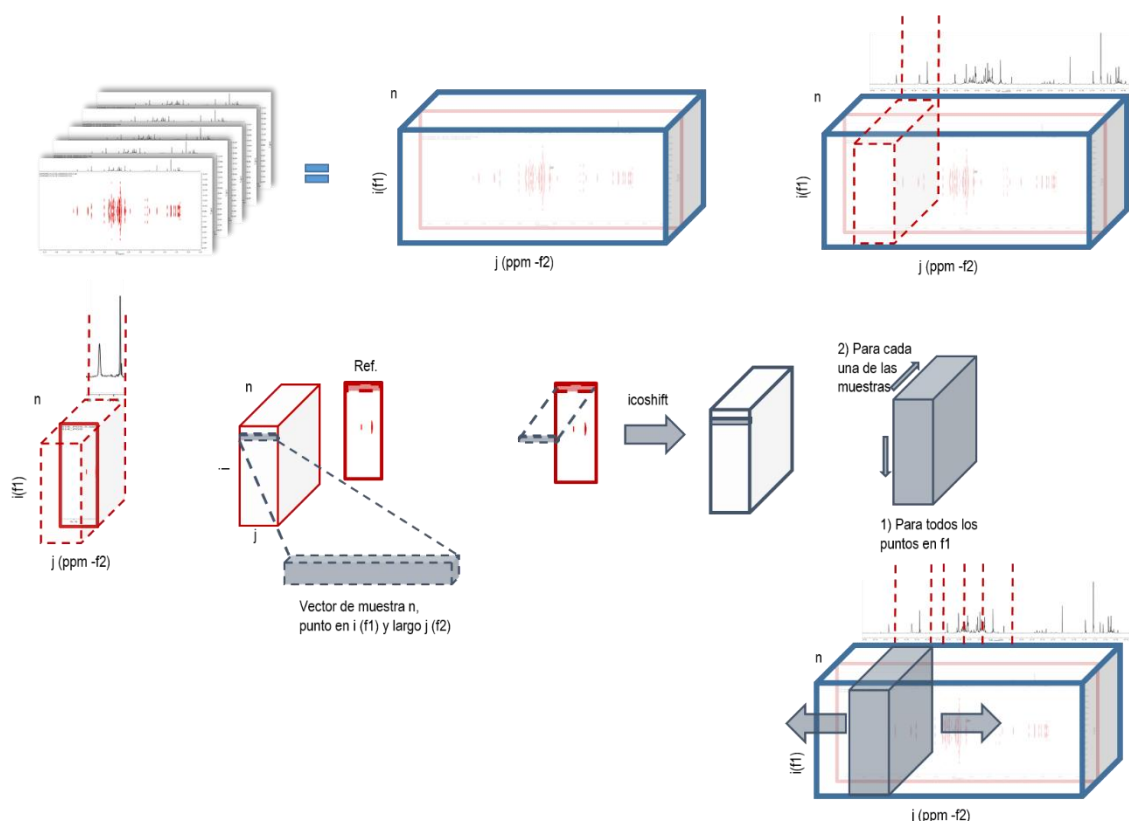


Figura 44. Representación gráfica del funcionamiento de JIBA.

4.3.2 Alineación empleando JIBA

JIBA permite observar los datos antes y después de aplicar el alineamiento bidimensional para el bloque delimitado en f_2 . Para evaluar el desempeño tenemos dos posibles visualizaciones, la de las proyecciones en f_2 y una vista en 2D (**Figura 45**). Para esta visualización podemos optar por ver un espectro en concreto en comparación con nuestra referencia seleccionada, o todos los espectros juntos (**Figuras 45a y 45b**).

Los espectros sin inclinar e inclinados no presentaron la mejor performance al aplicar JIBA, dado que se produjo un mayor grado de deformación de las señales (ver suplementario, Figura S10). Por esta razón se decidió continuar trabajando con los espectros inclinados/simetrizados. Con

estos se pudo observar cómo las regiones fueron alineadas correctamente. Esto es evidente al examinar las proyecciones en f2 y también el espectro alineado en 2D (Figuras 45a y 45b).

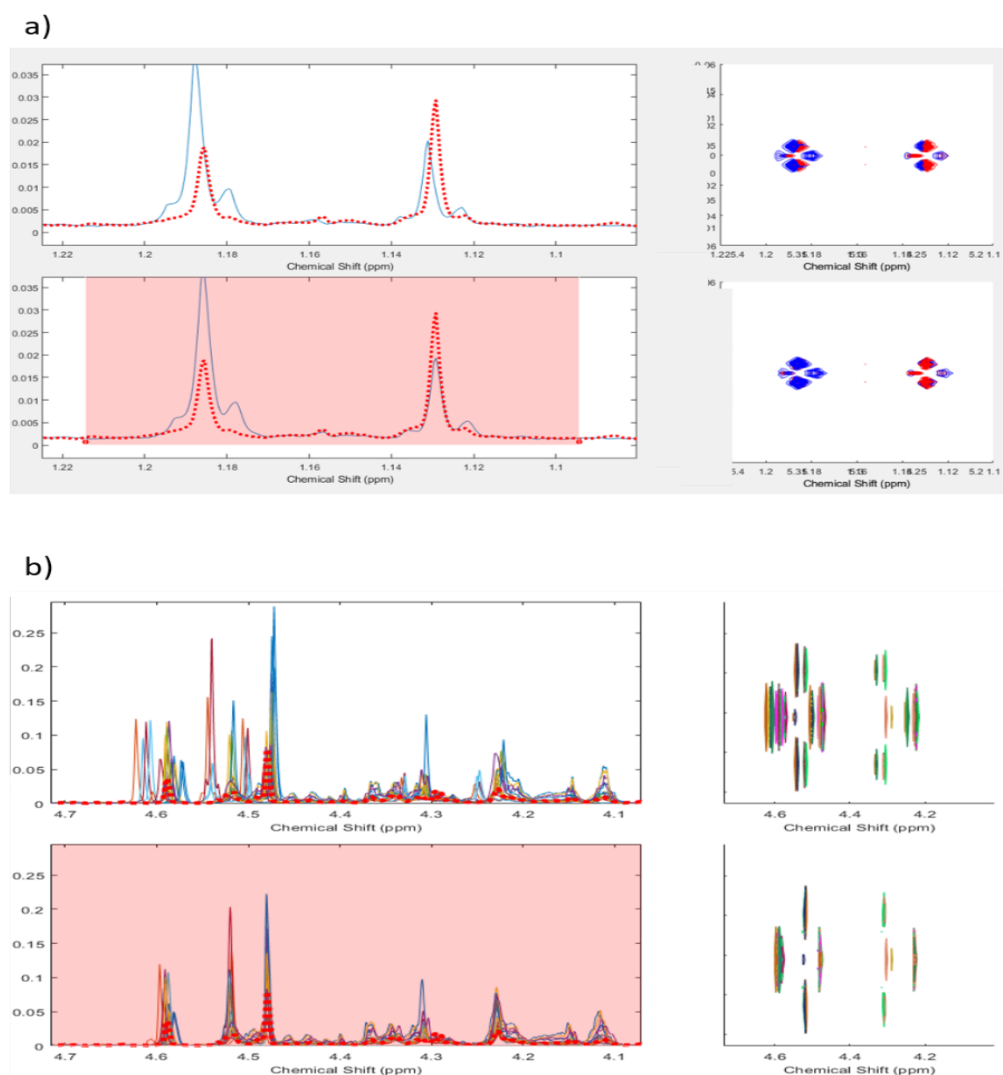


Figura 45. Visualización en 2D y proyecciones de dos espectros HOMO-2D J en un segmento alineado por JIBA **(a)**. Visualización en 2D y proyecciones de un conjunto de espectros HOMO-2D J en un segmento alineado por JIBA **(b)**. El espectro en rojo punteado indica el espectro de referencia.

4.3.3 Estrategias para el uso de espectros HOMO 2D J en MVDA

Para poder realizar los análisis estadísticos fue necesario transformar los datos 2D en un vector lineal (1D) de largo f_1 (J) por f_2 (ppm). Con los vectores 2D estirados se construyó la matriz de datos para realizar los MVDA de manera análoga a si se tratara de datos en una dimensión. A partir de los resultados obtenidos, se pudo reconstruir un gráfico de valores de carga en 2D análogo a un HOMO 2DJ (**Figura 46b**). Para esto se desarrollaron los algoritmos correspondientes que permiten tanto estirar un espectro HOMO 2DJ, así como reconstruirlo (**Anexo 1**). Este último algoritmo fue el que se empleó para generar el gráfico de valores de cargas análogo a un espectro HOMO 2D J a partir del vector de valores de cargas obtenido del análisis de OPLS de los espectros estirados (**Figura 46a**).

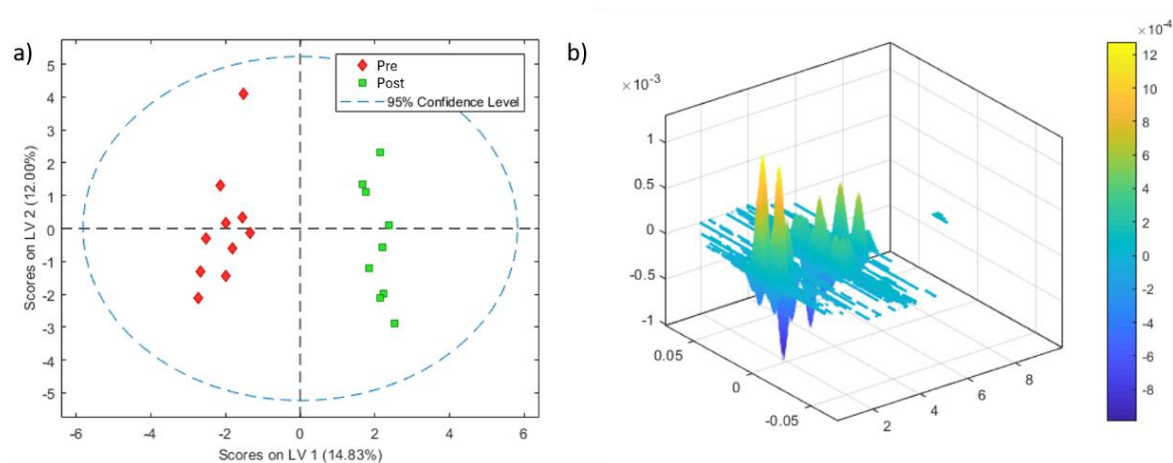


Figura 46. Gráfico de puntuación de OPLS-DA de los espectros estirados (**a**). Gráfico de valores de carga bidimensional construido a partir del vector lineal de valores de carga obtenido del OPLS-DA realizado con los espectros bidimensionales (**b**).

Posteriormente se realizó la comparación con los gráficos de valores de cargas obtenidos de realizar los OPLS-DA con los espectros de ^1H RMN CPMG y las proyecciones de los HOMO 2DJ (**Figura 47a y 47b**). Estos a su vez se compararon con el gráfico de valores de carga obtenido de los espectros HOMO 2DJ. Para poder comparar los gráficos de valores de cargas de datos 1D con el gráfico de valores de carga 2D se desarrolló un algoritmo que genera una proyección en f2 del gráfico de valores de cargas de los HOMO 2DJ (**Figura 47c**). Al comparar los diferentes gráficos se puede observar claramente como al trabajar con los espectros HOMO 2DJ se obtienen gráficos más limpios que permiten identificar mejor las señales de importancia que no se pudieron apreciar en los otros gráficos de valores de carga, las cuales fueron señaladas en verde en **Figuras 47c**.

Por otra parte, observamos en los gráficos de valores de carga la desaparición de la señal de la colina y ambigüedades en la señal del acetato, indicadas en rojo en la **Figura 47**. Ambas son señales muy intensas en los espectros HOMO 2DJ y presentan ruido de T_1 (**Figura S11**). En el preprocesado este artefacto se elimina, alterando así los valores de área para estos metabolitos. Esto lleva a que su contribución en los análisis de OPLS-DA se vea afectada.

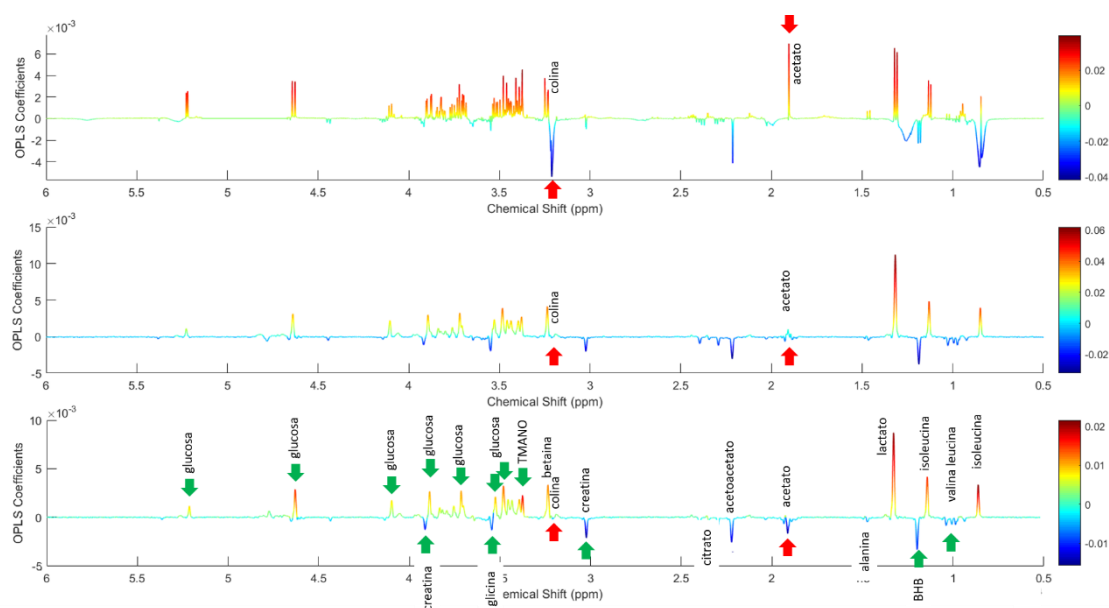


Figura 47. Gráfico de valores de carga obtenido de espectros CPMG (a), de las proyecciones de espectros HOMO 2D (b), y proyección del gráfico de valores de carga en 2D obtenido a partir de espectros HOMO 2D (c), ver Figura 46b. En rojo se resaltan las señales que se aprecian mejor al pasar de trabajar con espectros CPMG a los HOMO 2D.

4.3.4 Fusión de datos

Como prueba de concepto de fusión de datos se emplearon espectros ^1H RMN CPMG de suero y NOESY-1D de orina de una cohorte de vacas en torno al periparto. Los algoritmos desarrollados permiten la fusión de ambas matrices buscando un identificador común en el nombre que identifica cada muestra de ambos sets de datos. Una vez que ambas matrices fueron fusionadas se les pueden realizar los mismos procesados previos a los MVDA y obtener de estos un gráfico de valores de cargas de los datos fusionados (**Figura 48 y 49**). En la matriz fusionada se le sumaron 10 unidades al vector de corrimiento químico de suero para identificar las variables y posteriormente poder construir los gráfico de valores de carga y de STOCSY. Los modelos para las matrices individuales y concatenadas tuvieron una buena performance al clasificar los animales antes y

después del parto. Sin embargo, podemos observar una mejora en la clasificación y una menor dispersión de los datos para el modelo concatenado (**Figura 48**). En esta **Figura 48b** se observa que la muestra de orina indicada con una flecha azul se encuentra mal agrupada en el PCA. Por otro lado, esta muestra se agrupa correctamente en el PCA obtenido con datos concatenados de suero y orina (**Figura 48c**).

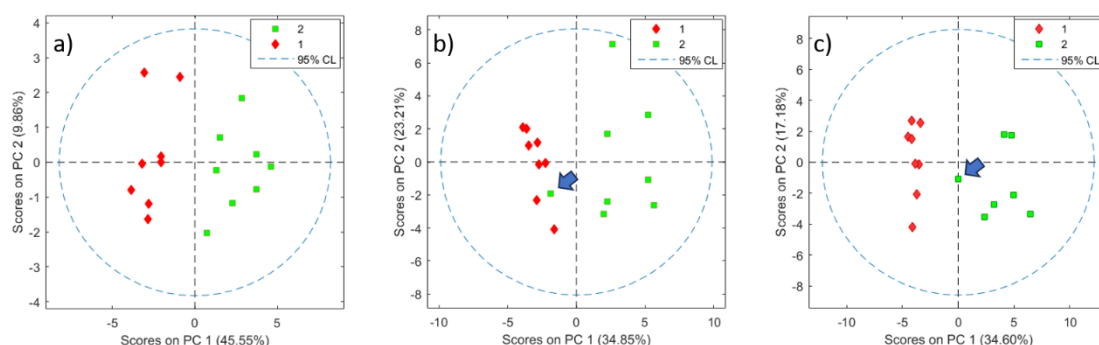


Figura 48. PCA de muestras de suero (a), orina (b), y datos concatenados suero-orina (c) para vacas antes (1) y después del parto (2)

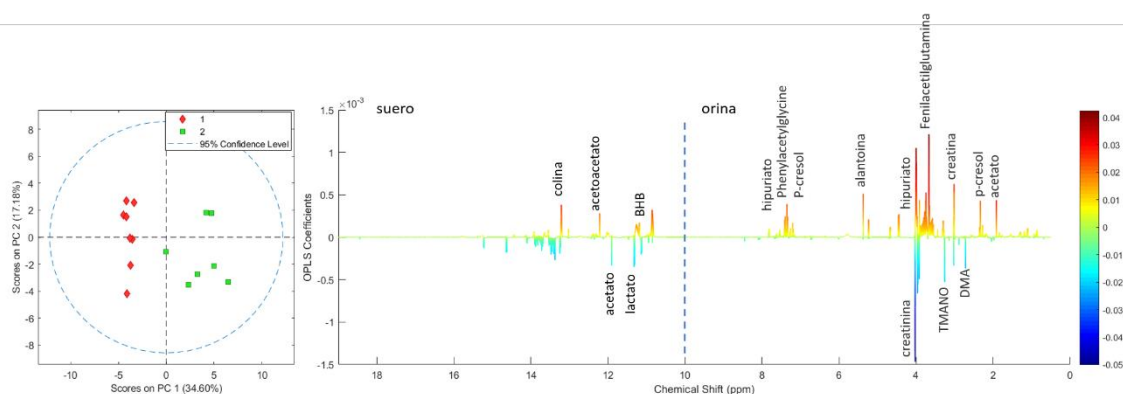


Figura 49. PCA de datos concatenados suero-orina para vacas antes (1) y después del parto (2), y gráfico de valores de carga del OPLS comparando matrices concatenadas suero-orina, antes y después del parto. Las señales positivas de color rojo oscuro se correlacionan con las vacas posparto y las negativas de color azul oscuro con las vacas preparto. Se construyó un nuevo eje de corrimiento químico concatenado para construir el gráfico. Se le adicionaron 10ppm al eje de corrimiento químico correspondiente al suero.

Sobre los datos fusionados también es posible la aplicación de algoritmos de STOCSY para buscar correlaciones estructurales y fisiológicas. En este caso particular entre suero y orina, por ejemplo, al emplear como pico directriz a la señal de la urea en suero a 5,77 ppm (15,77 ppm en nuestro vector de corrimiento químico concatenado) se pueden observar correlaciones biológicas con la fenilglutamina e hipurato en orina (**Figura 50**).

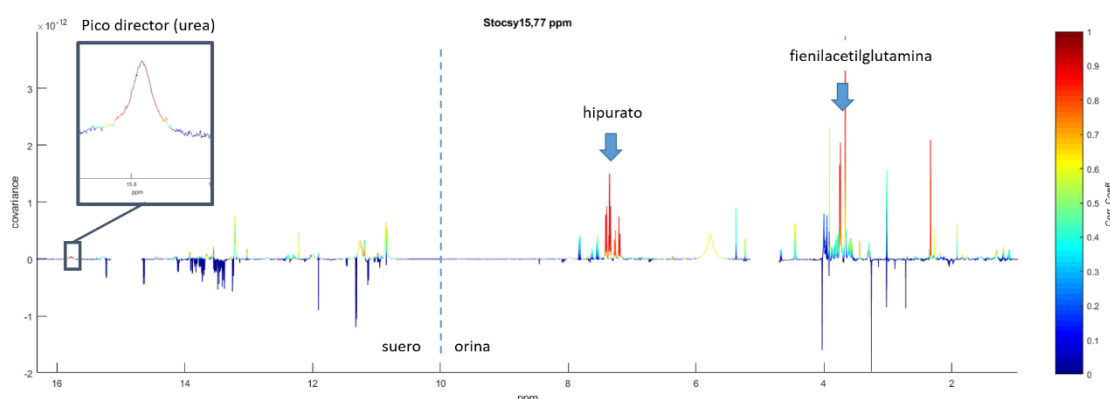


Figura 50. Ejemplo de gráfico de STOCSY concatenado datos de suero-orina para la señal de urea en suero a 5.7 ppm. A los efectos de construir el grafico se le sumaron 10 unidades al vector de corrimiento químico de suero al fusionar los datos. Se observan correlaciones de la urea en suero con señales de hipurato y fenilacetilglutamina en orina.

Estos resultados son concordantes con reportes bibliográficos que describen las posibles rutas de eliminación de la urea. Como se muestra en la **Figura 51**, el nitrógeno puede eliminarse por tres rutas posibles, incluyendo la de la urea, la del benzoato, o la del fenilacetato [168,169].

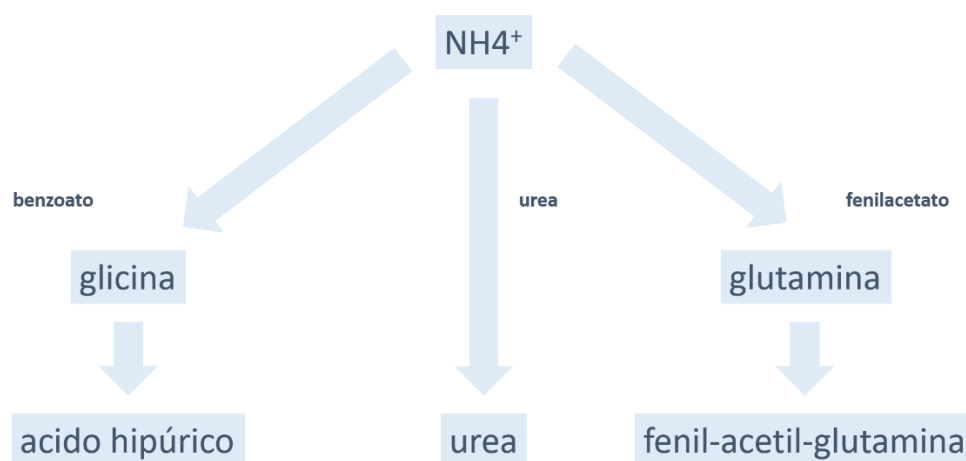


Figura 51. Esquema de rutas de eliminación del nitrógeno.

4.4 Conclusiones y Perspectivas

Como se demostró en la primera sección de este capítulo, la estrategia JIBA de alineación de espectros HOMO 2DJ permite que estos sean utilizados en esquemas de trabajo clásicos de análisis metabolómico. Al presentar una mejor resolución de las señales y minimizar la superposición de las mismas, el uso de datos de HOMO 2DJ alineados mediante JIBA nos permite obtener gráficos de valores de cargas más definidos y fáciles de interpretar. Además, los modelos generados con estos datos permiten identificar señales discriminantes que no se aprecian al emplear datos de sus homólogos monodimensionales. Estos resultados, junto con otras estrategias de identificación de metabolitos reportadas en la bibliografía como COCOA-POD y el desarrollo de herramientas de búsqueda en bases de datos a partir de espectros bidimensionales [157,158], constituyen una poderosa combinación para el análisis metabolómico. La innovación en estas estrategias de análisis de datos facilita la comprensión de la

dinámica metabólica de diversos sistemas biológicos y mejora la identificación de metabolitos que resulten relevantes.

Existen varios aspectos a mejorar de la estrategia JIBA. Por un lado, el espectro de referencia podría ser seleccionado de forma inteligente (i.e. el espectro con mayor número de máximos en cada bloque $n \times i \times j$). Otra opción a evaluar es la de trabajar con cortes en planos en lugar de alinear vectores de pares. Esto permitiría mejorar el funcionamiento del algoritmo y optimizar el tiempo de procesado. Es relevante destacar que, tanto para espectros 1D como 2D, el tamaño espectral puede tener incidencia en cómo se desempeña icoshift, por lo cual también es un aspecto a evaluar. Por último, se podría investigar el desempeño de otros algoritmos de alineamiento (i.e. PAFFT, COW, etc).

Para abordar algunos de los problemas relacionados al ruido en T_1 que fueron identificados en este trabajo podemos encontrar dos alternativas. En primer lugar, podríamos buscar disminuirlo mejorando el preprocesado de datos empleado. Para esto se puede, por ejemplo, modificar el llenado con ceros o las funciones de corrección de ventana, entre otros [25,170,171]. Esta estrategia debería ser cuidadosamente evaluada a fin de evitar la introducción artefactos o distorsiones a los datos. Una segunda estrategia, y posiblemente más efectiva, sería modificar los parámetros de adquisición de datos para minimizar el ruido en T_1 . Para esto, por ejemplo, podríamos emplear tiempos de adquisición y de reciclado entre escaneos más largos [172]. Si bien este enfoque es el más prometedor, también puede implicar mayores tiempos de uso de equipo y tiempos de adquisición por muestra.

Por otra parte, se logró desarrollar una herramienta que permite la fusión de datos de RMN de diferentes matrices para realizar los análisis estadísticos comúnmente empleados en metabolómica. Estas herramientas nos brindan la capacidad de desarrollar modelos discriminantes más precisos, y explorar el metabolóma en forma más profunda y completa. Además, facilitan la identificación de correlaciones biológicas dentro del metabolismo y contribuyen significativamente a la elucidación de los mecanismos y rutas metabólicas intrínsecas de un organismo. Finalmente, estas estrategias no se limitan solamente a la concatenación de datos de RMN, sino que pueden aplicarse a la fusión de datos de múltiples plataformas analíticas, como MS y RMN.

Referencias

1. Lindon, C.; Nicholson, J. K.; Holmes, E. *The Handbook of Metabonomics and Metabolomics*. Imperial College, London, **2007**.
2. Worley, B.; Powers, R. Multivariate Analysis in Metabolomics. *Curr. Metabolomics* **2013**, 1, 92-107.
3. Verpoorte, R.; Choi, Y. H.; Mustafa, N. R.; Kim, H. K. Metabolomics: Back to basics. *Phytochem. Rev.* **2008**, 7, 525-537.
4. Pauli, W. Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren. *Z. Physik.* **1925**, 31, 765-783.
5. Frisch, R.; Stern, O. Über diemagnetische Ablenkung von Wasserstoffmolekülen und das magnetische Moment des Protons. *I.Z. Phys.* **1933**, 85, 4-16.
6. Rabi, I.; Millman, S.; Kusch, P.; Zacharias, J. The Molecular Beam Resonance Method for Measuring Nuclear Magnetic Moments. The Magnetic Moments of ${}^6\text{Li}$, ${}^7\text{Li}$ and ${}^{19}\text{F}$. *Phys. Rev.* **1939**, 55, 526-535.
7. Purcell, E. M.; Torrey, H.; Pound, R. V. Resonance absorption by nuclear magnetic moments in a solid. *Phys. Rev.* **1946**, 69, 3-7.
8. Bloch, F. Nuclear Induction. *Phys. Rev.* **1946**, 70, 460-474.
9. Hahn, E. L. Nuclear Induction Due to Free Larmor Precession. *Phys. Rev.* **1950**, 77, 297-298.

10. Gutowsky, H.; Hoffman, C. Chemical Shifts in the Magnetic Resonance of F^{19} . *Phys. Rev.* **1950**, *80*, 110-111.
11. Ernst, R. R.; Anderson, W. A. Application of Fourier Transform Spectroscopy to Magnetic Resonance. *Rev. Sci. Instr.* **1966**, *37*, 93-102.
12. Jeener, J.; Alewaeters, G. "Pulse pair technique in high resolution NMR" a reprint of the historical 1971 lecture notes on two-dimensional spectroscopy. *Prog Nucl. Magn. Reson. Spectrosc.* **2016**, *94-95*, 75-80.
13. Aue, W. P.; Bartholdi, E.; Ernst, R. R. Two-dimensional spectroscopy. Application to nuclear magnetic resonance. *J. Chem. Phys.* **1976**, *64*, 2229-2246.
14. Wüthrich, K. Nuclear magnetic resonance in protein research. *Experientia* **1974**, *30*, 577-585.
15. Hoult, D. I.; Busby, S. J.; Gadian, D. G.; Radda, G. K.; Richards, R. E.; Seeley, P. J. Observation of tissue metabolites using ^{31}P nuclear magnetic resonance. *Nature* **1974**, *252*, 285-287.
16. Bales, J. R.; Higham, D. P.; Howe, I.; Nicholson, J. K.; Sadler, P. J. Use of high resolution proton nuclear magnetic resonance spectroscopy for rapid multicomponent analysis of urine. *Clin. Chem.* **1984**, *30*, 426-432.
17. Nicholson, J. K.; Lindon, J. C.; Holmes, E. 'Metabonomics': Understanding the Metabolic Responses of Living Systems to Pathophysiological Stimuli Via Multivariate Statistical Analysis of Biological NMR Spectroscopic Data. *Xenobiotica* **1999**, *29*, 1181-1189.

18. Lindon, J. C.; Nicholson, J. K.; Holmes, E.; Everett, J. R. Metabonomics: Metabolic processes studied by NMR spectroscopy of biofluids. *Concepts Magn. Reson.* **2000**, *12*, 289-320.
19. Nicholson, J. K. Molecular Phenomic Approaches to Deconvolving the Systemic Effects of SARS-CoV-2 Infection and Post-acute COVID-19 Syndrome. *Phenomix* **2021**, *1*, 143-150.
20. Girsztowt, A. The book collection of the Loitz family in the Elbląg Library: current state of research, overview of the collection, and a supplement to the list of prints. *Porta Aurea* **2022**, 279-292.
21. Baker, M. Metabolomics: from small molecules to big ideas. *Nat. Methods* **2011**, *8*, 117-121.
22. Dunn, W. B.; David I. E. Metabolomics: Current analytical platforms and methodologies. *Trends Anal. Chem.* **2005**, *24*, 285-294.
23. Díaz-Galiano, F. J.; Murcia-Morales, M.; Fernández-Alba, A. R. From sound check to encore: A journey through high-resolution mass spectrometry-based food analyses and metabolomics. *Compr. Rev. Food Sci. Food Saf.* **2024**, *23*, e13325.
24. Sheberstov, K. F.; Sistaré Guardiola, E.; Jeannerat, D. Everything you wanted to know about phase and reference frequency in one- and two-dimensional NMR spectroscopy. *Magn. Reson. Chem.* **2020**, *58*, 376-389.
25. Boden, N.; Fukushima, E.; Roeder, S. B. W. (1982). *Experimental pulse NMR: A nuts and bolts approach*. Addison-Wesley. Biochemistry Education, *Biochem. Educ.* **1982**

26. Vu, T. N.; Laukens, K. Getting your peaks in line: a review of alignment methods for NMR spectral data. *Metabolites* **2013**, 3, 259-276.
27. Savorani, F.; Tomasi, G.; Engelsen, S. B. icoshift: A versatile tool for the rapid alignment of 1D NMR spectra. *J Magn. Reson.* **2010**, 202, 190-202.
28. Dieterle, F.; Ross, A.; Schlotterbeck, G.; Senn, H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in ¹H NMR metabonomics. *Anal. Chem.* **2006**, 78, 4281-4290.
29. Van den Berg, R. A.; Hoefsloot, H. C.; Westerhuis, J. A.; Smilde, A.K.; van der Werf, M. J. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* **2006**, 7, 142-156.
30. Murtagh, F.; Contreras, P. Algorithms for hierarchical clustering: An overview. *WIREs Data Min. Knowl.* **2012**, 2, 86-97.
31. Jolliffe, I. T.; Cadima, J. Principal component analysis: a review and recent developments. *Phil. Trans. R. Soc.* **2016**, A374: 20150202.
32. Cunningham, P.; Delany, S. k-Nearest neighbour classifiers. *Mult. Classif. Syst.* **2007**, 54, 1-16.
33. Bylesjo, M.; Rantalainen, M.; Cloarec, O.; Nicholson, J. K.; Holmes, E.; Trygg, J. OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *J. Chemom.* **2006**, 20, 341-351.
34. Evgeniou, T.; Pontil, M., Support Vector Machines: Theory and Applications. *MLWA*. **2001**, 2049, 249-257.

35. Graupe, D. *Principles of Artificial Neural Networks* (2nd ed.). World Scientific Publishing Company, 2019.
36. Mishra, S.; Sarkar, U.; Taraphder, S.; Datta, S.; Swain, D.; Saikhom, R.; Panda, S.; Laishram, M., Principal Component Analysis. *Int. J. Livest. Res.* **2017**, 7, 60-78.
37. Worley, B.; Powers, R. Multivariate Analysis in Metabolomics. *Curr. Metabolomics* **2013**, 1, 92-107.
38. Trygg, J.; Holmes, E.; Lundstedt, T.; Chemometrics in metabonomics. *J Proteome Res.* **2007**, 6, 469-79.
39. Geladi, P.; Kowalski, B. R., Partial least-squares regression: a tutorial. *Anal. Chim. Acta* **1986**, 185, 1-17.
40. Haaland, D. M.; Thomas, E. V., Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. *Anal. Chem.* **1988**, 60, 1193-1202.
41. Gramatica, P. On the Development and Validation of QSAR Models. *Methods mol. Biol.* **2013**, 930, 499-526.
42. Bylesjo, M.; Rantalainen, M.; Cloarec, O.; Nicholson, J.; Holmes, E.; Trygg, J., OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *J. Chemom.* **2006**, 20, 341-351.
43. Jr., H.; Ringle, C.; Sarstedt, M. PLS-sem: Indeed a silver bullet. *J. Mark. Theory Pract.* **2011**, 19, 139-151.
44. Eriksson, L.; Trygg, J.; Wold, S., CV-ANOVA for significance testing of PLS and OPLS® models. *J. Chemom.* **2008**, 22, 594-600.

45. Triba, M. N.; Le Moyec, L.; Amathieu, R.; Goossens, C.; Bouchemal, N.; Nahon, P.; Rutledge, D. N.; Savarin, P. PLS/OPLS models in metabolomics: the impact of permutation of dataset rows on the K-fold cross-validation quality parameters. *Mol. Biosyst.* **2015**, *11*, 13-19.
46. Florkowski, C. M. Sensitivity, specificity, receiver-operating characteristic (ROC) curves and likelihood ratios: communicating the performance of diagnostic tests. *Clin. Biochem. Rev.* **2008**, *29*, 83-87.
47. Cloarec, O.; Dumas, M. E.; Craig, A.; Barton, R. H.; Trygg, J.; Hudson, J.; Blancher, C.; Gauguier, D.; Lindon, J. C.; Holmes, E.; Nicholson, J. K., Statistical Total Correlation Spectroscopy: An Exploratory Approach for Latent Biomarker Identification from Metabolic ¹H NMR Data Sets. *Anal. Chem.* **2005**, *77*, 1282-1289.
48. Dona, A. C.; Kyriakides, M.; Scott, F.; Shephard, E. A.; Varshavi, D.; Veselkov, K.; Everett, J. R., A guide to the identification of metabolites in NMR-based metabonomics / metabolomics experiments *Comput. Struct. Biotechnol. J.* **2016**, *14*, 135-153.
49. Johnson, S. R.; Lange, B. M., Open-Access Metabolomics Databases for Natural Product Research: Present Capabilities and Future Potential. *Front. Bioeng. Biotechnol.* **2015**, *3*, 1-10.
50. Grummer, R. R. Impact of changes in organic nutrient metabolism on feeding the transition dairy cow. *J. Anim. Sci.* **1995**, *73*, 2820-2833.

51. Bauman, D. E.; Currie, W. B. Partitioning of nutrients during pregnancy and lactation: A review of mechanisms involving homeostasis and homeorhesis. *J. Dairy Sci.* **1980**, *63*, 1514–1529.
52. Ingvarlsen, K. L. Feeding- and management-related diseases in the transition cow: Physiological adaptations around calving and strategies to reduce feeding-related diseases. *Anim. Feed Sci. Technol.* **2006**, *126*, 175–213.
53. Butler, W. R. Nutritional interactions with reproductive performance in dairy cattle. *Anim. Reprod. Sci.* **2000**, *60–61*, 449–457.
54. Drackley, J. K.; Overton, T. R.; Douglas, G. N. Adaptations of glucose and long-chain fatty acid metabolism in liver of dairy cows during the periparturient period. *J. Dairy Sci.* **2001**, *84*, E100–E112.
55. Meikle, A.; Kulcsar, M.; Chilliard, Y.; Febel, H.; Delavaud, C.; Cavestany, D.; Chilbroste, P. Effects of parity and body condition at parturition on endocrine and reproductive parameters of the cow. *Reproduction* **2004**, *127*, 727–737.
56. Ingvarlsen, K. L.; Moyes, K. Nutrition, immune function and health of dairy cattle. *Animal* **2012**, *7*, 112–122.
57. Mulligan, F. J.; Doherty, M. L. Production diseases of the transition cow. *Vet. J.* **2008**, *176*, 3–9.
58. Ingvarlsen, K. L.; Dewhurst, R. J.; Friggens, N. C. On the relationship between lactational performance and health: Is it yield or metabolic

- imbalance that cause production diseases in dairy cattle? A position paper. *Livest. Prod. Sci.* **2003**, *83*, 277–308.
59. Oltenacu, P. A.; Broom, D. M. The impact of genetic selection for increased milk yield on the welfare of dairy cows. *Anim. Welf.* **2010**, *19*, 39–49.
 60. Ingvarlsen, K. L.; Andersen, J. B. Integration of metabolism and intake regulation: A review focusing on periparturient animals. *J. Dairy Sci.* **2000**, *83*, 1573–1597.
 61. Burke, C.; Roche, R. Effects of pasture feeding during the periparturient period on postpartum anovulation in grazed dairy cows. *J. Dairy Sci.* **2007**, *90*, 4304–4312.
 62. Enjalbert, F.; Nicot, M. C.; Bayourthe, C.; Moncoulon, R. Ketone bodies in milk and blood of dairy cows: Relationship between concentrations and utilization for detection of subclinical ketosis. *J. Dairy Sci.* **2001**, *84*, 583–589.
 63. Giannesella, M.; Perillo, L.; Fiori, E.; Giudice, E.; Zumbo, A.; Morgante, M.; Piccione, G. Transition period in healthy and diseased dairy cows: Evaluation of metabolic modifications. *Large Anim. Rev.* **2018**, *24*, 107–111.
 64. Ceciliani, F.; Lecchi, C.; Urh, C.; Sauerwein, H. Proteomics and metabolomics characterizing the pathophysiology of adaptive reactions to the metabolic challenges during the transition from late pregnancy to early lactation in dairy cows. *J. Proteome Res.* **2018**, *178*, 92–106.

65. Zhang, H.; Wu, L.; Xu, C.; Xia, C.; Sun, L.; Shu, S. Plasma metabolomic profiling of dairy cows affected with ketosis using gas chromatography/mass spectrometry. *BMC Vet. Res.* **2013**, *9*, 186–199.
66. Imhasly, S.; Naegeli, H.; Baumann, S.; von Bergen, M.; Luch, A.; Jungnickel, H.; Potratz, S.; Gerspach, C. Metabolomic biomarkers correlating with hepatic lipidosis in dairy cows. *BMC Vet. Res.* **2014**, *10*, 122-130.
67. Dervishi, E.; Zhang, G.; Dunn, S. M.; Mandal, R.; Wishart, D. S.; Ametaj, B. N. GC-MS metabolomics identifies metabolite alterations that precede subclinical mastitis in the blood of transition dairy cows. *J. Proteome Res.* **2017**, *16*, 433–446.
68. Zhang, G.; Dervishi, E.; Dunn, S. M.; Mandal, R.; Liu, P.; Han, B.; Wishart, D. S.; Ametaj, B. N. Metabotyping reveals distinct metabolic alterations in ketotic cows and identifies early predictive serum biomarkers for the risk of disease. *Metabolomics* **2017**, *13*, 43–58.
69. Zhang, G.; Ametaj, B. N. Ketosis: An old story under a new approach. *Dairy* **2020**, *1*, 42–60.
70. Sun, L. W.; Zhang, H. Y.; Wu, L.; Shu, S.; Xia, C.; Xu, C.; Zheng, J. S. ¹H-Nuclear magnetic resonance-based plasma metabolic profiling of dairy cows with clinical and subclinical ketosis. *J. Dairy Sci.* **2014**, *97*, 1552–1562.
71. Xu, C.; Sun, L.-W.; Xia, C.; Zhang, H.-Y.; Zheng, J.-S.; Wang, J.-S. ¹H-Nuclear magnetic resonance-based plasma metabolic profiling of dairy cows with fatty liver. *Asian Australas. J. Anim. Sci.* **2016**, *29*, 219–229.

72. Dervishi, E.; Zhang, G.; Hailemariam, D.; Mandal, R.; Wishart, D. S.; Ametaj, B. N. Urine metabolic fingerprinting can be used to predict the risk of metritis and highlight the pathobiology of the disease in dairy cows. *Metabolomics* **2018**, *14*, 1-15.
73. Zhang, G.; Dervishi, E.; Zwierzchowski, G.; Mandal, R.; Wishart, D. S.; Ametaj, B. N. Urinary metabolomics around parturition identifies metabolite alterations in dairy cows affected postpartum by lameness: Preliminary study. *Diary* **2020**, *1*, 6-19.
74. Hailemariam, D.; Mandal, R.; Saleem, F.; Dunn, S. M.; Wishart, D. S.; Ametaj, B. N. Identification of predictive biomarkers of disease state in transition dairy cows. *J. Dairy Sci.* **2014**, *97*, 2680-2693.
75. Hailemariam, D.; Mandal, R.; Saleem, F.; Dunn, S. M.; Wishart, D. S.; Ametaj, B. N. Metabolomics approach reveals altered amino acid and sphingolipid profiles associated with pathological state in transition dairy cows. *Curr. Metab.* **2014**, *2*, 184-195.
76. Kenéz, Á.; Dänicke, S.; Rolle-Kampczyk, U.; von Bergen, M.; Huber, K. A metabolomics approach to characterize phenotypes of metabolic transition from late pregnancy to early lactation in dairy cows. *Metabolomics* **2016**, *12*, 165-176.
77. Luo, Z. Z.; Shen, L. H.; Jiang, J.; Huang, Y. X.; Bai, L. P.; Yu, S. M.; Yao, X. P.; Ren, Z. H.; Yang, Y. X.; Cao, S. Z. Plasma metabolite changes in dairy cows during parturition identified using untargeted metabolomics. *J. Dairy Sci.* **2019**, *102*, 1-12.

78. Luke, T. D. W.; Pryce, J. E.; Elkins, A. C.; Wales, W. J.; Rochfort, S. J. Use of large and diverse datasets for ^1H NMR serum metabolic profiling of early lactation dairy cows. *Metabolites* **2019**, *10*, 180-200.
79. Fariña, S. R.; Chillibroste, P. Opportunities and challenges for the growth of milk production from pasture: The case of farm systems in Uruguay. *Agric. Syst.* **2019**, *176*, 102631.
80. Kolver, E. S.; Muller, L. D. Performance and nutrient intake of high producing Holstein cows consuming pasture or a total mixed ration. *J. Dairy Sci.* **1998**, *81*, 1403-1411.
81. Duffield, T. F.; Leslie, K. E.; Sandals, D.; Lissemore, K.; McBride, B. W.; Lumsden, J. H.; Dick, P.; Bagg, R. Effect of prepartum administration of a monensin controlled release capsule on cow health and reproduction. *J. Dairy Sci.* **1999**, *82*, 2377-2384.
82. LeBlanc, S. Postpartum uterine disease and dairy herd reproductive performance: A review. *Vet. J.* **2008**, *176*, 102-114.
83. Galvão, K. N.; Greco, L. F.; Vilela, J. M.; Sá Filho, M. F.; Santos, J. E. Effect of intrauterine infusion of ceftiofur on uterine health and fertility in dairy cows. *J. Dairy Sci.* **2009**, *92*, 1532-1542.
84. Ruegg, P. L. New perspectives in udder health management. *Vet. Clin. Food Anim.* **2012**, *28*, 149-163.
85. Duffield, T. F.; Lissemore, K. D.; McBride, B. W.; Leslie, K. E. Impact of hyperketonemia in early lactation dairy cows on health and production. *J. Dairy Sci.* **2009**, *92*, 571-580.

86. Nicholson, J. K.; Foxall, P. J. D.; Spraul, M.; Farrant, R. D.; Lindon, J. C. 750 MHz ^1H and ^{13}C NMR spectroscopy of human blood plasma. *Anal. Chem.* **1995**, 67, 793–811.
87. Viant, M. R. Improved methods for the acquisition and interpretation of NMR metabolomic data. *Biochem. Biophys. Res. Commun.* **2003**, 310, 943–948.
88. Simundic, A.-M. Diagnostic accuracy–Part 1. Basic concepts: Sensitivity and specificity, ROC analysis, STARD statement. *Point Care* **2012**, 11, 6–8.
89. Ekelund, S. ROC curves–What are they and how are they used? *Point Care* **2012**, 11, 16–21.
90. Ni, Y.; Su, M.; Lin, J.; Wang, X.; Qiu, Y.; Zhao, A.; Chen, T.; Jia, W. Metabolic profiling reveals disorder of amino acid metabolism in four brain regions from a rat model of chronic unpredictable mild stress. *FEBS Lett.* **2008**, 582, 2627–2636.
91. Sun, H. Z.; Wang, D. M.; Wang, B.; Wang, J. K.; Liu, H. Y.; Guan, L. L.; Liu, J. X. Metabolomics of four biofluids from dairy cows: Potential biomarkers for milk production and quality. *J. Proteome Res.* **2015**, 14, 1287–1298.
92. Thornalley, P.J. Modification of the glyoxalase system in human red blood cells by glucose in vitro. *Biochem. J.* **1988**, 254, 751–755.
93. Omole, O. O.; Nappert, G.; Naylor, J.M.; Zello, G. A. Both L- and D-lactate contribute to metabolic acidosis in diarrheic calves. *J. Nutr.* **2001**, 131, 2128–2131.

94. Sako, T.; Urabe, S.; Kusaba, A.; Kimura, N.; Yoshimura, I.; Tazaki, H.; Imai, S.; Ono, K.; Arai, T. Comparison of plasma metabolite concentrations and lactate dehydrogenase activity in dogs, cats, horses, cattle and sheep. *Vet. Res. Commun.* **2007**, *31*, 413–417.
95. Krauze, K. M.; Oetzel, G. R. Understanding and preventing subacute ruminal acidosis in dairy herds: A review. *Anim. Feed Sci. Technol.* **2006**, *126*, 215–236.
96. Asanuma, N.; Hino, T. Regulation of fermentation in a ruminal bacterium, *Streptococcus bovis*, with special reference to rumen acidosis. *Anim. Sci. J.* **2002**, *73*, 313–325.
97. Overton, T. R.; Waldron, M. R. Nutritional management of transition dairy cows: Strategies to optimize metabolic health. *J. Dairy Sci.* **2004**, *87*, E105–E119.
98. Aschenbach, J. R.; Kristensen, N. B.; Donkin, S. S.; Hammon, H. M.; Penner, B. G. Gluconeogenesis in dairy cows: The secret of making sweet milk from sour dough. *IUBMB Life* **2010**, *62*, 869–877.
99. Gonzáles, F. D.; Muiño, R.; Pereira, V.; Campos, R.; Bendito, J.L. Relationship among blood indicators of lipomobilization and hepatic function during early lactation in high-yielding dairy cows. *J. Vet. Sci.* **2011**, *12*, 251–255.
100. Vercoe, J. E. Urinary allantoin excretion and digestible dry-matter intake in cattle and buffalo. *J. Agric. Sci.* **1976**, *86*, 613–615.

101. Koch, F.; Lamp, O.; Eslamizad, M.; Weitzel, J.; Kuhla, B. Metabolic response to heat stress in late-pregnant and early lactation dairy cows: Implications to liver-muscle crosstalk. *PLoS ONE* **2016**, *11*, e0160912.
102. Bell, A. W. Regulation of organic nutrient metabolism during transition from late pregnancy to early lactation. *J. Anim. Sci.* **1995**, *73*, 2804–2819.
103. Hayirli, A.; Grummer, R. R.; Nordheim, E. V.; Crump, P. M. Animal and dietary factors affecting feed intake during the prefresh transition period in Holsteins. *J. Dairy Sci.* **2002**, *85*, 3430–3443.
104. Grant, R. J.; Albright, J. L. Effect of animal grouping on feeding behavior and intake of dairy cattle. *J. Dairy Sci.* **2001**, *85*, E156–E163.
105. Meikle, A.; de Brun, V.; Carriquiry, M.; Soca, P.; Sosa, C.; Adrien, M.L.; Chilibraste, P.; Abecia, J.A. Influences of nutrition and metabolism on reproduction of the female ruminant. *Anim. Reprod.* **2018**, *15*, 899–911.
106. Kim, H. S.; Kim, E. D. T.; Eom, J. S.; Choi, Y. Y.; Lee, S. J.; Lee, S. S.; Chung, C. D.; Lee, S. S. Exploration of metabolite profiles in the biofluids of dairy cows by proton nuclear magnetic resonance analysis. *PLoS ONE* **2021**, *16*, e0246290.
107. Akar, T.; Avci, M.; Dusunceli, F. *BARLEY: Post-Harvest Operations (2004)*. <https://www.fao.org/3/au997e/au997e.pdf>.
108. Ullrich SE. Barley: production, improvement, and uses. *John Wiley & Sons*, **2010**.
109. Busa, L.; Bertins, M.; Viksna, A.; Legzdina, L.; Kobzarevs, D., Evaluation of carbon, nitrogen, and oxygen isotope ratio measurement data for

- characterization of organically and conventionally cultivated spring barley (*Hordeum vulgare* L.) grain. *Agronomy Research* **2021**, 19, 1364-1372.
110. Gozukirmizi, N.; Karlik, E. Barley (*Hordeum vulgare* L.) Improvement Past, Present and Future. *Brewing Technology*, InTechOpen. **2017**
 111. Bian, J.; Deng, P.; Zhan, H.; Wu, X.; Nishantha, M.; Yan, Z.; Du, X.; Nie, X.; Song, W., Transcriptional Dynamics of Grain Development in Barley (*Hordeum vulgare* L.). *Int. J. Mol. Sci.* **2019**, 20, 1-16.
 112. Ribeiro, R. V.; Machado, R. S.; Machado, E. C.; Machado, D. F. S. P.; Magalhaes Filho, J. R.; Landell, M. G. A., Revealing drought-resistance and productive patterns in sugarcane genotypes by evaluating both physiological responses and stalk. *Exp. Agric.* **2013**, 49, 212-224.
 113. Dughetti, A. **2012**. Pulgones: Clave para identificar las formas ápteras que atacan a los cereales. RIAN, *Red de información agropecuaria nacional*. INTA y Ministerio de Agricultura, Ganadería y Pesca. 44
 114. Salvadori, J. R.; Lau, D.; Valle da Silva, P. R. **2009**. Cultivo de trigo; pragas e métodos de controle. Brasil, EMBRAPA. *Sistemas de producao* 4. sp.
 115. Bentancourt, C. **2004**. *Manual de entomología*. Uruguay: Universidad de la República, Facultad de Agronomía
 116. Kaleem Ullah RM, Gao F, Sikandar A, Wu H. Insights into the Effects of Insecticides on Aphids (Hemiptera: Aphididae): Resistance Mechanisms and Molecular Basis. *Int. J. Mol. Sci.* **2023**, 24, 6750.
 117. Bass, C.; Nauen, R. The molecular mechanisms of insecticide resistance in aphid crop pests. *Insect Biochem. Mol. Biol.* **2023**, 156, 103937.

118. FAO (Organización de las Naciones Unidas para la Agricultura y la Alimentación, IT). World Food and Agriculture, *Statistical pocketbook*. Rome, **2022**.
119. Painter, R. H. *Insect Resistance in Crop Plants*. The McMillan Co, New York, **1951**.
120. Razzaq, A.; Sadia, B.; Raza, A.; Khalid Hameed, M.; Saleem, F.; Metabolomics: A Way Forward for Crop Improvement. *Metabolites* **2019**, 9, 1-37.
121. Leiss, K. A.; Choi, Y. H.; Verpoorte, R.; Klinkhamer, P. G. An overview of NMR-based metabolomics to identify secondary plant compounds involved in host plant resistance. *Phytochem Rev.* **2011**, 10, 205-216.
122. Sanchez-Arcos, C.; Kai, M.; Svatoš, A.; Gershenzon, J.; Kunert, G., Untargeted Metabolomics Approach Reveals Differences in Host Plant Chemistry Before and After Infestation With Different Pea Aphid Host Races. *Front. Plant Sci.* **2019**, 10, 1-13.
123. Ma, N. L.; Lah, W. A. C.; Kadir, N. A.; Mustaqim, M.; Rahmat, Z.; Ahmad, A.; Lam, S. D.; Ismail, M. R. Susceptibility and tolerance of rice crop to salt threat: Physiological and metabolic inspections. *PLoS ONE* **2018**, 13, 1-17.
124. Sanchez, D. H.; Siahpoosh, M. R.; Roessner, U.; Udvardi, M.; Kopka, J. Plant metabolomics reveals conserved and divergent metabolic responses to salinity. *Physiol. Plant.* **2008**, 132, 209-219.
125. Litvinov, D. Y.; Karlov, G. I.; Divashuk, M. G. Metabolomics for Crop Breeding: General Considerations. *Genes* **2021**, 12, 1-24.

126. Seong, G.-U.; Yun, D.-Y.; Shin, D.-H.; Cho, J.-S.; Lee, G.; Choi, J. H.; Park, K.-J.; Ku, K.-H.; Lim, J.-H. Comparative ^1H NMR-Based Metabolomics of Traditional Landrace and Disease-Resistant Chili Peppers (*Capsicum annuum* L.). *Foods* **2024**, *13*, 1-16.
127. Copley, T. R.; Aliferis, K. A.; Kliebenstein, D. J.; Jabaji, S. H. An integrated RNAseq- ^1H NMR metabolomics approach to understand soybean primary metabolism regulation in response to Rhizoctonia foliar blight disease. *BMC Plant Biol.* **2017**, *17*, 1-18.
128. Razzaq, A.; Wishart, D. S.; Wani, S. H.; Hameed, M. K.; Mubin, M.; Saleem, F. Advances in Metabolomics-Driven Diagnostic Breeding and Crop Improvement. *Metabolites* **2022**, *12*, 1-22.
129. Manickam, S.; Rajagopalan, V.R.; Kambale, R.; Rajasekaran, R.; Kanagarajan, S.; Muthurajan, R. Plant Metabolomics: Current Initiatives and Future Prospects. *Curr. Issues Mol. Biol.* **2023**, *45*, 8894-8906.
130. Sakurai N. Recent applications of metabolomics in plant breeding. *Breed. Sci.* **2022**, *72*, 56-65.
131. Ramachandran, K. M.; Tsokos, C. P. Design of experiments. *Mathematical Statistics with Applications in R (Third Edition)* Academic Press, Hoboken, **2021**
132. Locatelli, A.; Cuesta-Marcos, A.; Gutiérrez, L.; Hayes, P.; Smith, K.; Castro, A., Genome-wide association mapping of agronomic traits in relevant barley in Uruguay. *Mol. Breed.* **2013**, *31*, 631-654.

133. Choi, H.-K.; Yoon, J.-H.; Kim, Y.-S.; Kwon, D. Y., Metabolomic profiling of Cheonggukjang during fermentation by ^1H NMR spectrometry and principal components analysis. *Process Biochem.* **2007**, *42*, 263-266.
134. Trygg, J.; Wold, S., Orthogonal projections to latent structures (O-PLS). *J. Chem.* **2002**, *16*, 119-128.
135. Wold, S.; Esbensen, K.; Geladi, P., Principal component analysis. *Chemometr. Intell. Lab. Syst.* **1987**, *2*, 37-52.
136. Peng, D. X.; Lai, F., Using partial least squares in operations management research: A practical guideline and summary of past research. *J. Oper. Manag.* **2012**, *30*, 467-480.
137. McNeil, S. D.; Nuccio, M. L.; Ziemak, M. J.; Hanson, A. D., Enhanced synthesis of choline and glycine betaine in transgenic tobacco plants that overexpress phosphoethanolamine *N*-methyltransferase. *PNAS* **2001**, *98*, 10001-10005.
138. Nuccio, M. L.; Russell, B. L.; Nolte, K. D.; Rathinasabapathi, B.; Gage, D. A.; Hanson, A. D., The endogenous choline supply limits glycine betaine synthesis in transgenic tobacco expressing choline monooxygenase. *Plant J.* **1998**, *16*, 487-96.
139. Annunziata, M. G.; Ciarmiello, L. F.; Woodrow, P.; Dell'Aversana, E.; Carillo, P., Spatial and Temporal Profile of Glycine Betaine Accumulation in Plants Under Abiotic Stresses. *Front. in Plant Sci.* **2019**, *10*, 1-13.
140. Tomaiuolo, M.; Bertram, R.; Houle, D., Enzyme isoforms may increase phenotypic robustness. *Evol. Int. J. Org. Evol.* **2008**, *62*, 2884-93.

141. Powell, G.; Tosh, C. R.; Hardie, J., Host Plant Sselection by Aphids: Behavioral, Evolutionary, and Applied Perspectives. *Annu. Rev. Entomol.* **2005**, *51*, 309-330.
142. Wu, J.; Lan, H.; Zhang, Z. F.; Cao, H. H.; Liu, T. X., Performance and Transcriptional Response of the Green Peach Aphid *Myzus persicae* to the Restriction of Dietary Amino Acids. *Front. Physiol.* **2020**, *11*, 487.
143. Dinant, S.; Bonnemain, J.-L.; Girusse, C.; Kehr, J., Phloem sap intricacy and interplay with aphid feeding. *C. R. Biol.* **2010**, *333*, 504-515.
144. Kansman, J. T.; Basu, S.; Casteel, C. L.; Crowder, D. W.; Lee, B. W.; Nihranz, C. T.; Finke, D. L. Plant Water Stress Reduces Aphid Performance: Exploring Mechanisms Driven by Water Stress Intensity. *Front. Ecol. Evol.* **2022**, *10*, 1-11.
145. Zúñiga, G. E.; Corcuera, L. J. Glycine-betaine accumulation influences susceptibility of water-stressed barley to the aphid *Schizaphis graminum*. *Phytochem.* **1987**, *26*, 367-369.
146. Michelmore, R. W.; Meyers, B. C., Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Gen. Res.* **1998**, *8*, 13-30.
147. Sharma, A.; Jones, J. B.; White, F. F., Recent advances in developing disease resistance in plants. *F1000Research* **8**, **2019**, 1934.
148. Munns, R.; James, R. A.; Läuchli, A., Approaches to increasing the salt tolerance of wheat and other cereals. *Journal of Experimental Botany*, **2006**, *57*, 1025-1043.

149. Castro, A. M.; Ramos, S.; Vasicek, A.; Worland, A.; Giménez, D.; Clúa, A. A.; Suárez, E., Identification of wheat chromosomes involved with different types of resistance against greenbug (*Schizaphis graminum*, Rond.) and the Russian wheat aphid (*Diuraphis noxia*, Mordvilko). *Euphytica* **2001**, 118, 321-330.
150. Wang, W.; Luo, L.; Lu, H. et al. Angiotensin-converting enzymes modulate aphid-plant interactions. *Sci. Rep.* **2015**, 5, 8885.
151. Carolan, J. C.; Fitzroy, C. I.; Ashton, P. D.; Douglas, A. E.; Wilkinson, T. L. The secreted salivary proteome of the pea aphid *Acyrtosiphon pisum* characterised by mass spectrometry. *Proteomics* **2009**, 9, 2457-2467.
152. Rahfeld, P.; Kirsch, R.; Kugel, S.; Wielsch, N.; Stock, M.; Groth, M.; Boland, W.; Burse, A.; Independently recruited oxidases from the glucose-methanol-choline oxidoreductase family enabled chemical defences in leaf beetle larvae (subtribe Chrysomelina) to evolve. *Proc. Biol. Sci.* **2014**, 281, 1-9.
153. Ludwig, C.; Viant, MR. Two-dimensional J-resolved NMR spectroscopy: review of a key methodology in the metabolomics toolbox. *Phytochem. Anal.* **2010**, 21, 22-32.
154. Verhoeven, A.; Slagboom, E.; Wuhrer, M.; Giera, M.; Mayboroda, O. A., Automated quantification of metabolites in blood-derived samples by NMR. *Anal. Chim. Acta.* **2017**, 976, 52-62.
155. Fonville, J. M.; Maher, A. D.; Coen, M.; Holmes, E.; Lindon, J. C.; Nicholson, J. K.; Evaluation of full-resolution J-resolved ¹H NMR

- projections of biofluids for metabonomics information retrieval and biomarker identification. *Anal. Chem.* **2010**, 82, 1811–1821.
156. Ludwig, C.; Viant, M. Two-Dimensional J-Resolved NMR Spectroscopy: Review of a Key Methodology in the Metabolomics Toolbox. *Phytochem. Anal.* **2010**, 21, 22-32.
 157. Charris-Molina, A.; Riquelme, G.; Burdisso, P.; Hoijemberg, P. A., Consecutive Queries to Assess Biological Correlation in NMR Metabolomics: Performance of Comprehensive Search of Multiplets over Typical 1D ^1H NMR Database Search. *J. Proteome Res.* **2020**, 19, 2977-2988.
 158. Charris-Molina, A.; Riquelme, G.; Burdisso, P.; Hoijemberg, P. A., Tackling the Peak Overlap Issue in NMR Metabolomics Studies: 1D Projected Correlation Traces from Statistical Correlation Analysis on Nontilted 2D ^1H NMR J-Resolved Spectra. *J. Proteome Res.* **2019**, 18, 2241-2253.
 159. Charris Molina, A.; Burdisso, P.; Hoijemberg, P. Multiplet-Assisted Peak Alignment for ^1H NMR-Based Metabolomics. *J. Proteome Res.* **2023**, 23, 430-448.
 160. Vu, T. N.; Laukens, K. Getting your peaks in line: a review of alignment methods for NMR spectral data. *Metabolites* **2013**, 3, 259–276.
 161. Savorani, F.; Tomasi, G.; Engelsen, S.B. icoshift: A versatile tool for the rapid alignment of 1D NMR spectra. *J. Magn. Reson.* **2010**, 202, 190-202.

162. Korifi, R.; Le Dréau, Y.; Dupuy, N. Comparative study of the alignment method on experimental and simulated chromatographic data. *J. Sep. Sci.* **2014**, *37*, 3276-91.
163. Marshall, D. D.; Powers, R. Beyond the paradigm: Combining mass spectrometry and nuclear magnetic resonance for metabolomics. *Prog. Nucl. Magn. Reson. Spectrosc.* **2017**, *100*, 1-16.
164. Monakhova, Y. B.; Godelmann, R.; Hermann, A.; Kuballa, T.; Cannet, C.; Schäfer, H.; Spraul, M.; Rutledge, D. N., Synergistic effect of the simultaneous chemometric analysis of ^1H NMR spectroscopic and stable isotope (SNIF-NMR, ^{18}O , ^{13}C) data: Application to wine analysis. *Anal. Chim. Acta* **2014**, *833*, 29-39.
165. Leal, A. L.; Silva, A. M. S.; Ribeiro, J. C.; Martins, F. G., Data driven models exploring the combination of NIR and ^1H NMR spectroscopies in the determination of gasoline properties. *Microchem. J.* **2022**, *175*, 107217.
166. Azcarate, S. M.; Ríos-Reina, R.; Amigo, J. M.; Goicoechea, H. C. Data handling in data fusion: Methodologies and applications. *Trends Anal. Chem.* **2021**, *143*, 116355.
167. Forshed, J.; Idborg, H.; Jacobsson, S. P., Evaluation of different techniques for data fusion of LC/MS and ^1H -NMR. *Chemometr. Intell. Lab. Syst.* **2007**, *85*, 102-109.
168. Redant, S.; Empain, A.; Mugisha, A.; Kamgang, P.; Attou, R.; Honoré, P. M.; De Bels, D., Management of late onset urea cycle disorders—a remaining challenge for the intensivist? *Ann. Intensive Care* **2021**, *11*, 1-10.

169. Sandlers, Y. Amino Acids Profiling for the Diagnosis of Metabolic Disorders, Bobbarala, V.; Sarwar Zaman, G.; Nasir Mohd Desa, M.; Md Akim, A. (Eds.) *Fundamentals of Medical and Laboratory Science, Clinical Biochemistry Clinical Correlation and Diagnosis*, IntechOpen **2019**.
170. Mo, H.; Harwood, J. S.; Yang, D.; Post, C. B. A simple method for NMR t_1 noise suppression. *J. Magn. Reson.* **2017**, 276, 43-50.
171. Manoleras, N.; Norton, R. S. Spectral processing methods for the removal of t_1 noise and solvent artifacts from NMR spectra. *J. Biomol. NMR* **1992**, 2, 485-494.
172. Giraudeau, P.; Akoka, S. A. New detection scheme for ultrafast 2D J-resolved spectroscopy. *J. Magn. Reson.* **2007**, 186, 352-357.

Anexo - Material Suplementario

Tabla S1. Dieta postparto.

Mes	Composición ^a			Oferta total
	Pelets de soja	Maiz ensilado	Pastura	
Febrero	5	2	11	18
Marzo	5	11	2	18
Abril	5	9	5	19
Mayo	6	11	2	19
Junio	6	11	2	19

^a Kg de materia seca por animal por día.

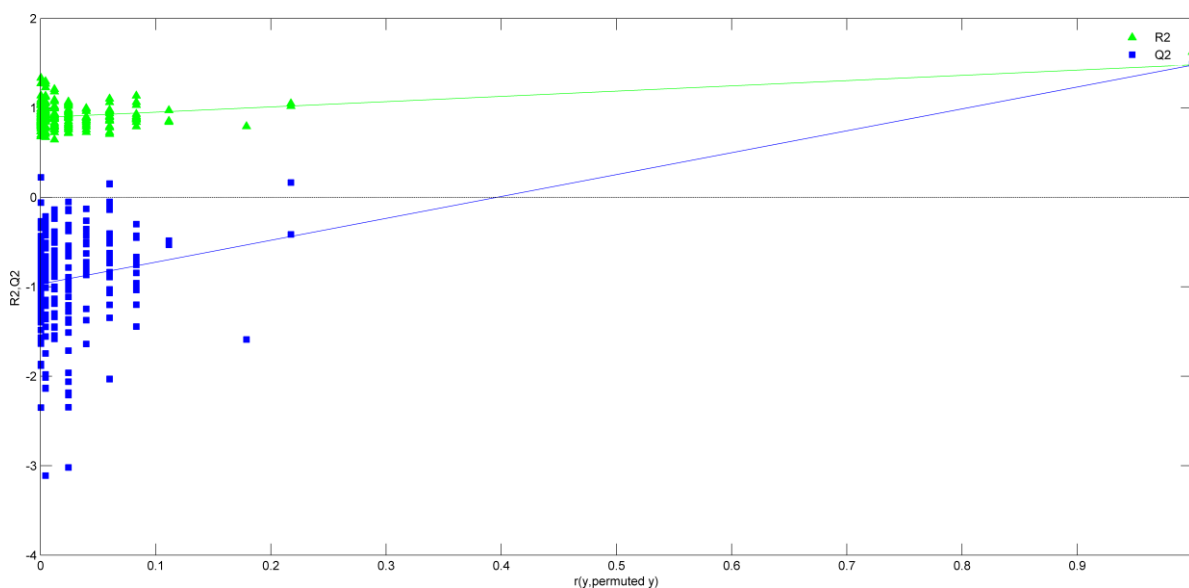


Figure S1. Gráficos de prueba de permutación para el modelo OPLS-DA comparando las etapas C-4 y C+1 ($R^2Y = 0.80$ and $Q^2Y = 0.72$).

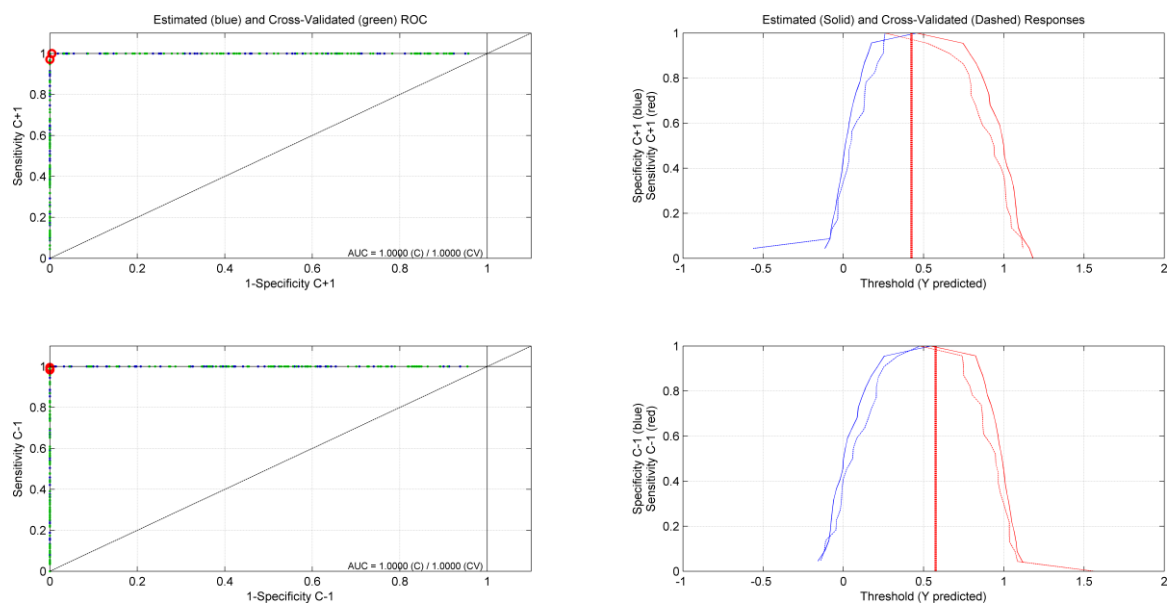


Figure S2. Curvas de análisis ROC derivadas del modelo OPLS-DA comparando los estadios C-4 y C+1 (AUC = 1.00).

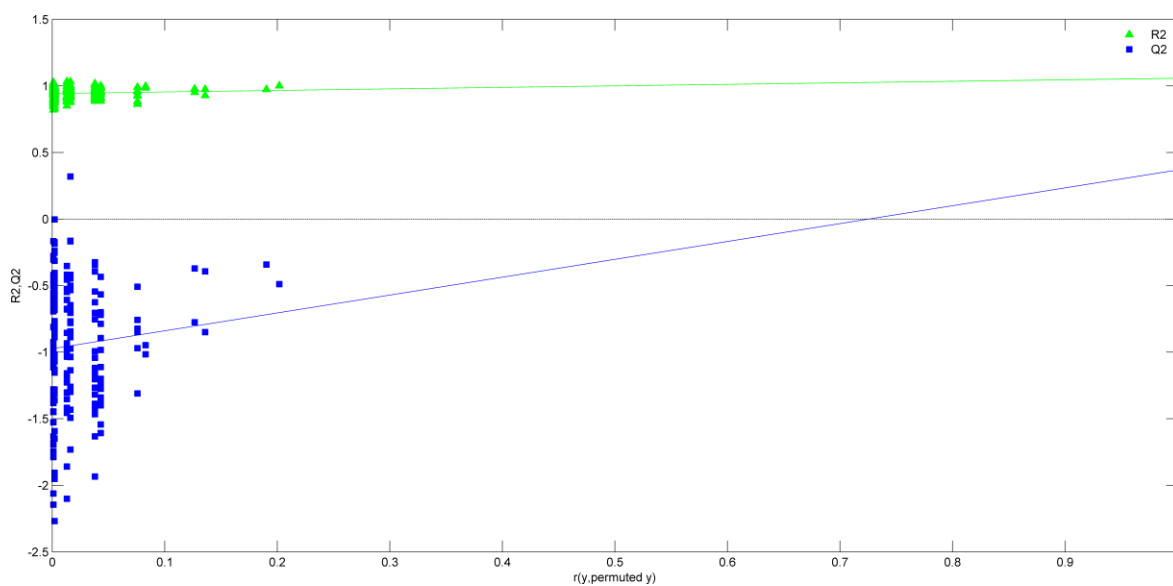


Figure S3. Gráficos de prueba de permutación para el modelo OPLS-DA comparando los estadios C-4 y C+4 ($R^2Y = 0.89$ and $Q^2Y = 0.37$).

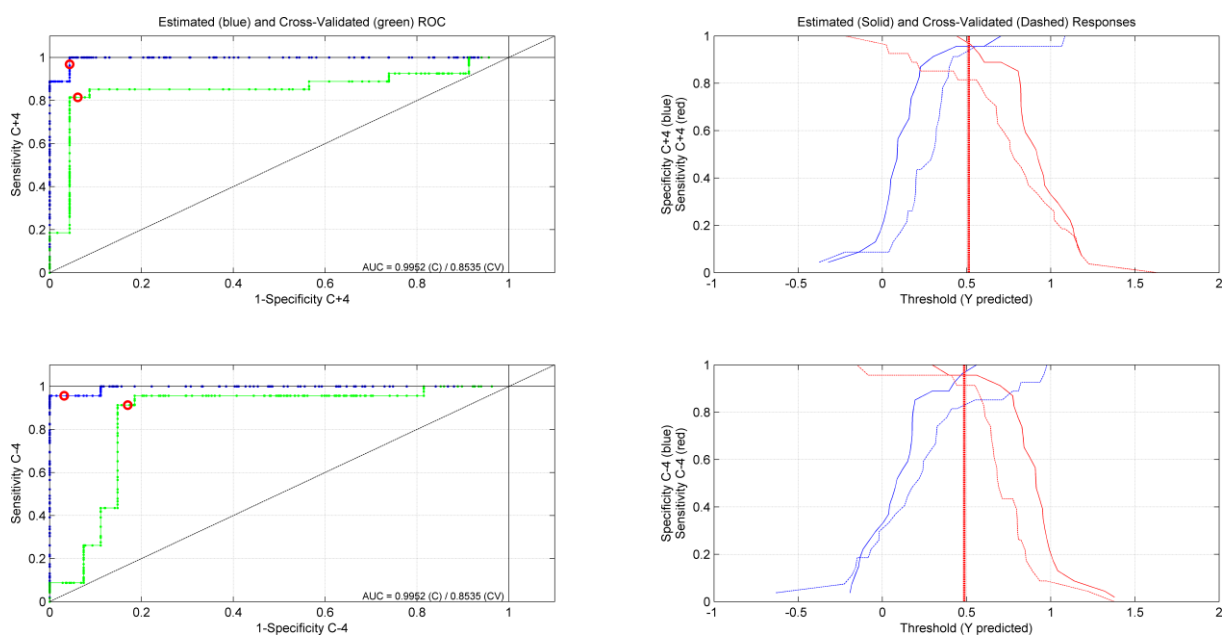


Figure S4. Curvas de análisis ROC derivadas del modelo OPLS-DA comparando los estadios C-4 y C+4 ($AUC = 0.85$).

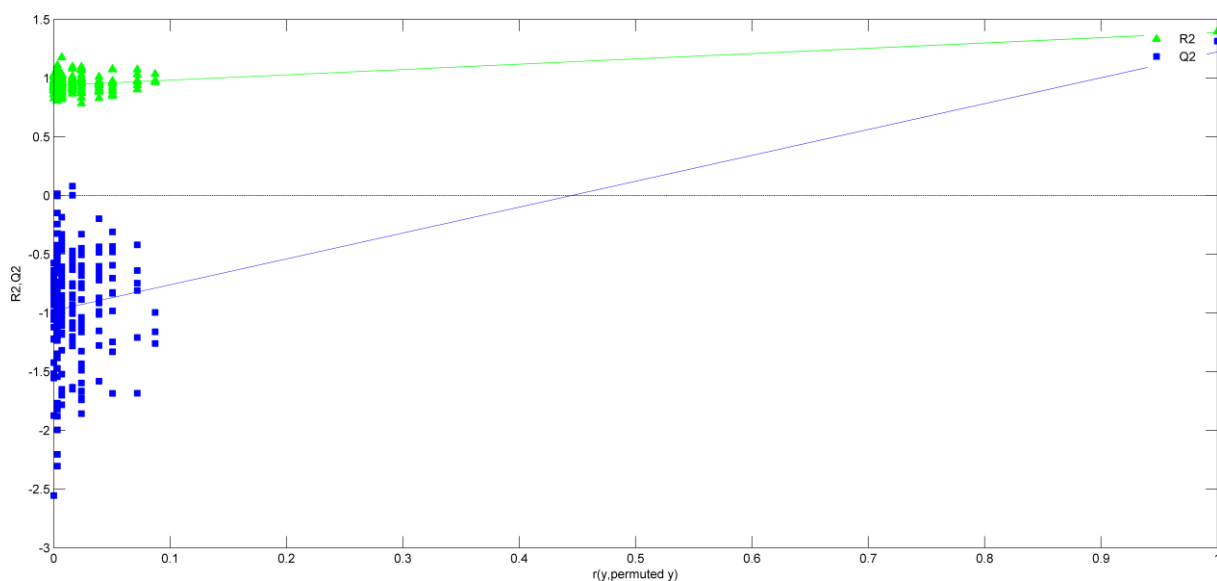


Figure S5. Gráficos de prueba de permutación para el modelo OPLS-DA comparando los estadios C-4 y C+8 ($R^2Y = 0.94$ and $Q^2Y = 0.87$).

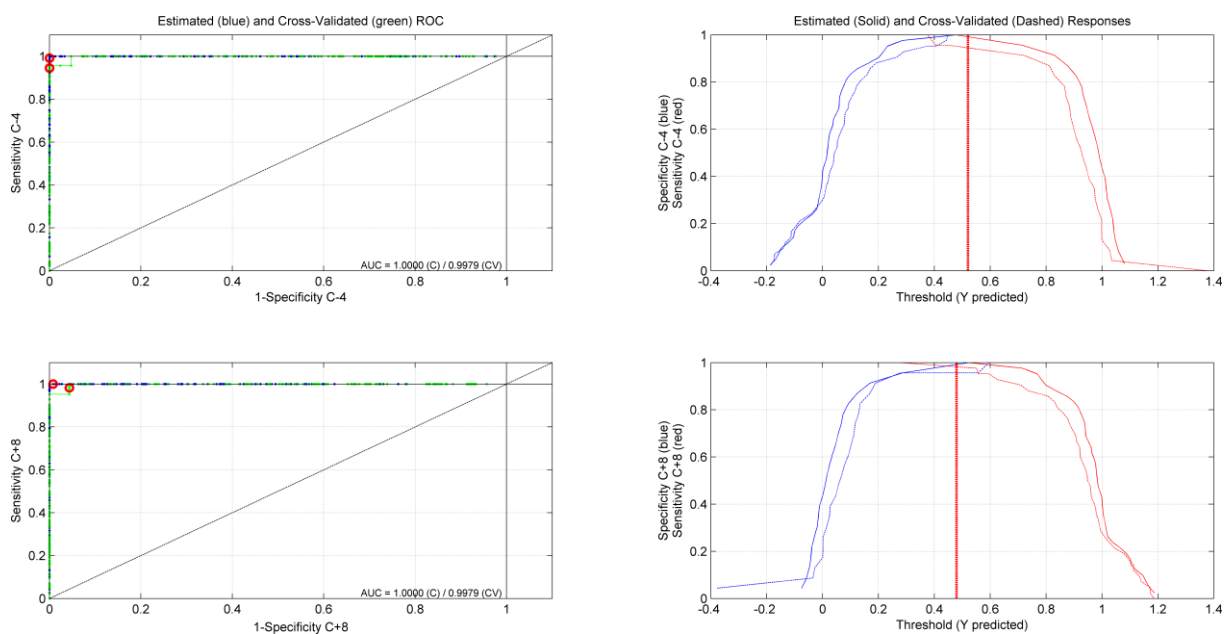


Figure S6. Curvas de análisis ROC derivadas del modelo OPLS-DA comparando los estadios C-4 y C+8 ($AUC = 0.99$).

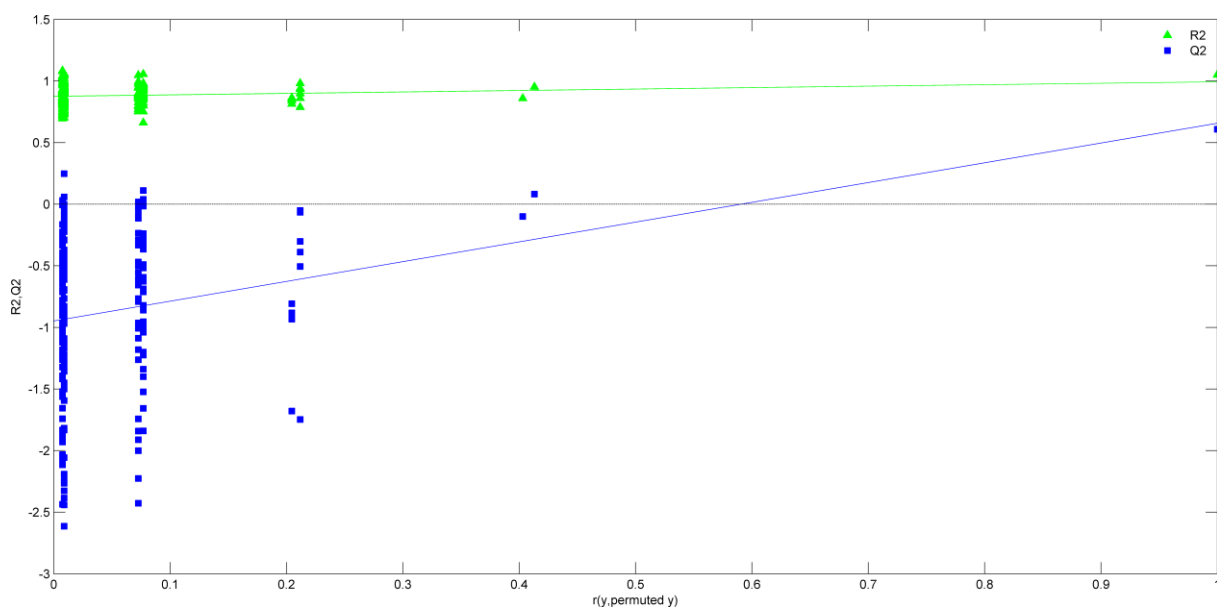


Figure S7. Gráficos de prueba de permutación para el modelo OPLS-DA comparando vacas lecheras multiparas y primíparas dentro del estadio C-4 ($R^2Y = 0.82$ and $Q^2Y = 0.56$).

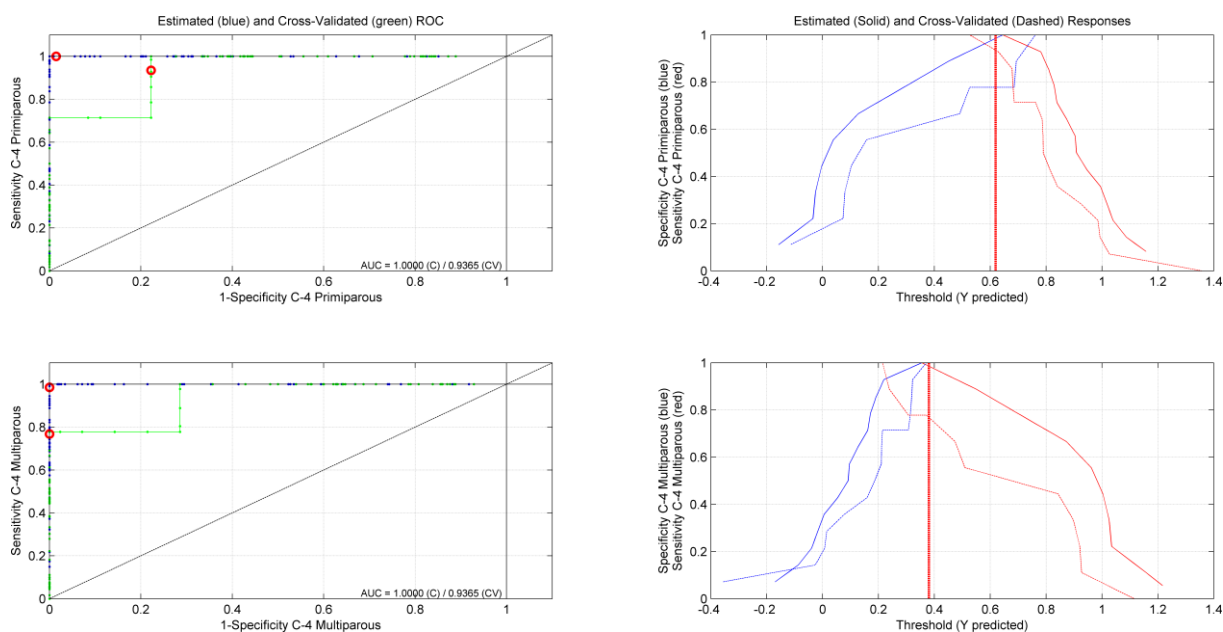


Figure S8. Curvas de análisis ROC derivadas del modelo OPLS-DA comparando vacas lecheras multiparas y primíparas dentro del estadio C-4 (AUC = 0.93).

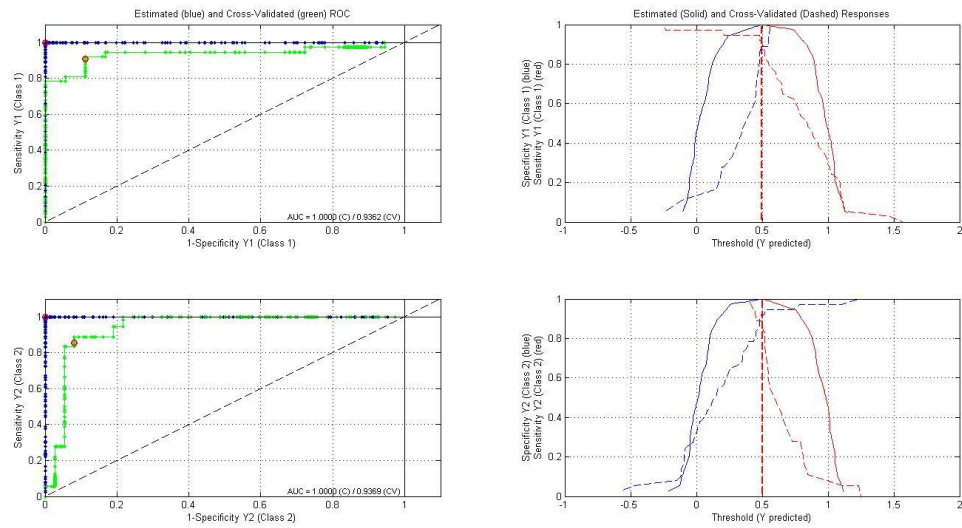


Figure S9. Curvas de análisis ROC derivadas del modelo OPLS-DA comparando genotipos de cebada sencibles y resistentes.

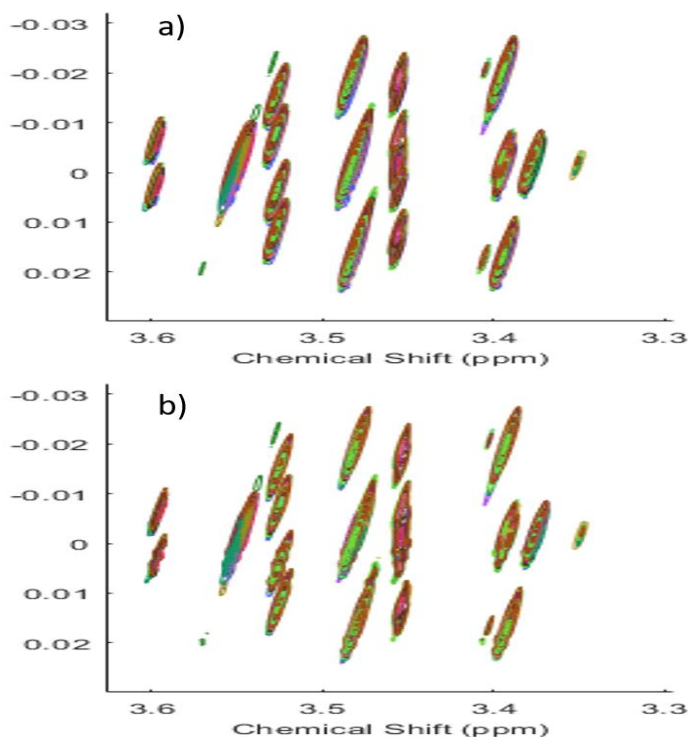


Figure S10. Superposición de un conjunto de espectros HOMO 2DJ inclinados antes (a) y después (b) de alinear con JIBA. Se pueden apreciar las deformaciones en los picos.

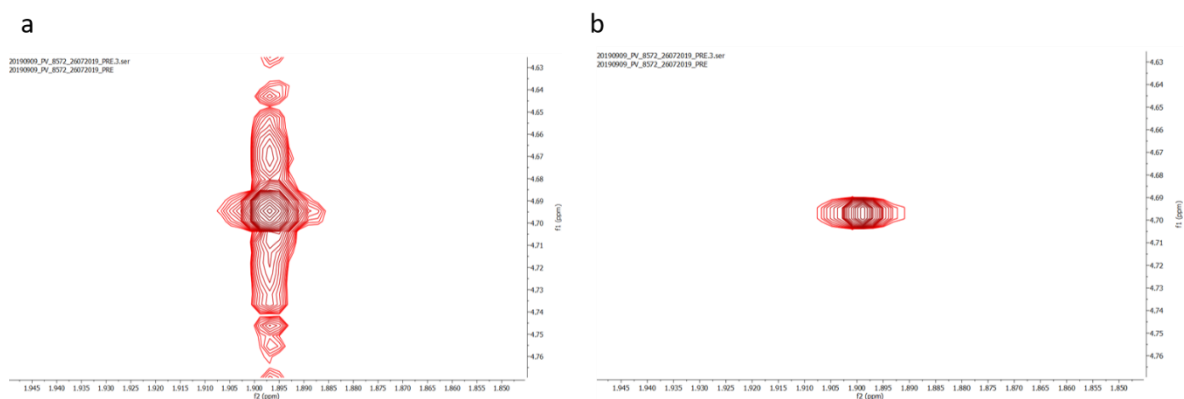


Figure S10. Ampliación de la región del acetato de un espectro HOMO 2DJ inclinado (a) e inclinado/simetrizado (b). Se puede apreciar la presencia artefacto en T_1 importantes para la señal en 1,9 en (a), que al simetrizar lleva a una cancelación del área de la misma.

Algoritmo de JIBA

```
function varargout = Guie_Aliganament(varargin)
% GUIE_ALIGANAMENT MATLAB code for Guie_Aliganament.fig
%   GUIE_ALIGANAMENT, by itself, creates a new GUIE_ALIGANAMENT
%   or raises the existing
%   singleton*.
%
%   H = GUIE_ALIGANAMENT returns the handle to a new
%   GUIE_ALIGANAMENT or the handle to
%   the existing singleton*.
%
%   GUIE_ALIGANAMENT('CALLBACK',hObject,eventData,handles,...) calls
%   the local
%   function named CALLBACK in GUIE_ALIGANAMENT.M with the given
%   input arguments.
%
%   GUIE_ALIGANAMENT('Property','Value',...) creates a new
%   GUIE_ALIGANAMENT or raises the
%   existing singleton*. Starting from the left, property value pairs are
%   applied to the GUI before Guie_Aliganament_OpeningFcn gets called. An
%   unrecognized property name or invalid value makes property application
%   stop. All inputs are passed to Guie_Aliganament_OpeningFcn via varargin.
%
%   *See GUI Options on GUIDE's Tools menu. Choose "GUI allows only one
%   instance to run (singleton)".
%
% See also: GUIDE, GUIDATA, GUIHANDLES

% Edit the above text to modify the response to help Guie_Aliganament

% Last Modified by GUIDE v2.5 22-Nov-2019 10:18:52

% Begin initialization code - DO NOT EDIT
gui_Singleton = 1;
gui_State = struct('gui_Name',    mfilename, ...
    'gui_Singleton', gui_Singleton, ...
    'gui_OpeningFcn', @Guie_Aliganament_OpeningFcn, ...
    'gui_OutputFcn', @Guie_Aliganament_OutputFcn, ...
    'gui_LayoutFcn', [] , ...
    'gui_Callback', []);
if nargin && ischar(varargin{1})
    gui_State.gui_Callback = str2func(varargin{1});
end
```

```

if nargout
    [varargout{1:nargout}] = gui_mainfcn(gui_State, varargin{:});
else
    gui_mainfcn(gui_State, varargin{:});
end
% End initialization code - DO NOT EDIT

% --- Executes just before Guie_Alignament is made visible.
function Guie_Alignament_OpeningFcn(hObject, eventdata, handles, varargin)
% This function has no output args, see OutputFcn.
% hObject    handle to figure
% eventdata  reserved - to be defined in a future version of MATLAB
% handles     structure with handles and user data (see GUIDATA)
% varargin    command line arguments to Guie_Alignament (see VARARGIN)

% Choose default command line output for Guie_Alignament
handles.output = hObject;
% Update handles structure
guidata(hObject, handles);

% UIWAIT makes Guie_Alignament wait for user response (see UIRESUME)
% uiwait(handles.figure1);

% --- Outputs from this function are returned to the command line.
function varargout = Guie_Alignament_OutputFcn(hObject, eventdata, handles)
% varargout  cell array for returning output args (see VARARGOUT);
% hObject    handle to figure
% eventdata  reserved - to be defined in a future version of MATLAB
% handles     structure with handles and user data (see GUIDATA)

% Get default command line output from handles structure
varargout{1} = handles.output;

% --- Executes on button press in load_data.
function load_data_Callback(hObject, eventdata, handles)
% hObject    handle to load_data (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles     structure with handles and user data (see GUIDATA)

a=2;

if a==1

% format_file_load= {'*.csv'}; % type
% [filenames, direction] = uigetfile(format_file_load,'MultiSelect', 'on');

```

else

```
    format_file_load = {'*.mat'}; % type
    [filenames, direction] = uigetfile(format_file_load, 'MultiSelect', 'on');

    load(strcat(direction, filenames));

    hObject.UserData.jres_f2 = jres_f2;
    hObject.UserData.jres_data = jres_data;
    hObject.UserData.names = names;
    hObject.UserData.jres_f1 = jres_f1;

    %
    jres_data = zeros(length(hObject.UserData.jres_f1), length(hObject.UserData.jres_f2), length(hObject.UserData.names));
    % for k=1:length(hObject.UserData.names)
    %
    % % para hacer el tilt
    % % [f2mat, f1mat] = meshgrid(jres_f2, jres_f1);
    % % jres_data_tilt(:, :, k) = interp2(f2mat, f1mat, jres_data(:, :, k), f2mat - f1mat, f1mat);
    %
    % % para hacer el simetrizado
    %
    %
    hObject.UserData.jres_data(:, :, k) = jresSymmetrize(hObject.UserData.jres_data(:, :, k));
    %
    % end

    hObject.UserData.names = names;
    hObject.UserData.jres_f1 = jres_f1;
    hObject.UserData.Jres_sum = zeros(length(hObject.UserData.names), length(hObject.UserData.jres_f2));
    hObject.UserData.ax1 = axes();
    hObject.UserData.ax1.Position = [0.18 0.55 0.8 0.4];

    for i = 1:length(hObject.UserData.names)
        jres = hObject.UserData.jres_data(:, :, i);
        jressum = sum(jres);
        hObject.UserData.Jres_sum(i, :) = jressum;
        plot(hObject.UserData.jres_f2, jressum);
        hold on
    end
    hold on
```

```

hObject.UserData.selec_ref=str2double(get(handles.Reference,'String'));
hObject.UserData.Ref_y=hObject.UserData.Jres_sum(hObject.UserData.selec_ref,:);
plot(hObject.UserData.jres_f2,hObject.UserData.Ref_y,'LineWidth',3,'LineStyle','-', 'Color','red');

```

```

hObject.UserData.ax1.XDir='reverse';
hObject.UserData.ax1.XLabel.String='Chemical Shift (ppm)';
hObject.UserData.ax1.YLabel.String='Intensity';
hObject.UserData.ax1.XLim=[0.5 10];
hObject.UserData.MaxInten=max(sum(hObject.UserData.Jres_sum));
hObject.UserData.minInten=min(sum(hObject.UserData.Jres_sum));
end

```

```

function Reference_Callback(hObject, eventdata, handles)

```

```

% hObject   handle to Reference (see GCBO)
% eventdata reserved - to be defined in a future version of MATLAB
% handles   structure with handles and user data (see GUIDATA)

```

```

% Hints: get(hObject,'String') returns contents of Reference as text
%       str2double(get(hObject,'String')) returns contents of Reference as a double

```

```

% --- Executes during object creation, after setting all properties.

```

```

function Reference_CreateFcn(hObject, eventdata, handles)

```

```

% hObject   handle to Reference (see GCBO)
% eventdata reserved - to be defined in a future version of MATLAB
% handles   empty - handles not created until after all CreateFcns called

```

```

% Hint: edit controls usually have a white background on Windows.

```

```

%       See ISPC and COMPUTER.

```

```

if ispc && isequal(get(hObject,'BackgroundColor'), get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end

```

```

% --- Executes on button press in Select_reg.

```

```

function Select_reg_Callback(hObject, eventdata, handles)

```

```

% hObject   handle to Select_reg (see GCBO)
% eventdata reserved - to be defined in a future version of MATLAB
% handles   structure with handles and user data (see GUIDATA)

```

```

maxI=zeros(1,length(handles.load_data.UserData.jres_f2));
for r=1:length(handles.load_data.UserData.jres_f2)
    max_r=handles.load_data.UserData.Jres_sum(:,r);
    max_rT=max_r';
    maxI(:,r)=max(max_rT);
end

```

```

end
MaxI=max(maxI);

hObject.UserData.Reg_AL_JresProj = []
Intency = []; % Matrices de salida a usar para alinear
while 0<1 %ver bien xq
    [x,y,b] = ginput(1); %funcion seleccion de pares x,y grafica
    hObject.UserData.Reg_AL_JresProj=[hObject.UserData.Reg_AL_JresProj;x]
    Intency=[Intency;y];
    if mod(length(hObject.UserData.Reg_AL_JresProj),2)==0
        % %      for q=1:(length(hObject.UserData.Reg_AL_JresProj)/2)
        %
        ppm_in_ppm_fin=reshape(((hObject.UserData.Reg_AL_JresProj'),2,(length(hObject.UserData.Reg_AL_JresProj)/2));
        %      ppm_in_ppm_fin_flip=flip(ppm_in_ppm_fin,2)
        %      hObject.UserData.ppmtable=ppm_in_ppm_fin_flip'
        %      hObject.UserData.ppmtable2= [(ppm_in_ppm_fin_flip (1,q))
        (ppm_in_ppm_fin_flip(2,q))]

        handles.uitable3.Data=((hObject.UserData.Reg_AL_JresProj);
        set(handles.uitable3,'data', hObject.UserData.Reg_AL_JresProj);

    %      end
end
%      h = uitable('data', raw, 'ColumEditable', true(1, size(raw,2)));

if isempty(b)
    break; %para terminar la seleccion "enter"
elseif b==91

[~,hObject.UserData.loc]=ismembertol(x,hObject.UserData.Reg_AL_JresProj,0.004);
hObject.UserData.Reg_AL_JresProj(hObject.UserData.loc)=[];
zoom on; %si apreto "[" hago zoom y mantiene
pause()

elseif b==93 % si apreto "]" sale del zoom y vuelve a seleccionar puntos
    zoom off;
else % Enter o cerando grafico nos da los bloques [ppm inetncidad]
    % Reg_AL_JresProj=[Reg_AL_JresProj;x];
    % Intency=[Intency;y];
    Y=0.0007;
    hold on
    scatter(x,Y,'o','r','filled');
    if mod(length(hObject.UserData.Reg_AL_JresProj),2)==0
        for q=1:(length(hObject.UserData.Reg_AL_JresProj)/2)

```

```

ppm_in_ppm_fin=reshape((hObject.UserData.Reg_AL_JresProj'),2,(length(hObject.UserData.Reg_AL_JresProj)/2));
    ppm_in_ppm_fin_flip=flip(ppm_in_ppm_fin,2)
    hObject.UserData.ppmtable=ppm_in_ppm_fin_flip'
    rectangle('Position',[ppm_in_ppm_fin_flip(2,q)
(ppm_in_ppm_fin_flip(1,q)-ppm_in_ppm_fin_flip(2,q))
MaxI],'EdgeColor','w','FaceColor',[1 0 0,0.2]);
%     handles.uitable3.Data=(hObject.UserData.ppmtable');
end
dd= hObject.UserData.Reg_AL_JresProj
%     table3=uitable3('Position';
%     table_as_cell = num2cell(hObject.UserData.ppmtable);
%     disp(table3,table_as_cell);
%     uitable
%     table_handle = uitable3('Position',[2.8 30.923 37.2 18.154]);
%     table_as_cell = num2cell(ppmtable);
%     set(table_handle, 'Data', table_as_cell);
end

end
end

```

```

(hObject.UserData.Reg_AL_JresProj=[(hObject.UserData.Reg_AL_JresProj;x]
Intensity=[Intensity;y];
%     hObject.UserData.Interval_ppm_I=[(hObject.UserData.Reg_AL_JresProj
Intensity];

```

% --- Executes on button press in Aline.

function Aline_Callback(hObject, eventdata, handles)

% hObject handle to Aline (see GCBO)

% eventdata reserved - to be defined in a future version of MATLAB

% handles structure with handles and user data (see GUIDATA)

% Alineado Jres en f2 por bloques y por punto de f1 muestra por muestra vs referencia

% Nota icoshift:

% [Mcortef1AL, intervals, indexes] = icoshift
(ref_AL,Matrizcortef1_reg(p,:),Reg_AL_JresProjT,'b',[0 1 1 0.09 1],jres_f2);%

alineado icoshift muestra por muestra vs ref

% Mcortef1AL = matriz alineada

% ref_AL = espectro referencia para alinear

% Matrizcortef1_reg(p,:)=fila(f2) p(muestra) del bloque i(f1) a alinear. la

% toma de la matriz cortada

% Reg_AL_JresProjT=regions delimitadas seleccionadas previamente de las

```

% Jresproj
% 'b',[0 1 1 0.09 1] = metodo icoshift(b)[ 0(sin salida grafica) 1
% (completar con punto anterior) 1(Co-Shift, AL global) 0,09(tolerancia de
% corrimiento en ppm) 1(datos en ppm)

Matrizcortef1_reg=zeros(length(handles.load_data.UserData.names),length(handles.load_data.UserData.jres_f2)); %matriz 2D corte f2xMuestra
corte_f1_reg=zeros(length(handles.load_data.UserData.names),length(handles.load_data.UserData.jres_f2));% matriz corte 3D dimension (punto de f1:f2:Muestras)
MJresALf1_reg=zeros(length(handles.load_data.UserData.jres_f1),length(handles.load_data.UserData.jres_f2),length(handles.load_data.UserData.names));%Matriz de salida 3D alineada en f1 por bloques

ppm_reg=handles.load_data.UserData.jres_f2;
Mkortef1AL_p=zeros(length(handles.load_data.UserData.names),length(handles.load_data.UserData.jres_f2));

h = waitbar(0,'Please wait loading Data....');
for i=1:length(handles.load_data.UserData.jres_f1)
    corte_f1_reg=handles.load_data.UserData.jres_data(i,:);% matriz de corte 3D

    for j=1:length(handles.load_data.UserData.names)
        cortef1=corte_f1_reg(1,:j);% vector de corte dimensiones(punto f1:muetsra(j): f2)
        Matrizcortef1_reg(j,:)=cortef1; %matriz corte 2D a alinear "feta(i) de f2:muestras
    end
    for p=1:length(handles.load_data.UserData.names)
        Reg_AL_JresProjT=handles.Select_reg.UserData.Reg_AL_JresProj;

        [~,interval_loc]=ismembertol(Reg_AL_JresProjT,handles.load_data.UserData.jres_f2,0.0004);%Intervalos=reshape(Reg_AL_JresProjT,1,2*length(Reg_AL_JresProjT));
        ref_AL=Matrizcortef1_reg((handles.load_data.UserData.selec_ref,:));
        %Referencia
        [Mkortef1AL, intervals, indexes] = icoshift(ref_AL,Matrizcortef1_reg(p,:),Reg_AL_JresProjT,'b',[0 1 1 0.09 1],handles.load_data.UserData.jres_f2);% alineado icoshift muestra por muestra vs ref
        Mkortef1AL_p(p,:)=Mkortef1AL; %coloco cada traza de muetsra alineada, armo matriz traza-muetsra que fue alineada muetsra por muetsra vs ref
    end

    for k=1:length(handles.load_data.UserData.names)

```

```

        MJresALf1_reg(i,:k)=Mcorfef1AL_p(k,:); %rearmado matríz por punto
en f1.

                                %Por cada punto de f1(j), coloca por cada
muestras(m)
        end                                %el vector de la matriz alineada correspondiente(k).

waitbar(i/(length(handles.load_data.UserData.jres_f1)));
%Matriz de salida 3D

end

close(h)

hObject.UserData.ax2=axes();
hObject.UserData.ax2.Position=[0.18 0.08 0.8 0.4];

hObject.UserData.MJresumAL_reg=zeros(size(MJresALf1_reg,3),size(MJresAL
f1_reg,2));

for t=1:length(handles.load_data.UserData.names)
    JresumAL = sum(MJresALf1_reg(:,:,t));
    hObject.UserData.MJresumAL_reg(t,:)=JresumAL;
    plot(handles.load_data.UserData.jres_f2,JresumAL)
    hold on
end

hold on
plot(handles.load_data.UserData.jres_f2,handles.load_data.UserData.Ref_y,'Lin
eWidth',3,'LineStyle','-', 'Color','red');

hObject.UserData.ax2.XDir='reverse';
hObject.UserData.ax2.XLabel.String='Chemical Shift (ppm)';
hObject.UserData.ax2.YLabel.String='Intensity';
hObject.UserData.ax2.XLim=[0.5 10];

linkaxes([handles.load_data.UserData.ax1,hObject.UserData.ax2],'xy');

[hObject.UserData.DataAL_MJresALf1_reg]=MJresALf1_reg;
% Hint: get(hObject,'Value') returns toggle state of Aline

% % --- Executes on button press in plot_imagesc.
% function plot_reg_Callback(hObject, eventdata, handles)
% % hObject    handle to plot_imagesc (see GCBO)
% % eventdata  reserved - to be defined in a future version of MATLAB
% % handles    structure with handles and user data (see GUIDATA)
%
```

```

% ax1
% hold on
%
%
[~,Loc_ppm_intervalos]=ismembertol(handles.Select_reg.Reg_AL_JresProj,handles.load_data.jres_f2,0.0004);
%
[~,Loc_Inetn_max]=ismembertol(handles.load_data.MaxInten,handles.load_data.jres_f1,0.0004);
%
[~,Loc_Inten_min]=ismembertol(handles.load_data.minInten,handles.load_data.jres_f1,0.0004);
%
ppm_in_ppm_fin=reshape((Reg_AL_JresProj'),2,(length(Loc_ppm_intervalos)/2));
% ppm_in_ppm_finT=ppm_in_ppm_fin';
%
Loc_ppm_in_ppm_fin=reshape(Loc_ppm_intervalos,(length(Loc_ppm_intervalos)/2),2);
% Loc_ppm_in_ppm_finT=Loc_ppm_in_ppm_fin';
% Loc_ppm_in_ppm_finT_flip=flip(Loc_ppm_in_ppm_finT,2);
% ppm_in_ppm_fin_flip=flip(ppm_in_ppm_fin,2);
% for q=1:length(ppm_in_ppm_fin_flip)
%     rectangle('Position',[ppm_in_ppm_fin_flip(2,q)      -0
(ppm_in_ppm_fin_flip(1,q)-ppm_in_ppm_fin_flip(2,q))
MaxI],'EdgeColor','w','FaceColor',[1 0 0 ,0.2]);
% end
%
%
% % Hint: get(hObject,'Value') returns toggle state of plot_imagesc

% --- Executes on button press in Save.
function Save_Callback(hObject, eventdata, handles)
% hObject    handle to Save (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

dlmwrite('DataAL_MJresALf1_reg.mat',
handles.Aline.UserData.DataAL_MJresALf1_reg);

% --- Executes on button press in radiobutton2.
function radiobutton2_Callback(hObject, eventdata, handles)
% hObject    handle to radiobutton2 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

```

```

if isfield(handles.load_data.UserData,'ax1')
    delete(handles.load_data.UserData.ax1);
    delete(handles.Aline.UserData.ax2);
end

if isfield(hObject.UserData,'ax5')
    delete(hObject.UserData.ax5);
    delete(hObject.UserData.ax6);

end

h=get(hObject,'Value');

if h==1
    hObject.UserData.ax3=axes();
    hObject.UserData.ax3.Position=[0.18 0.55 0.8 0.4];
    imagesc(handles.load_data.UserData.Jres_sum)
    hObject.UserData.ax3.XDir='reverse';
    hObject.UserData.ax3.XLabel.String='Chemical Shift (ppm)';
    hObject.UserData.ax3.YLabel.String='Espectra number';
    hObject.UserData.ax4=axes();
    hObject.UserData.ax4.Position=[0.18 0.08 0.8 0.4];
    imagesc(handles.Aline.UserData.MJressumAL_reg)
    hObject.UserData.ax4.XDir='reverse';
    hObject.UserData.ax4.XLabel.String='Chemical Shift (ppm)';
    hObject.UserData.ax4.YLabel.String='Espectra number';

    linkaxes([hObject.UserData.ax3,hObject.UserData.ax4],'xy');

%     delete(hObject.UserData.ax5)
%     delete(hObject.UserData.ax6)
else

    delete(hObject.UserData.ax3)
    delete(hObject.UserData.ax4)

    hObject.UserData.ax5=axes();
    hObject.UserData.ax5.Position=[0.18 0.55 0.8 0.4];

    plot(handles.load_data.UserData.jres_f2,handles.load_data.UserData.Jres_sum);
    hold on

    plot(handles.load_data.UserData.jres_f2,handles.load_data.UserData.Ref_y,'Lin
eWidth',3,'LineStyle',':', 'Color','red');

    hObject.UserData.ax5.XDir='reverse';

```

```

hObject.UserData.ax5.XLabel.String='Chemical Shift (ppm)';
hObject.UserData.ax5.YLabel.String='Intensity';
hObject.UserData.ax5.XLim=[0.5 10];
hObject.UserData.ax6=axes();
hObject.UserData.ax6.Position=[0.18 0.08 0.8 0.4];

plot(handles.load_data.UserData.jres_f2,handles.Aline.UserData.MJressumAL_
reg);

hold on

plot(handles.load_data.UserData.jres_f2,handles.load_data.UserData.Ref_y,'Lin
eWidth',3,'LineStyle','-', 'Color','red');

hObject.UserData.ax6.XDir='reverse';
hObject.UserData.ax6.XLabel.String='Chemical Shift (ppm)';
hObject.UserData.ax6.YLabel.String='Intensity';
hObject.UserData.ax6.XLim=[0.5 10];

linkaxes([hObject.UserData.ax5,hObject.UserData.ax6],'xy');

end

% hObject.UserData.ax1.XLim=[0.5 10];
% Hint: get(hObject,'Value') returns toggle state of radiobutton2

% --- Executes on button press in eval_contur.
function eval_contur_Callback(hObject, eventdata, handles)
% hObject    handle to eval_contur (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

if isfield(handles.load_data.UserData,'ax1')
    delete(handles.load_data.UserData.ax1);
    delete(handles.Aline.UserData.ax2);
end

if isfield(handles.radiobutton2.UserData,'ax5')
    delete(handles.radiobutton2.UserData.ax5);
    delete(handles.radiobutton2.UserData.ax6);
end

end

if isfield(handles.radiobutton2.UserData,'ax3')
    delete(handles.radiobutton2.UserData.ax3);
    delete(handles.radiobutton2.UserData.ax4);
end

```

```

end

f=get(hObject,'Value');
if f==1
    hObject.UserData.ax7=axes();
    hObject.UserData.ax7.Position=[0.18 0.55 0.5 0.4];

plot(handles.load_data.UserData.jres_f2,handles.load_data.UserData.Jres_sum);
    hold on

plot(handles.load_data.UserData.jres_f2,handles.load_data.UserData.Ref_y,'LineWidth',3,'LineStyle','-', 'Color','red');

    hObject.UserData.ax7.XDir='reverse';
    hObject.UserData.ax7.XLabel.String='Chemical Shift (ppm)';
    hObject.UserData.ax7.YLabel.String='Intensity';
    hObject.UserData.ax7.XLim=[0.5 10];

    hObject.UserData.ax8=axes();
    hObject.UserData.ax8.Position=[0.18 0.08 0.5 0.4];

plot(handles.load_data.UserData.jres_f2,handles.Aline.UserData.MJressumAL_
reg);

    hold on

plot(handles.load_data.UserData.jres_f2,handles.load_data.UserData.Ref_y,'LineWidth',3,'LineStyle','-', 'Color','red');

    hObject.UserData.ax8.XDir='reverse';
    hObject.UserData.ax8.XLabel.String='Chemical Shift (ppm)';
    hObject.UserData.ax8.YLabel.String='Intensity';
    hObject.UserData.ax8.XLim=[0.5 10];

linkaxes([hObject.UserData.ax7,hObject.UserData.ax8],'xy');

if isfield(hObject.UserData,'ax9')
    delete(hObject.UserData.ax9);
    delete(hObject.UserData.ax10);
end

if isfield(handles.plotcontourreg.UserData,'ax15')
    delete(handles.plotcontourreg.UserData.ax15);
    delete(handles.plotcontourreg.UserData.ax16);
end

```

```

else
    if isfield(hObject.UserData,'ax7')
        delete(hObject.UserData.ax7);
        delete(hObject.UserData.ax8);
    end
    % delete(handles.load_data.radiobutton2.UserData.ax5);
    % delete(handles.load_data.radiobutton2.UserData.ax6);

    % delete(hObject.UserData.ax3)
    % delete(hObject.UserData.ax4)

    hObject.UserData.ax9=axes();
    hObject.UserData.ax9.Position=[0.18 0.55 0.5 0.4];

    plot(handles.load_data.UserData.jres_f2,handles.load_data.UserData.Jres_sum);
    hold on

    plot(handles.load_data.UserData.jres_f2,handles.load_data.UserData.Ref_y,'Lin
eWidth',3,'LineStyle',':', 'Color','red');

    hObject.UserData.ax9.XDir='reverse';
    hObject.UserData.ax9.XLabel.String='Chemical Shift (ppm)';
    hObject.UserData.ax9.YLabel.String='Intensity';
    hObject.UserData.ax9.XLim=[0.5 10];

    hObject.UserData.ax10=axes();
    hObject.UserData.ax10.Position=[0.18 0.07 0.5 0.4];

    plot(handles.load_data.UserData.jres_f2,handles.Aline.UserData.MJressumAL_
reg);

    hold on

    plot(handles.load_data.UserData.jres_f2,handles.load_data.UserData.Ref_y,'Lin
eWidth',3,'LineStyle',':', 'Color','red');

    hObject.UserData.ax10.XDir='reverse';
    hObject.UserData.ax10.XLabel.String='Chemical Shift (ppm)';
    hObject.UserData.ax10.YLabel.String='Intensity';
    hObject.UserData.ax10.XLim=[0.5 10];

    linkaxes([hObject.UserData.ax9,hObject.UserData.ax10],'xy');

end
% Hint: get(hObject,'Value') returns toggle state of eval_contur

```

```

% --- Executes on button press in reg_contour.
function reg_contour_Callback(hObject, eventdata, handles)
% hObject    handle to reg_contour (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

maxI=zeros(1,length(handles.load_data.UserData.jres_f2));
    for r=1:length(handles.load_data.UserData.jres_f2)
        max_r=handles.load_data.UserData.Jres_sum(:,r);
        max_rT=max_r';
        maxI(:,r)=max(max_rT);
    end
    MaxI=max(maxI);

hObject.UserData.Reg_corte = []; Intensity = []; % Matrices de salida a usar para
alinear
while 0<1 %ver bien xq
    [x,y,s] = ginput(1); %funcion seleccion de pares x,y grafica
    hObject.UserData.Reg_corte=[hObject.UserData.Reg_corte;x];
    Intensity=[Intensity;y];

    % if isempty(b)
    % break; %para terminar la seleccion "enter"
    if s==91
        zoom on; %si apreto "[" hago zoom y mantiene
        pause();
        [~,loc]=ismembertol(x,hObject.UserData.Reg_corte,0.0004);
        hObject.UserData.Reg_corte(loc)=[];
    elseif s==93 % si apreto "]" sale del zoom y vuelve a seleccionar puntos
        zoom off;
    elseif isempty(s)
        break;
    else % Enter o cerando grafico nos da los bloques [ppm inetncidad]
        % Reg_AL_JresProj=[Reg_AL_JresProj;x];
        % Intensity=[Intensity;y];
        Y=0.0007;
        hold on
        scatter(x,Y,'o','r','filled');
        if mod(length(hObject.UserData.Reg_corte),2)==0
            for q=1:(length(hObject.UserData.Reg_corte)/2)

ppm_in_ppm_fin_c=reshape((hObject.UserData.Reg_corte'),2,(length(hObject.U
serData.Reg_corte)/2));
            hObject.UserData.ppm_in_ppm_fin_flip_c=flip(ppm_in_ppm_fin_c,2);

```

```

        rectangle('Position',[hObject.UserData.ppm_in_ppm_fin_flip_c(2,q)      -0
(hObject.UserData.ppm_in_ppm_fin_flip_c(1,q)-
hObject.UserData.ppm_in_ppm_fin_flip_c(2,q))
MaxI],'EdgeColor','w','FaceColor',[1 0 0,0.2]);
    end

%
[~,corte_ppm_menor]=ismembertol(ppm_in_ppm_fin_flip_c(2,1),handles.load_
data.UserData.jres_f2,0.0004);
%
[~,corte_ppm_mayor]=ismembertol(ppm_in_ppm_fin_flip_c(1,1),handles.load_
data.UserData.jres_f2,0.0004);
%    [~,Jmax]=max(handles.load_data.UserData.jres_f1);
%    [~,Jmin]=min(handles.load_data.UserData.jres_f1);
%
jres_cortada=handles.load_data.UserData.jres_data(Jmin:Jmax,corte_ppm_men
or:corte_ppm_mayor,1:(length(handles.load_data.UserData.names)));
%
jres_cortadaAL=handles.Aline.UserData.DataAL_MJresALf1_reg(Jmin:Jmax,cor
te_ppm_menor:corte_ppm_mayor,1:(length(handles.load_data.UserData.names
)));
%
f2_corte=handles.load_data.UserData.jres_f2(corte_ppm_menor:corte_ppm_ma
yor);
%    c=rand(length(handles.load_data.UserData.names),3);
%
%    hObject.UserData.ax9=axes();
%    hObject.UserData.ax9.Position=[0.75 0.55 0.23 0.4];
%
%    for g=1:length(handles.load_data.UserData.names)
%
nmrContourPlot(f2_corte,handles.load_data.UserData.jres_f1,jres_cortada(:,g),'
Color',c(g,:));
%    hold on
%    end
%
%    hObject.UserData.ax9.XDir='reverse';
%    hObject.UserData.ax9.XLabel.String='Chemical Shift (ppm)';
%    hObject.UserData.ax9.YLabel.String='Espectra number';
%
%    hObject.UserData.ax10=axes();
%    hObject.UserData.ax10.Position=[0.75 0.07 0.23 0.4];
%
%    for g=1:length(handles.load_data.UserData.names)

```

```

%
nmrContourPlot(f2_corte,handles.load_data.UserData.jres_f1,jres_cortadaAL(:,g),
'Color',c(g,:));
%     hold on
%     end
%
%     hObject.UserData.ax10.XDir='reverse';
%     hObject.UserData.ax10.XLabel.String='Chemical Shift (ppm)';
%     hObject.UserData.ax10.YLabel.String='Espectra number';
%
%     linkaxes ([hObject.UserData.ax9,hObject.UserData.ax10],'xy')
end

end

end

hObject.UserData.Reg_corte=[hObject.UserData.Reg_corte;x];
Intensity=[Intensity;y];
hObject.UserData.Interval_ppm_I=[hObject.UserData.Reg_corte Intensity];

% --- Executes on button press in plotcontour.
function plotcontour_Callback(hObject, eventdata, handles)
% hObject    handle to plotcontour (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

[~,corte_ppm_menor]=ismembertol(handles.reg_contour.UserData.ppm_in_pp
m_fin_flip_c(2,1),handles.load_data.UserData.jres_f2,0.0004);

[~,corte_ppm_mayor]=ismembertol(handles.reg_contour.UserData.ppm_in_pp
m_fin_flip_c(1,1),handles.load_data.UserData.jres_f2,0.0004);
    [~,Jmax]=max(handles.load_data.UserData.jres_f1);
    [~,Jmin]=min(handles.load_data.UserData.jres_f1);

jres_cortada=handles.load_data.UserData.jres_data(Jmin:Jmax,corte_ppm_men
or:corte_ppm_mayor,1:(length(handles.load_data.UserData.names)));

jres_cortadaAL=handles.Aline.UserData.DataAL_MJresALf1_reg(Jmin:Jmax,cor
te_ppm_menor:corte_ppm_mayor,1:(length(handles.load_data.UserData.names
)));

f2_corte=handles.load_data.UserData.jres_f2(corte_ppm_menor:corte_ppm_ma
yor);
    c=rand(length(handles.load_data.UserData.names),3);

```

```

hObject.UserData.ax11=axes();
hObject.UserData.ax11.Position=[0.75 0.55 0.23 0.4];

for g=1:length(handles.load_data.UserData.names)

nmrContourPlot(f2_corte,handles.load_data.UserData.jres_f1,jres_cortada(:,g),'
Color',c(g,:));
hold on
end

hObject.UserData.ax11.XDir='reverse';
hObject.UserData.ax11.XLabel.String='Chemical Shift (ppm)';
hObject.UserData.ax11.YLabel.String='Espectra number';

hObject.UserData.ax12=axes();
hObject.UserData.ax12.Position=[0.75 0.07 0.23 0.4];

for g=1:length(handles.load_data.UserData.names)

nmrContourPlot(f2_corte,handles.load_data.UserData.jres_f1,jres_cortadaAL(:,
g),'Color',c(g,:));
hold on
end

hObject.UserData.ax12.XDir='reverse';
hObject.UserData.ax12.XLabel.String='Chemical Shift (ppm)';
hObject.UserData.ax12.YLabel.String='Espectra number';

linkaxes ([hObject.UserData.ax11,hObject.UserData.ax12],'xy')

function samplenum_Callback(hObject, eventdata, handles)
% hObject handle to samplenum (see GCBO)
% eventdata reserved - to be defined in a future version of MATLAB
% handles structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'String') returns contents of samplenum as text
% str2double(get(hObject,'String')) returns contents of samplenum as a double

% --- Executes during object creation, after setting all properties.
function samplenum_CreateFcn(hObject, eventdata, handles)
% hObject handle to samplenum (see GCBO)
% eventdata reserved - to be defined in a future version of MATLAB
% handles empty - handles not created until after all CreateFcns called

% Hint: edit controls usually have a white background on Windows.
% See ISPC and COMPUTER.

```

```

if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end

% --- Executes on button press in plot_proj.
function plot_proj_Callback(hObject, eventdata, handles)
% hObject    handle to plot_proj (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

if isfield(handles.plotcontour.UserData,'ax11')
    delete(handles.plotcontour.UserData.ax11);
    delete(handles.plotcontour.UserData.ax12);
end

if isfield(hObject.UserData,'ax13')
    delete(hObject.UserData.ax13);
    delete(hObject.UserData.ax14);
end

if isfield(handles.eval_contur.UserData,'ax9')
    delete(handles.eval_contur.UserData.ax9);
    delete(handles.eval_contur.UserData.ax10);
end

if isfield(handles.eval_contur.UserData,'ax7')
    delete(handles.eval_contur.UserData.ax7);
    delete(handles.eval_contur.UserData.ax8);
end

if isfield(handles.plotcontourreg.UserData,'ax15')
    delete(handles.plotcontourreg.UserData.ax15);
    delete(handles.plotcontourreg.UserData.ax16);
end
    hObject.UserData.sampnum=str2double(get(handles.sampnum,'String'));
    hObject.UserData.ax13=axes();
    hObject.UserData.ax13.Position=[0.18 0.55 0.5 0.4];

Jres_sum_sampnum=handles.load_data.UserData.Jres_sum(hObject.UserData.s
ampnum,:);
    plot(handles.load_data.UserData.jres_f2,Jres_sum_sampnum);
    hold on

plot(handles.load_data.UserData.jres_f2,handles.load_data.UserData.Ref_y,'Lin
eWidth',3,'LineStyle','-', 'Color','red');

```

```

hObject.UserData.ax13.XDir='reverse';
hObject.UserData.ax13.XLabel.String='Chemical Shift (ppm)';
hObject.UserData.ax13.YLabel.String='Intensity';
hObject.UserData.ax13.XLim=[0.5 10];

hObject.UserData.ax14=axes();
hObject.UserData.ax14.Position=[0.18 0.08 0.5 0.4];

MJressumAL_reg_sampnum=handles.Aline.UserData.MJressumAL_reg(hObject
t.UserData.sampnum,:);
plot(handles.load_data.UserData.jres_f2,MJressumAL_reg_sampnum);

hold on

plot(handles.load_data.UserData.jres_f2,handles.load_data.UserData.Ref_y,'Lin
eWidth',3,'LineStyle','-', 'Color','red');

hObject.UserData.ax14.XDir='reverse';
hObject.UserData.ax14.XLabel.String='Chemical Shift (ppm)';
hObject.UserData.ax14.YLabel.String='Intensity';
hObject.UserData.ax14.XLim=[0.5 10];

linkaxes([hObject.UserData.ax13,hObject.UserData.ax14],'xy');

% --- Executes on button press in selecprojreg.
function selecprojreg_Callback(hObject, eventdata, handles)
% hObject    handle to selecprojreg (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

if isfield(handles.load_data.UserData,'ax11')
    delete(handles.load_data.UserData.ax11);
    delete(handles.Aline.UserData.ax12);
end
if isfield(handles.plotcontourreg.UserData,'ax15')
    delete(handles.plotcontourreg.UserData.ax15);
    delete(handles.plotcontourreg.UserData.ax16);
end

maxI=zeros(1,length(handles.load_data.UserData.jres_f2));
for r=1:length(handles.load_data.UserData.jres_f2)
    max_r=handles.load_data.UserData.Jres_sum(:,r);
    max_rT=max_r';
    maxI(:,r)=max(max_rT);
end

```

```

MaxI=max(maxI);

hObject.UserData.Reg_corte2 = []; Intencity = []; % Matrices de salida a usar para
alinear
while 0<1 %ver bien xq
    [x,y,s] = ginput(1); %funcion seleccion de pares x,y grafica
    hObject.UserData.Reg_corte2=[hObject.UserData.Reg_corte2;x];
    Intencity=[Intencity;y];
    % if isempty(b)
    % break; %para terminar la seleccion "enter"
    if s==91
        zoom on; %si apreto "[" hago zoom y mantiene
        pause();
        [~,loc]=ismembertol(x,hObject.UserData.Reg_corte2,0.0004);
        hObject.UserData.Reg_corte2(loc)=[];
    elseif s==93 % si apreto "]" sale del zoom y vuelve a seleccionar puntos
        zoom off;
    elseif isempty(s)
        break;
    else % Enter o cerando grafico nos da los bloques [ppm inetncidad]
        % Reg_AL_JresProj=[Reg_AL_JresProj;x];
        % Intencity=[Intencity;y];
        Y=0.0007;
        hold on
        scatter(x,Y,'o','r','filled');
        if mod(length(hObject.UserData.Reg_corte2),2)==0
            for q=1:(length(hObject.UserData.Reg_corte2)/2)

                ppm_in_ppm_fin_c2=reshape((hObject.UserData.Reg_corte2'),2,(length(hObject
                .UserData.Reg_corte2)/2));
                hObject.UserData.ppm_in_ppm_fin_flip_c2=flip(ppm_in_ppm_fin_c2,2);
                rectangle('Position',[hObject.UserData.ppm_in_ppm_fin_flip_c2(2,q) -0
                (hObject.UserData.ppm_in_ppm_fin_flip_c2(1,q)-
                hObject.UserData.ppm_in_ppm_fin_flip_c2(2,q))
                MaxI],'EdgeColor','w','FaceColor',[1 0 0,0.2]);
            end
        end

        %
    end

end

end

hObject.UserData.Reg_corte2=[hObject.UserData.Reg_corte2;x];
Intencity=[Intencity;y];
hObject.UserData.Interval_ppm_I2=[hObject.UserData.Reg_corte2 Intencity];

```

```

% --- Executes on button press in plotcontourreg.
function plotcontourreg_Callback(hObject, eventdata, handles)
% hObject    handle to plotcontourreg (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

[~,corte_ppm_menor2]=ismembertol(handles.selecprojreg.UserData.ppm_in_p
pm_fin_flip_c2(2,1),handles.load_data.UserData.jres_f2,0.0004);

[~,corte_ppm_mayor2]=ismembertol(handles.selecprojreg.UserData.ppm_in_p
pm_fin_flip_c2(1,1),handles.load_data.UserData.jres_f2,0.0004);
    [~,Jmax2]=max(handles.load_data.UserData.jres_f1);
    [~,Jmin2]=min(handles.load_data.UserData.jres_f1);

jres_cortada2=handles.load_data.UserData.jres_data(Jmin2:Jmax2,corte_ppm_
menor2:corte_ppm_mayor2,1:(length(handles.load_data.UserData.names)));

jres_cortadaAL2=handles.Aline.UserData.DataAL_MJresALf1_reg(Jmin2:Jmax2
,corte_ppm_menor2:corte_ppm_mayor2,1:(length(handles.load_data.UserData.
names)));

f2_corte=handles.load_data.UserData.jres_f2(corte_ppm_menor2:corte_ppm_m
ayor2);
%     c=rand(length(handles.load_data.UserData.names),3);

    hObject.UserData.ax15=axes();
    hObject.UserData.ax15.Position=[0.75 0.55 0.23 0.4];

%     for g=1:length(handles.load_data.UserData.names)

nmrContourPlot(f2_corte,handles.load_data.UserData.jres_f1,jres_cortada2(:,h
andles.load_data.UserData.selec_ref),'Color',[1 0 0])
    alpha 0.2
    hold on

nmrContourPlot(f2_corte,handles.load_data.UserData.jres_f1,jres_cortada2(:,h
andles.plot_proj.UserData.sampnum),'Color',[0 0 1]);

%     end

    hObject.UserData.ax15.XDir='reverse';
    hObject.UserData.ax15.XLabel.String='Chemical Shift (ppm)';
    hObject.UserData.ax15.YLabel.String='Espectra number';

    hObject.UserData.ax16=axes();

```

```

hObject.UserData.ax16.Position=[0.75 0.07 0.23 0.4];

%   for g=1:length(handles.load_data.UserData.names)

nmrContourPlot(f2_corte,handles.load_data.UserData.jres_f1,jres_cortada2(:,h
andles.load_data.UserData.selec_ref),'Color',[1 0 0])
    hold on

nmrContourPlot(f2_corte,handles.load_data.UserData.jres_f1,jres_cortadaAL2(:,
handles.plot_proj.UserData.sampnum),'Color',[0 0 1]);

hObject.UserData.ax16.XDir='reverse';
hObject.UserData.ax16.XLabel.String='Chemical Shift (ppm)';
hObject.UserData.ax16.YLabel.String='Espectra number';

linkaxes ([hObject.UserData.ax15,hObject.UserData.ax16],'xy')

% --- Executes when entered data in editable cell(s) in uitable3.
function uitable3_CellEditCallback(hObject, eventdata, handles)
% hObject   handle to uitable3 (see GCBO)
% eventdata   structure with the following fields (see
MATLAB.UI.CONTROL.TABLE)
% Indices: row and column indices of the cell(s) edited
% PreviousData: previous data for the cell(s) edited
% EditData: string(s) entered by the user
% NewData: EditData or its converted form set on the Data property. Empty if
Data was not changed
% Error: error string when failed to convert EditData to appropriate value for
Data
% handles   structure with handles and user data (see GUIDATA)

% hObject.UserData.uitable3=(handles.Selec_reg.UserData.Reg_AL_JresProj);

% --- Executes when selected cell(s) is changed in uitable3.
function uitable3_CellSelectionCallback(hObject, eventdata, handles)
% hObject   handle to uitable3 (see GCBO)
% eventdata   structure with the following fields (see
MATLAB.UI.CONTROL.TABLE)
% Indices: row and column indices of the cell(s) currently selecteds
% handles   structure with handles and user data (see GUIDATA)

hObject.UserData.indices=eventdata.Indices(1);

% --- Executes on button press in delpoint.
function delpoint_Callback(hObject, eventdata, handles)
% hObject   handle to delpoint (see GCBO)

```

```

% eventdata reserved - to be defined in a future version of MATLAB
% handles structure with handles and user data (see GUIDATA)
indi=(handles.uitable3.UserData.indices);

handles.Select_reg.UserData.Reg_AL_JresProj(indi)=[];
handles.Select_reg.UserData.Reg_AL_JresProj(indi-1)=[];
%
handles.Select_reg.UserData.Reg_AL_JresProj(handles.Select_reg.UserData.loc)
=[];
% maxI=zeros(1,length(handles.load_data.UserData.jres_f2));
%     for r=1:length(handles.load_data.UserData.jres_f2)
%         max_r=handles.load_data.UserData.Jres_sum(:,r);
%         max_rT=max_r';
%         maxI(:,r)=max(max_rT);
%     end
%     MaxI=max(maxI);
%
delete(rectangle('Position',[(handles.Select_reg.UserData.Reg_AL_JresProj(indi)
) -0 (handles.Select_reg.UserData.Reg_AL_JresProj(indi-1))-
(handles.Select_reg.UserData.Reg_AL_JresProj(indi)) MaxI]));

set(handles.uitable3,'data',handles.Select_reg.UserData.Reg_AL_JresProj)
%handles.Select_reg.UserData.Reg_AL_JresProj)

% handles.uitable3.Data(handles.uitable3.UserData.indices)=[];

```

Algoritmo gráfico de valores de carga bidimensional

```
% Desescalar y genrar vector y %
wt=w';
srsd=sqrt(sd);
y=srsd.*wt;

%% Datos Jres

%ppm=Jres_f2_corte;
J=jres_f1;
% MJres=Jres_1.Intensity; promedio de todas las lineas ppm

%% Des-reShape

% szJ=size(jres_cortadaf1,1);
% szppm=size(jres_cortadaf1,2);
% sz_estirado=szppm*szJ;
szppm=length(ppm);
szJ=length(J);

y_2D2=reshape(y,szppm,szJ);
w_2D2=reshape(w,szppm,szJ);
y_2D2T=y_2D2';
w_2D2T=w_2D2';
%% Plot

figure(1)
contour(ppm,J,y_2D2T,100,'LineStyle','-');

figure(2);
contour(w_2D2T);

figure(3);
contour3(ppm,J,w_2D2,100,'LineWidth',2);

figure(4);
nmrContourPlot(ppm,J,y_2D2T,w_2D2T);

figure(5);
ax=axes();
contour(ppm,J,w_2D2T,100,'LineWidth',2);
ax.XDir='reverse';

%%
```

```

y(end) = NaN;

figure
patch(varIDf1,y,w,'EdgeColor','interp','LineWidth',1);
colorbar;
set(gca,'xDir','reverse');

% x-axis label

xlabel('Chemical Shift (ppm)');

% y-axis label

ylabel('OPLS Coefficients');

% END %

```

Algoritmo proyección gráfico de valores de carga bidimensional

%%% proj loading desescalado

```
proj_y2D2_rT=sum(y_2D2_rT);  
proj_w_2D2_rTmax=max(w_2D2_rT);  
proj_w_2D2_rTmin=min(w_2D2_rT);  
proj_w2D2_rTMm=[proj_w_2D2_rTmax;proj_w_2D2_rTmin];  
sumproj=sum(proj_w2D2_rTMm);
```

```
loading_proj=zeros(1,length(sumproj));  
for i=1:length(sumproj)  
    n=proj_w_2D2_rTmax(i)+proj_w_2D2_rTmin(i);  
    if n>0  
        loading_proj(:,i)=proj_w_2D2_rTmax(i);  
    else  
        loading_proj(:,i)=proj_w_2D2_rTmin(i);  
    end  
end
```

```
proj_y2D2_rT(end) = NaN;
```

```
figure(1)  
patch(ppm,proj_y2D2_rT,loading_proj,'EdgeColor','interp','LineWidth',1);  
colorbar;  
set(gca,'xDir','reverse');
```

```
% proj_y2D2_rT(end) = NaN;  
%  
% figure(2)  
%  
patch(ppm,proj_y2D2_rT,proj_w_2D2_rTmin,'EdgeColor','interp','LineWidth',1)  
;  
% colorbar;  
% set(gca,'xDir','reverse');
```

Algoritmo STOCSY y funciones asociadas

%%Función para extracción de datos

```
function [X,ppm]= myext (data)
```

```
X1=data(:,2:end); %% Matriz sin la primera fila %%  
X=X1';  
ppm1= data(:,1);%% Primera fila %%  
ppm=ppm1';  
end
```

Función para calculo traza covarianza

```
function [COVXYM]=myCOV4(c,data)  
[X,ppm]= myext(data); %Extrae ppm y matriz cruda en forma vertical%
```

```
ppm1=ppm';  
idx = findindx(ppm,c);  
Cdp = X(:,idx);  
Dp_media = mean(Cdp);  
sumx = Cdp-Dp_media;
```

```
COVXY=cell(length(X(1,:)),1);
```

```
for i=1:size(X,2)  
    Vy=X(:,i);  
    Vymedia=mean(Vy);  
    sumy=Vy-Vymedia;  
    prod_sumxy=sumy.*sumx;  
    mediasumxy=(sum(prod_sumxy))/(size(X,2));  
    COVXY{i}=mediasumxy;
```

```
end
```

```
COVXYT=COVXY';  
COVXYM=cell2mat(COVXYT);  
end
```

%%Función para calculo traza correlación

```
function [CORRM]=myCORR4(c,data)
```

```
[X,ppm]= myext(data); %Extrae ppm y matriz cruda en forma vertical%
```

```

idx= findindx(ppm,c);
Cdp = X(:,idx);
Dp_media = mean(Cdp);
sumx = Cdp-Dp_media;
sumx2=sumx.*sumx;
sumx2suma=sum(sumx2);

CORR=cell(length(X(1,:)),1);

for i=1:size(X,2)
    Vy=X(:,i);
    Vymedia=mean(Vy);
    sumy=Vy-Vymedia;
    sumy2=sumy.*sumy;
    sumy2suma=sum(sumy2);
    sumxy=sumy.*sumx;
    sumxysuma=sum(sumxy);
    denominador=sqrt(sumx2suma*sumy2suma);
    CORRvariable=sumxysuma/denominador;
    CORR{i}=CORRvariable;

end

CORRt=CORR';
CORRM=cell2mat(CORRt);

end

```

%%Función STOC SY

```

function [COV,CORR]=my_Stocsy_trazas6(c,data)
[D,ppm1]=myext(data);
ppm2=ppm1';
[COV]=myCOV4(c,data);
[CORR]=myCORR4(c,data);

y=COV;
c=CORR;
y(end) = NaN;

figure

patch(ppm1,y,c,'EdgeColor','interp');
d=colorbar;
set(gca,'xDir','reverse')

```

```
xlabel('Chemical Shift (ppm)')
ylabel('intensity')
title(d,'corr.coeff.')
title(gca,'Stocsy')
end

%%END%
```

Algoritmo de grafico de valores de carga monodimensional

```
% Datos a extraer

% w = factores de carga (Weights on LV1 se saca de la tabla del grafico de loading)
% sd = desviación estandar (se saca de los datos crudos)
% ppm = corrimiento quimico, eje x (se saca de los datos crudos)

% Solo para escalado usando Pareto
% Revisen los extremos de los vectores si hay que sacar datos no
% correspondan. Todos los vectores tienen que tener las mismas dimensiones
% Trasponer loading %

wt=w';

% Desescalar y genrar vector y %

srsd=sqrt(sd);
y=srsd.*wt;

% Graficar ppm vs y, colorar por loading %

y(end) = NaN;

figure
%patch(ppm,y,wt,'EdgeColor',c,'interp','LineWidth',1);%
patch(ppm,y,wt,'EdgeColor','interp','LineWidth',1);
colorbar;
colormap('jet');
set(gca,'xDir','reverse');

% x-axis label

xlabel('Chemical Shift (ppm)') ;

% y-axis label

ylabel('OPLS Coefficients') ;

% END %
```

Article

Monitoring the Transition Period in Dairy Cows through ¹H NMR-Based Untargeted Metabolomics

Andrés López Radcenco ¹, María de Lourdes Adrien ², Gretel Rupprechter ³, Elena de Torres ⁴, Ana Meikle ^{3,*} and Guillermo Moyna ^{1,*}

¹ Departamento de Química del Litoral, CENUR Litoral Norte, Universidad de la República, Ruta 3 Km 363, Paysandú 60000, Uruguay; alopez@fq.edu.uy

² Departamento de Salud en los Sistemas Pecuarios, Facultad de Veterinaria, Universidad de la República, Ruta 3 Km 363, Paysandú 60000, Uruguay; lourdesadrien@gmail.com

³ Laboratorio de Endocrinología y Metabolismo Animal, Facultad de Veterinaria, Universidad de la República, Avenida Alberto Lasplacas 1550, Montevideo 11600, Uruguay; gcrupprechter@gmail.com

⁴ Campo Experimental 2, Facultad de Veterinaria, Universidad de la República, Ruta 1 Km 42.5, Libertad, San José 80100, Uruguay; elena.detorres@gmail.com

* Correspondence: meikleana@gmail.com (A.M.); gmoyna@fq.edu.uy (G.M.)

Abstract: The metabolic alterations associated with the increase in milk production make the transition period critical to the health of dairy cows, usually leading to a higher incidence of disease in periparturient animals. In this manuscript, we describe the use of NMR-based untargeted metabolomics to follow how these changes impact the serum metabolome in a group of 28 transition dairy cows with no initial clinical diseases. Principal component analysis (PCA) of serum ¹H NMR data from four weeks before calving to 8 weeks after parturition allowed us to clearly identify four stages during the transition period. Pairwise comparisons using orthogonal partial least square discriminant analysis (OPLS-DA) and univariate data analysis led to the identification of 18 metabolites that varied significantly through these stages. Species such as acetate, betaine, and creatine are observed early after calving, while other markers of metabolic stress, including acetone, β-hydroxybutyrate (BHB), and choline, accumulate significantly at the height of milk production. Furthermore, marked variations in the levels of lactate, allantoin, alanine, and other amino acids reveal the activation of different gluconeogenic pathways following parturition. Concomitant with a return to homeostasis, a gradual normalization of the serum metabolome occurs 8 weeks after calving. Correlations of metabolite levels with dietary and metabolic adaptations based on animal parity could also be identified. Overall, these results show that NMR-based chemometric methods are ideally suited to monitor manifestations of metabolic diseases throughout the transition period and to assess the impact of nutritional management schemes on the metabolism of dairy cows.

Keywords: dairy cattle; metabolic stress; NMR-based metabolomics; serum metabolomics; transition period



Citation: López Radcenco, A.; Adrien, M.d.L.; Rupprechter, G.; de Torres, E.; Meikle, A.; Moyna, G. Monitoring the Transition Period in Dairy Cows through ¹H NMR-Based Untargeted Metabolomics. *Dairy* **2021**, *2*, 356–366. <https://doi.org/10.3390/dairy2030028>

Academic Editor: Luciano Caixeta

Received: 19 March 2021

Accepted: 18 June 2021

Published: 2 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

During the transition period from late pregnancy to early lactation, dairy cows suffer dramatic changes in their metabolism due to the decrease in dry matter intake (DMI) occurring even before calving and the increased energetic needs associated with milk production [1]. These changes lead to a state of negative energy balance (NEB), which results in the activation of lipid and protein catabolic pathways [2,3]. If the NEB is severe or prolonged, it affects milk production and, more importantly, has effects on the immune and endocrine systems that are detrimental to animal health and reproductive performance [4–7]. Indeed, the transition period is characterized by an increase in the incidence of production diseases such as mastitis, metritis, ketosis, and hepatic lipidosis [3,8]. Furthermore, the process of continuous genetic selection that has been employed over recent decades to improve milk production yields has also led to an exacerbation of problems associated with metabolic stress [9,10].

Since the animal health issues observed during the transition period translate into important economic losses, considerable efforts have been dedicated to better understand the metabolic changes of dairy cows in peripartum. Several studies have focused on monitoring metabolites associated with NEB, particularly increases in the concentration of non-esterified fatty acids (NEFA) in blood [11,12], which are usually followed by increases in β -hydroxybutyrate (BHB), acetone, and acetoacetate levels [13]. The variation in the concentration of glucose, insulin and other hormones, and blood proteins also correlates with the metabolic challenges experienced by the animals around calving [6,11,14]. However, production diseases are etiologically interrelated and cannot be considered in isolation [8], and thus integrative analytical approaches in which groups of metabolites are monitored simultaneously are desirable. Metabolomic methods offer this possibility [15], and consequently they have been successfully employed to investigate dairy cattle health. For example, targeted metabolomics employing mass spectrometry (MS) have been used to investigate mastitis, metritis, ketosis, and hepatic lipidosis in periparturient cows [16–20], and untargeted approaches based on nuclear magnetic resonance (NMR) spectroscopy were also applied to the study of several of these conditions [20–24].

Targeted and untargeted MS-based metabolomics have also been applied to follow metabolite levels in transition cows in an attempt to identify biomarkers that could evidence early stages of disease [25–29]. Most of these studies have been conducted under confined dairy production systems. However, maximizing the proportion of pastures in the diet is a pivotal factor for minimizing production costs [30]. It is well known that grazing dairy cows do not have sufficient DMI to sustain the high milk production that could be achieved with their genetic potential [31], and supplements and concentrates are therefore given to achieve the desired production levels and minimize the magnitude and duration of the NEB. In contrast with total mixed ration (TMR) diet schemes, the prediction of nutrient availability in mixed systems where herbage is the primary diet component is complex because it includes uncertainties associated with grazing behavior. In this scenario, the availability of propionate and other glucogenic metabolites for liver demands may be limited so as to maintain adequate glucose supply for the mammary gland of the high-producing dairy cows. Thus, the characterization of alternative precursors and pathways involved in *de novo* glucose synthesis and the identification of liver and peripheral tissue adaptations to lactation could contribute to the optimization of herd nutritional management strategies. Moreover, this knowledge could be instrumental in the design of preventive medicine programs aimed at decreasing costly production diseases. With this in mind, we investigated the suitability of NMR-based untargeted metabolomics as a tool to follow changes in the serum metabolome of a group of dairy cows with no initial clinical conditions and access to pastures following calving.

2. Materials and Methods

2.1. Animals and Sample Collection

The experimental protocol was approved by the Comisión Honoraria de Experimentación Animal, Universidad de la República (111130-002261-13). The study was performed on a group of 26 Holstein, Jersey, and Holstein $\times$ Jersey crossbred cows managed together at the dairy farm of the Facultad de Veterinaria, Universidad de la República (San José, Uruguay). It included 17 heifers and 9 multiparous cows that had between three and seven calvings. Their diet consisted of a concentrate and silage based TMR before calving, and grazing was offered as an option after that (Table S1). Monitoring was carried out by trained veterinarians from ~30 days prepartum to ~60 days postpartum. Animals with clinical manifestations of hypocalcemia, retained placenta, metritis, or moderate or severe mastitis (i.e., grades 2 or 3) were excluded from the study [32–35], while those developing mild clinical mastitis (i.e., grade 1) or subclinical ketosis that did not show systemic affection were not [35,36]. Bleeding was carried out every 15 days before calving, and once a week immediately after milking and before grazing after parturition. Blood

samples were drawn from the coccygeal vein into tubes without anticoagulant, centrifuged at 3000 g for 10 min, and the resulting serum stored at -20°C until analyzed.

2.2. Clinical Biochemistry

Cholesterol, total protein, albumin, urea, and calcium concentrations in serum were determined with commercial colorimetric kits (Weiner Laboratorios S.A.I.C., Rosario, Argentina), using a Vitalab Selectra 2 automated biochemistry analyzer. The globulin concentration was estimated as the difference between the total protein and albumin concentrations. NEFA concentrations were determined following the ACS ACOD method (NEFA-C kit, Wako Chemicals USA Inc., Richmond, VA, USA), and enzymatic methods were employed to obtain concentrations of BHB in all samples (Ranbut kit, Randox Laboratories Ltd., Crumlin, UK). Intra and inter-assay coefficients of variation for all commercial serum determinations were less or equal than 10%.

2.3. NMR Spectroscopy

Blood serum samples were allowed to thaw at room temperature, and 400 μL aliquots were mixed with 200 μL of a 0.9% NaCl solution in D_2O and transferred to 5 mm NMR tubes (NE HL5 7, New Era Enterprises Inc., Vineland, NJ, USA). All NMR spectra were recorded at 25°C on a Bruker AVANCE III 500 NMR spectrometer operating at ^1H and ^{13}C frequencies of 500.13 MHz and 125.76 MHz, respectively, and equipped with a z-gradient TXI probe. To attenuate broad signals from proteins and lipids, 1D ^1H spectra were obtained using a water suppressed Carr Purcell Meiboom-Gill sequence with a τ of 0.4 ms and an n of 80 [37,38]. A spectral width of 10 KHz, a data size of 32 K, and a total of 128 scans were employed to record each spectrum, using a relaxation delay of 4 s between scans. When required, gradient enhanced HSQC and 1D TOCSY spectra were acquired using standard pulse sequences provided with the spectrometer.

2.4. Data Processing

All free induction decays were zero-filled to 64 K points and apodized with a 0.3 Hz exponential window function prior to Fourier transformation. ^1H NMR spectra were manually phased and baseline corrected using MNova (version 10.0, MestreLab Research, S.L., Santiago de Compostela, Spain), and referenced to the lactate methyl resonance at 1.25 ppm present in all serum samples. Manually selected regions of the spectra were then aligned, and the data was normalized to the total spectral area after excluding the residual water resonance signal (4.60–5.00 ppm). Spectra were then segmented into bins of 0.001 ppm between 0.5 and 9.0 ppm [39]. The resulting data matrix was then exported as a text file for use in multivariate analyses.

2.5. Metabolite Concentrations

The molar ratios of unambiguously identified serum metabolites were obtained by integration of their ^1H signals in the NMR spectrum of each sample. Together with the concentration of BHB obtained through enzymatic methods for the same sample, these estimations were employed to compute absolute metabolite concentrations.

2.6. Statistical Analysis

One-way analysis of variance (ANOVA) was performed using GraphPad Prism (version 7.0, GraphPad Software Inc., San Diego, CA, USA). Multivariate statistical analyses, including principal component analysis (PCA), partial least squares discriminant analysis (PLS-DA), and orthogonal partial least squares discriminant analysis (OPLS-DA), were carried out with the PLS_Toolbox package (version 8.5, Eigenvector Research Inc., Manson, WA, USA) implemented for MATLAB (revision 2014a, The MathWorks Inc., Natick, MA, USA). For all models, the data was mean-centered and scaled using a Pareto factor [40]. Analysis of the data was first performed with PCA, which reduces the dimensionality and facilitates the identification of data clusters or trends [41–43]. The PCA scores plot was

also employed to identify strong outliers outside the 95% significance region of Hotelling's T2 ellipse. OPLS-DA models were derived from the complete data matrix, as well as from subsets in the 6.0 to 9.0 ppm range. Cross-validation of all OPLS-DA models was achieved using the random subset method, which involved 20 iterations over data split into 10 equally sized parts. Receiver operating characteristic (ROC) curves were plotted, and area under the curve (AUC) values were calculated to ensure the goodness of fit of the resulting models [44,45]. A permutation test with 200 iterations was also performed to determine the degree of over-fitting and further validate the discriminant analyses [46]. The results from these validations are provided as Supplementary Materials.

3. Results and Discussion

3.1. ^1H NMR Spectra of Serum Samples

A representative ^1H NMR spectrum of a serum sample used in the study is shown in Figure 1. In addition to signals corresponding to lipoproteins, sugars, and amino acids, resonances from metabolites such as acetate (1.84 ppm), acetone (2.16 ppm), allantoin (5.30 ppm), betaine (3.19 and 3.83 ppm), BHB (1.12, 2.22, 2.33, and 4.06 ppm), choline (3.08, 3.36, and 3.90 ppm), citrate (2.45 and 2.61 ppm), citrulline (1.43, 1.73, 3.00, and 3.63 ppm), creatine (2.96 and 3.86 ppm), creatinine (2.97 and 3.98 ppm), formate (8.38 ppm), glycerolphosphocholine (3.15, 3.54, 3.60, 3.83, and 4.23 ppm), isobutyrate (1.04 and 2.55 ppm), lactate (1.25 and 4.03 ppm), phosphocholine (3.17, 3.58, and 4.10 ppm), trimethylamine (TMA) (2.84 ppm), trimethylamine N-oxide (TMAO) (3.28 ppm), and tyramine (6.85 and 7.05 ppm) can be assigned by comparison to literature data [22,37,47]. These assignments were, in some cases, further corroborated with the aid of HSQC and 1D TOCSY spectra. As discussed earlier, several of these species are related to NEB-related metabolic stress occurring in the transition period of the lactating cow.

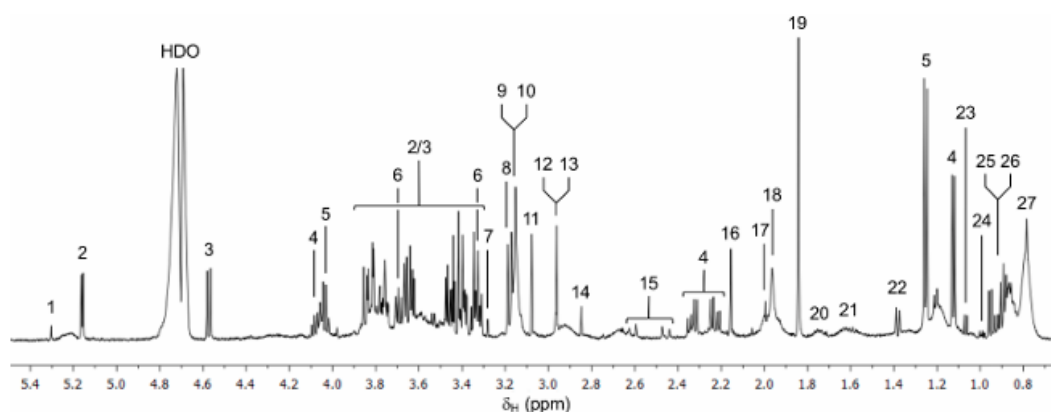


Figure 1. Representative ^1H NMR spectrum of a dairy cow serum sample. The signals corresponding to allantoin (1), α -glucose (2), β -glucose (3), BHB (4), lactate (5), myo-inositol (6), TMAO (7), betaine (8), phosphocholine (9), glycerolphosphocholine (10), choline (11), creatinine (12), creatine (13), TMA (14), citrate (15), acetone (16), *O*-acetylglycoprotein (17), *N*-acetylglycoprotein (18), acetate (19), citrulline (20), lysine (21), alanine (22), valine (23), isobutyrate (24), isoleucine (25), leucine (26), and L1 lipoprotein (27) are annotated.

3.2. NMR-Based Serum Metabolomics

The initial step in the study involved the identification of groups among the animals through peripartum. For that purpose, we carried out a PCA considering the data derived from the ^1H NMR spectra for all serum samples. The resulting score plot showed no differentiation according to parity or breed (Figure 2a,b, respectively). Similar results were

obtained when principal components capturing less variance were considered in the score plots (Figures S1 and S2). However, when a continuous coloring scheme based on days around calving was employed to color the scores, four clear clusters were identified. These groups involved samples from animals four weeks before calving ($C - 4$), and one, four, and eight weeks after calving ($C + 1$, $C + 4$, and $C + 8$, respectively), and are colored blue, light-blue, yellow, and red in Figure 2c, respectively. Blood biochemistry data and other information on animals involved in each of these stages are summarized in Table 1.

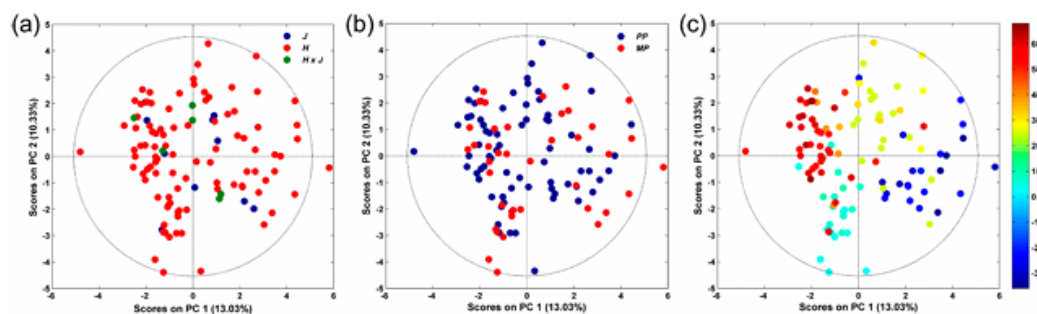


Figure 2. PCA score plot obtained from ^1H NMR spectral data for all serum samples. The plot was colored based on breed (a), parity (b), or using a continuous coloring scheme which varies as a function of the transition period day relative to calving (c).

Table 1. Animal and blood biochemistry data for cows within each stage identified through multivariate analysis of ^1H NMR data.

Animal Data	Stage			
	$C - 4$	$C + 1$	$C + 4$	$C + 8$
Days in stage	−36 to −22	3 to 11	23 to 34	39 to 68
N ($H, J, H \times J$) ^a	23 (19, 2, 2)	20 (18, 0, 2)	24 (20, 2, 2)	20 (17, 1, 2)
Parity (PP, MP) ^b	15, 8	12, 8	15, 9	13, 7
Disease (SCK, MCM) ^c	0, 9	2, 6	6, 10	2, 11
Milk production (L/day)	–	23 ± 7	28 ± 6	25 ± 7
Biochemistry				
Albumin (g/L)	31.00 ± 2.19	29.64 ± 2.91	29.44 ± 5.67	31.42 ± 2.78
Urea (mM)	2.40 ± 1.31	2.82 ± 1.10	2.65 ± 1.11	3.43 ± 1.46
Cholesterol (mM)	2.71 ± 0.43	2.29 ± 0.50	3.81 ± 0.75	5.09 ± 0.93
Total protein (g/L)	79.61 ± 7.00	74.23 ± 7.19	78.13 ± 7.72	80.52 ± 6.64
Globuline (g/L)	48.90 ± 6.50	44.59 ± 5.17	48.69 ± 5.12	49.10 ± 4.60
BHB (mM)	0.38 ± 0.21	0.56 ± 0.32	0.78 ± 0.63	0.65 ± 0.35
NEFA (mM)	0.29 ± 0.18	0.63 ± 0.22	0.68 ± 0.36	0.47 ± 0.20

^a H, J , and $H \times J$ represent Holstein, Jersey, and Holstein $\times$ Jersey crosses, respectively. ^b PP and MP indicate primiparous and multiparous, respectively. ^c SCK and MCM indicate subclinical ketosis and mild clinical mastitis (i.e., grade 1), respectively.

Having identified four peripartum stages, we investigated variations in the serum metabolome, taking prepartum metabolite levels as reference. This was achieved through a pairwise comparison of the $C - 4$ against the $C + 1$, $C + 4$, and $C + 8$ groups, respectively, using OPLS-DA and univariate data analysis (Figure 3, Figure S3 and Table 2). In all cases, the large positive peak at 1.25 ppm observed in the OPLS-DA loading plots corresponds to a sharp drop in lactic acid concentration after calving. While there are several sources of this metabolite, including dehydrogenation of pyruvate through the glyoxalase system in the metabolism of fats, carbohydrates, and proteins [48], anaerobic bacterial fermentation in rumen and colon likely leads to the high levels of lactate observed before calving in

this case [49,50]. As concluded in a recent large-scale NMR-based serum metabolomics study of clinically healthy dairy cattle [29], serum lactate can indeed be associated with an increased intake of fermentable carbohydrates such as those present in the TMR prepartum diet [51,52]. At the same time, the marked decrease in lactate levels following calving suggests that this metabolite is used to support glucose demands through gluconeogenesis by activation of the Cori cycle [53,54]. This, together with an increase in lipid metabolism through β -oxidation and the activation of the alanine cycle at the height of milk production (*vide infra*), is also consistent with the observed two- to three-fold rise in blood glucose levels after calving (Table 2), and could help explain why none of the animals had clinical manifestations of ketosis during the study.

One week after calving, negative values in the OPLS-DA loading plot for acetate, betaine, and creatine were observed (Figure 3a), indicating an increase in the levels of these metabolic stress markers [22,28]. The rise in the BHB and NEFA concentrations determined biochemically is also in agreement with this finding (Table 1). Overall, these data agree with the increase in lipid mobilization that reportedly takes place in early lactation [14,55]. As collected in Table 2, increases in asparagine, citrate, citrulline, glucose, and lysine are observed as well. This indicates alterations in the citric acid, urea, glycolytic, and gluconeogenic pathways immediately following parturition [14,28], and confirms that the animals enter a NEB condition. The rise in the levels of circulating allantoin (Table 2), a metabolite linked to purine catabolism, is consistent with higher ruminal microbial yields associated with the increase in DMI as well as the inclusion of forage in the diet after calving [56].

In addition to acetate and BHB, the loading plots for the OPLS-DA contrasting the C − 4 and C + 4 groups show that the levels of other metabolic stress markers, such as acetone, choline, and formate rise significantly four weeks postpartum (Figure 3b and Figure S3b). Similarly, univariate data analysis indicates two- to four-fold increases for the species identified in the C + 1 stage, including acetate, BHB, betaine, and glucose, as well as for creatine and creatinine (Table 2). Greater accumulation of allantoin is also observed, which at this stage is likely to be associated not only with higher bacterial protein yields caused by the increase in DMI, but with an increase in tissue damage [57]. Indeed, the accumulation of alanine points to an activation of the alanine cycle, where muscle protein is catabolized to fuel liver gluconeogenesis [2,58]. The higher levels of acetate and citrate in blood also reveal a further increase in lipomobilization through lipolysis and β -oxidation. This, in turn, triggers the ketogenic pathway, leading to the marked rise observed in the concentrations of BHB, lysine, and, ultimately, acetone [28,57]. Not surprisingly, these pronounced changes occur when animals reach the peak of milk production (Table 1) and corroborate with the large shifts in metabolism that are required to meet energy demands during this period. Indeed, based exclusively on BHB levels, nearly one out of every four animals developed subclinical ketosis at this stage [36].

Table 2. Average metabolite concentrations during peripartum determined by ^1H NMR (mean $\pm$ std. dev.) and fold changes with respect to prepartum levels.

Metabolite	Concentration (mM)				Fold Change					
	C − 4	C + 1	C + 4	C + 8	$C+1/C-4$	<i>p</i>	$C+4/C-4$	<i>p</i>	$C+8/C-4$	<i>p</i>
Acetate	0.83 $\pm$ 0.80	1.36 $\pm$ 1.22	1.99 $\pm$ 2.17	1.86 $\pm$ 1.13	1.63	0.0494	2.40	0.0091	2.25	0.0002
Acetone	0.11 $\pm$ 0.15	0.12 $\pm$ 0.08	0.17 $\pm$ 0.16	0.12 $\pm$ 0.09	1.14	0.3519	1.62	0.0724	1.10	0.3734
Alanine	0.18 $\pm$ 0.17	0.26 $\pm$ 0.22	0.43 $\pm$ 0.46	0.51 $\pm$ 0.39	1.50	0.0788	2.48	0.0071	2.88	0.0002
Allantoin	0.07 $\pm$ 0.07	0.16 $\pm$ 0.16	0.24 $\pm$ 0.27	0.20 $\pm$ 0.12	2.16	0.0167	3.41	0.0021	2.81	<0.0001
Asparagine	0.61 $\pm$ 0.42	1.71 $\pm$ 2.20	2.36 $\pm$ 2.90	1.92 $\pm$ 1.27	2.82	0.0116	3.88	0.0074	3.16	<0.0001
Betaine	0.18 $\pm$ 0.15	0.34 $\pm$ 0.38	0.56 $\pm$ 0.63	0.43 $\pm$ 0.26	1.92	0.0329	3.15	0.0035	2.44	0.0001
BHB	0.38 $\pm$ 0.20	0.56 $\pm$ 0.32	0.78 $\pm$ 0.62	0.65 $\pm$ 0.35	1.46	0.0175	2.04	0.0028	1.71	0.0009
Citrate	0.10 $\pm$ 0.09	0.26 $\pm$ 0.25	0.39 $\pm$ 0.50	0.22 $\pm$ 0.18	2.59	0.0097	4.01	0.0085	2.29	0.0035
Citrulline	0.13 $\pm$ 0.12	0.26 $\pm$ 0.31	0.48 $\pm$ 0.65	0.28 $\pm$ 0.23	2.01	0.0442	3.75	0.0069	2.18	0.0039
Creatine	0.04 $\pm$ 0.04	0.10 $\pm$ 0.14	0.15 $\pm$ 0.22	0.10 $\pm$ 0.09	2.51	0.0201	3.70	0.0105	2.33	0.0029
TMA	0.06 $\pm$ 0.04	0.23 $\pm$ 0.67	0.13 $\pm$ 0.19	0.10 $\pm$ 0.09	1.54	0.0325	2.13	0.0484	1.68	0.0229
α -Glucose	1.00 $\pm$ 0.88	2.27 $\pm$ 2.54	3.50 $\pm$ 4.15	2.91 $\pm$ 2.04	2.28	0.0145	3.51	0.0033	2.29	<0.0001
β -Glucose	0.92 $\pm$ 0.81	2.19 $\pm$ 2.61	3.49 $\pm$ 4.74	2.81 $\pm$ 2.08	2.38	0.0415	3.77	0.0068	3.04	<0.0001
Lactate	3.60 $\pm$ 3.10	1.16 $\pm$ 0.63	1.67 $\pm$ 1.13	1.60 $\pm$ 0.77	−2.23	0.0038	−2.16	0.0015	−2.26	0.0006
Leucine	0.12 $\pm$ 0.10	0.06 $\pm$ 0.03	0.06 $\pm$ 0.03	0.03 $\pm$ 0.03	−1.83	0.0166	−1.93	0.0036	−3.80	<0.0001
Lysine	0.08 $\pm$ 0.07	0.23 $\pm$ 0.31	0.37 $\pm$ 0.48	0.26 $\pm$ 0.20	2.63	0.0169	4.16	0.0037	2.97	0.0001
Valine	0.23 $\pm$ 0.17	0.22 $\pm$ 0.07	0.24 $\pm$ 0.08	0.26 $\pm$ 0.09	−1.02	0.4583	1.04	0.3938	1.14	0.1901

Metabolome data at C + 8 is of particular interest. Although cows are still at their peak of milk production, DMI also increases at this stage. As a result, animals tend to normalize their energy balance, which is evidenced by a decrease in the levels of NEFA (Table 1). This is also reflected in the OPLS-DA loading plot presented in Figure 3c, which reveals that the prepartum and C + 8 group are differentiated by metabolites such as acetate and glucose, but not by markers of severe metabolic stress present in the C + 4 group. In a similar way, univariate analysis shows that the metabolites with large fold changes for the C + 4 stage now present smaller variations when compared to their levels before calving. This indicates a trend towards the normalization of the metabolic pathways identified as altered by our analyses of the C + 1 and C + 4 groups and is consistent with the animals exiting the NEB condition and regaining homeostasis.

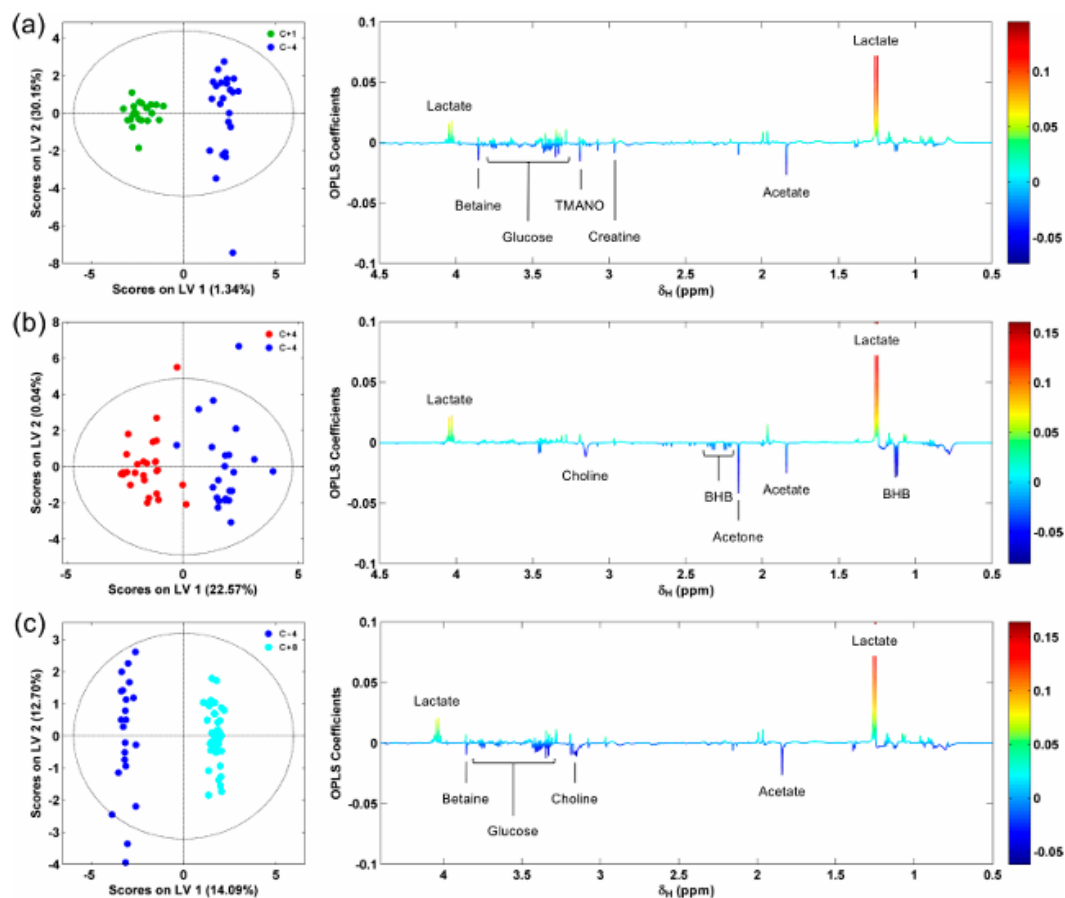


Figure 3. Score and loading factor plots obtained from the OPLS-DA between prepartum (C − 4) and first (C + 1) (a), second (C + 4) (b), and third (C + 8) (c) postpartum groups. The metabolites that differentiate the prepartum from each postpartum group are annotated in the loading factor plots. The R^2Y and Q^2Y coefficients for each model were 0.80 and 0.72, 0.89, and 0.37, and 0.94 and 0.87, respectively, and their ROC curves had AUC values of 1.00, 0.85, and 0.99 (see Figures S4–S9).

As stated above, no differences based on animal parity were observed when all the data were analyzed together. However, we decided to investigate if variations related to

this condition could be evidenced within each of the four stages. For that purpose, a PCA for each of the peripartum group datasets was carried out. While no significant effects were observed at any stage after calving, a clear difference between primiparous and multiparous cows was evident in the C – 4 group (Figure 4a). Indeed, an OPLS-DA of these data readily reveals that the concentrations of lactate, acetate, and BHB are higher in multiparous cows before calving, while accumulation of creatine, TMAO, choline, glucose, and lipoproteins is observed in heifers (Figure 4b). These data likely reflect the feeding and herd behaviors of multiparous cows, which are known to be better adapted to prepartum TMR diets and to dominate primiparous animals, and thus have a greater DMI [59,60]. The higher intake of fermentable nutrients in these animals leads, in turn, to higher levels of volatile fatty acids in their bloodstream. In addition, heifers continue growing through pregnancy, and thus the concentration of anabolic hormones, such as IGF-1, is usually greater. This impacts the availability of nutrients in the mammary gland and could contribute to lowering the levels of circulating lactate and ketone bodies in these animals [61]. Few studies have analyzed the effect of parity on the entire serum metabolome of the lactating cow [29], and this is to our knowledge one of the first reports linking the observed variations with a combination of dietary and metabolic adaptations.

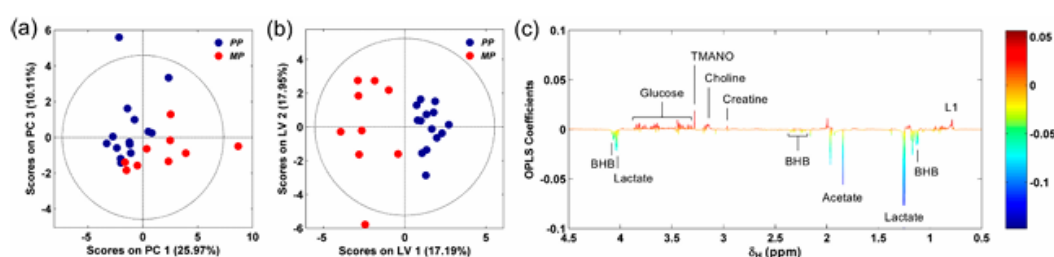


Figure 4. PCA score plot obtained from ^1H NMR spectral data for prepartum (C – 4) serum samples colored according to animal parity (a), and score and loading factor plots obtained from the OPLS-DA between primiparous and multiparous dairy cows within this group (b,c). The metabolites that differentiate the two groups are annotated in the loading factor plots. The R^2Y and Q^2Y coefficients for the model were 0.82 and 0.56, and its ROC curve had an AUC value of 0.93 (see Figures S16 and S17).

4. Conclusions

As detailed in the previous sections, NMR-based metabolomics allowed us to follow the changes of 18 serum metabolites during the transition period of dairy cows. The approach identified four clear stages through peripartum based on the metabolome composition and revealed alterations in important metabolic pathways. In particular, the data shows the sequential activation of several gluconeogenic routes after calving, including the Cori and alanine cycles. Analysis of metabolomic profiles also showed that the contribution of the former gluconeogenic pathway is affected by animal parity before calving. Although recent studies have yielded comparable results at two well-defined moments in the transition period [28], our findings indicate that markers of subacute metabolic diseases, such as ketosis and liver lipodosis, are present through most of the lactation period but peak four weeks after calving. Furthermore, the results presented here are consistent with data obtained using targeted MS-based metabolomics [27].

Overall, our findings indicate that NMR-based methods are ideally suited to readily evaluate and follow the impact of nutritional management on the metabolic health of dairy herds throughout the transition period. We are currently using this approach to assess the effects of different postpartum dietary schemes on animal condition. In addition, the combined use of metabolomic data from serum, urine, and other biofluids to achieve higher specificity will be evaluated [47,62]. Work in this and related areas is underway, and our findings will be reported in due course.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/dairy2030028/s1>, Table S1 and Figures S1–S17: Postpartum diet composition, additional PCA and OPLS-DA plots, and permutation test plots and ROC analysis curves for all OPLS-DA models.

Author Contributions: Conceptualization, A.M. and G.M.; experimental design, M.d.L.A., G.R., E.d.T. and A.M.; sample collection, M.d.L.A., G.R. and E.d.T.; data collection, data processing, and statistical analyses, A.L.R. and G.M.; writing—original draft preparation, G.M.; writing—review and editing, A.L.R., M.d.L.A., G.R., E.d.T., A.M. and G.M.; supervision, A.M. and G.M.; funding acquisition, M.d.L.A., A.M. and G.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Agencia Nacional de Investigación e Innovación (ANII), grant number FSA_1_2013_12442, and the Programa de Desarrollo de las Ciencias Básicas (PEDECIBA).

Institutional Review Board Statement: The study was conducted according to guidelines and regulations approved by the Comisión Honoraria de Experimentación Animal, Universidad de la República (11130-002261-13).

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding authors.

Acknowledgments: We wish to thank Lucas Irigoyen and Federico Hitta for their assistance in sample preparation and data collection, and Alberto Casal and Pablo Højemberg for critically reviewing the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Grummer, R.R. Impact of changes in organic nutrient metabolism on feeding the transition dairy cow. *J. Anim. Sci.* **1995**, *73*, 2820–2833. [\[CrossRef\]](#)
2. Bauman, D.E.; Currie, W.B. Partitioning of nutrients during pregnancy and lactation: A review of mechanisms involving homeostasis and homeorhesis. *J. Dairy Sci.* **1980**, *63*, 1514–1529. [\[CrossRef\]](#)
3. Ingvarsen, K.L. Feeding- and management-related diseases in the transition cow: Physiological adaptations around calving and strategies to reduce feeding-related diseases. *Anim. Feed Sci. Technol.* **2006**, *126*, 175–213. [\[CrossRef\]](#)
4. Butler, W.R. Nutritional interactions with reproductive performance in dairy cattle. *Anim. Reprod. Sci.* **2000**, *60–61*, 449–457. [\[CrossRef\]](#)
5. Drackley, J.K.; Overton, T.R.; Douglas, G.N. Adaptations of glucose and long-chain fatty acid metabolism in liver of dairy cows during the periparturient period. *J. Dairy Sci.* **2001**, *84*, E100–E112. [\[CrossRef\]](#)
6. Meikle, A.; Kulcsar, M.; Chilliard, Y.; Febel, H.; Delavaud, C.; Cavestany, D.; Chilibroste, P. Effects of parity and body condition at parturition on endocrine and reproductive parameters of the cow. *Reproduction* **2004**, *127*, 727–737. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Ingvarsen, K.L.; Moyes, K. Nutrition, immune function and health of dairy cattle. *Animal* **2012**, *7*, 112–122. [\[CrossRef\]](#)
8. Mulligan, F.J.; Doherty, M.L. Production diseases of the transition cow. *Vet. J.* **2008**, *176*, 3–9. [\[CrossRef\]](#)
9. Ingvarsen, K.L.; Dewhurst, R.J.; Friggens, N.C. On the relationship between lactational performance and health: Is it yield or metabolic imbalance that cause production diseases in dairy cattle? A position paper. *Livest. Prod. Sci.* **2003**, *83*, 277–308. [\[CrossRef\]](#)
10. Oltenacu, P.A.; Broom, D.M. The impact of genetic selection for increased milk yield on the welfare of dairy cows. *Anim. Welf.* **2010**, *19*, 39–49.
11. Ingvarsen, K.L.; Andersen, J.B. Integration of metabolism and intake regulation: A review focusing on periparturient animals. *J. Dairy Sci.* **2000**, *83*, 1573–1597. [\[CrossRef\]](#)
12. Burke, C.; Roche, R. Effects of pasture feeding during the periparturient period on postpartum anovulation in grazed dairy cows. *J. Dairy Sci.* **2007**, *90*, 4304–4312. [\[CrossRef\]](#)
13. Enjalbert, F.; Nicot, M.C.; Bayourthe, C.; Moncoulon, R. Ketone bodies in milk and blood of dairy cows: Relationship between concentrations and utilization for detection of subclinical ketosis. *J. Dairy Sci.* **2001**, *84*, 583–589. [\[CrossRef\]](#)
14. Giannesella, M.; Perillo, L.; Fiori, E.; Giudice, E.; Zumbo, A.; Morgante, M.; Piccione, G. Transition period in healthy and diseased dairy cows: Evaluation of metabolic modifications. *Large Anim. Rev.* **2018**, *24*, 107–111.
15. Cecilian, F.; Lecchi, C.; Urh, C.; Sauerwein, H. Proteomics and metabolomics characterizing the pathophysiology of adaptive reactions to the metabolic challenges during the transition from late pregnancy to early lactation in dairy cows. *J. Proteom.* **2018**, *178*, 92–106. [\[CrossRef\]](#)
16. Zhang, H.; Wu, L.; Xu, C.; Xia, C.; Sun, L.; Shu, S. Plasma metabolomic profiling of dairy cows affected with ketosis using gas chromatography/mass spectrometry. *BMC Vet. Res.* **2013**, *9*, 186–199. [\[CrossRef\]](#) [\[PubMed\]](#)

17. Imhasly, S.; Naegeli, H.; Baumann, S.; von Bergen, M.; Luch, A.; Jungnickel, H.; Potratz, S.; Gerspach, C. Metabolomic biomarkers correlating with hepatic lipidosis in dairy cows. *BMC Vet. Res.* **2014**, *10*, 122. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Dervishi, E.; Zhang, G.; Dunn, S.M.; Mandal, R.; Wishart, D.S.; Ametaj, B.N. GC-MS metabolomics identifies metabolite alterations that precede subclinical mastitis in the blood of transition dairy cows. *J. Proteom. Res.* **2017**, *16*, 433–446. [\[CrossRef\]](#)
19. Zhang, G.; Dervishi, E.; Dunn, S.M.; Mandal, R.; Liu, P.; Han, B.; Wishart, D.S.; Ametaj, B.N. Metabotyping reveals distinct metabolic alterations in ketotic cows and identifies early predictive serum biomarkers for the risk of disease. *Metabolomics* **2017**, *13*, 43–58. [\[CrossRef\]](#)
20. Zhang, G.; Ametaj, B.N. Ketosis: An old story under a new approach. *Dairy* **2020**, *1*, 42–60. [\[CrossRef\]](#)
21. Sun, L.W.; Zhang, H.Y.; Wu, L.; Shu, S.; Xia, C.; Xu, C.; Zheng, J.S. ¹H-Nuclear magnetic resonance-based plasma metabolic profiling of dairy cows with clinical and subclinical ketosis. *J. Dairy Sci.* **2014**, *97*, 1552–1562. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Xu, C.; Sun, L.-W.; Xia, C.; Zhang, H.-Y.; Zheng, J.-S.; Wang, J.-S. ¹H-Nuclear magnetic resonance-based plasma metabolic profiling of dairy cows with fatty liver. *Asian Australas. J. Anim. Sci.* **2016**, *29*, 219–229.
23. Dervishi, E.; Zhang, G.; Hailemariam, D.; Mandal, R.; Wishart, D.S.; Ametaj, B.N. Urine metabolic fingerprinting can be used to predict the risk of metritis and highlight the pathobiology of the disease in dairy cows. *Metabolomics* **2018**, *14*, 83. [\[CrossRef\]](#)
24. Zhang, G.; Dervishi, E.; Zwierchowski, G.; Mandal, R.; Wishart, D.S.; Ametaj, B.N. Urinary metabolomics around parturition identifies metabolite alterations in dairy cows affected postpartum by lameness: Preliminary study. *Dairy* **2020**, *1*, 6–19. [\[CrossRef\]](#)
25. Hailemariam, D.; Mandal, R.; Saleem, F.; Dunn, S.M.; Wishart, D.S.; Ametaj, B.N. Identification of predictive biomarkers of disease state in transition dairy cows. *J. Dairy Sci.* **2014**, *97*, 2680–2693. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Hailemariam, D.; Mandal, R.; Saleem, F.; Dunn, S.M.; Wishart, D.S.; Ametaj, B.N. Metabolomics approach reveals altered amino acid and sphingolipid profiles associated with pathological state in transition dairy cows. *Curr. Metab.* **2014**, *2*, 184–195. [\[CrossRef\]](#)
27. Kenéz, Á.; Dánicse, S.; Rolle-Kampczyk, U.; von Bergen, M.; Huber, K. A metabolomics approach to characterize phenotypes of metabolic transition from late pregnancy to early lactation in dairy cows. *Metabolomics* **2016**, *12*, 165–176. [\[CrossRef\]](#)
28. Luo, Z.Z.; Shen, L.H.; Jiang, J.; Huang, Y.X.; Bai, L.P.; Yu, S.M.; Yao, X.P.; Ren, Z.H.; Yang, Y.X.; Cao, S.Z. Plasma metabolite changes in dairy cows during parturition identified using untargeted metabolomics. *J. Dairy Sci.* **2019**, *102*, 1–12. [\[CrossRef\]](#)
29. Luke, T.D.W.; Pryce, J.E.; Elkins, A.C.; Wales, W.J.; Rochfort, S.J. Use of large and diverse datasets for ¹H NMR serum metabolic profiling of early lactation dairy cows. *Metabolites* **2019**, *10*, 180. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Fariña, S.R.; Chillibroste, P. Opportunities and challenges for the growth of milk production from pasture: The case of farm systems in Uruguay. *Agric. Syst.* **2019**, *176*, 102631. [\[CrossRef\]](#)
31. Kolver, E.S.; Muller, L.D. Performance and nutrient intake of high producing Holstein cows consuming pasture or a total mixed ration. *J. Dairy Sci.* **1998**, *81*, 1403–1411. [\[CrossRef\]](#)
32. Duffield, T.F.; Leslie, K.E.; Sandals, D.; Lissimore, K.; McBride, B.W.; Lumsden, J.H.; Dick, P.; Bagg, R. Effect of prepartum administration of a monensin controlled release capsule on cow health and reproduction. *J. Dairy Sci.* **1999**, *82*, 2377–2384. [\[CrossRef\]](#)
33. LeBlanc, S. Postpartum uterine disease and dairy herd reproductive performance: A review. *Vet. J.* **2008**, *176*, 102–114. [\[CrossRef\]](#)
34. Galvão, K.N.; Greco, L.F.; Vilela, J.M.; Sá Filho, M.F.; Santos, J.E. Effect of intrauterine infusion of ceftiofur on uterine health and fertility in dairy cows. *J. Dairy Sci.* **2009**, *92*, 1532–1542. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Ruegg, P.L. New perspectives in udder health management. *Vet. Clin. Food Anim.* **2012**, *28*, 149–163. [\[CrossRef\]](#)
36. Duffield, T.F.; Lissimore, K.D.; McBride, B.W.; Leslie, K.E. Impact of hyperketonemia in early lactation dairy cows on health and production. *J. Dairy Sci.* **2009**, *92*, 571–580. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Nicholson, J.K.; Foxall, P.J.D.; Spraul, M.; Farrant, R.D.; Lindon, J.C. 750 MHz ¹H and ¹³C NMR spectroscopy of human blood plasma. *Anal. Chem.* **1995**, *67*, 793–811. [\[CrossRef\]](#)
38. Viant, M.R. Improved methods for the acquisition and interpretation of NMR metabolomic data. *Biochem. Biophys. Res. Commun.* **2003**, *310*, 943–948. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Purohit, P.V.; Rocke, D.M.; Viant, M.R.; Woodruff, D.L. Discrimination models using variance-stabilizing transformation of metabolomic NMR data. *Omics* **2004**, *8*, 118–130. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Van den Berg, R.A.; Hoefsloot, H.C.J.; Westerhuis, J.A.; Smilde, A.K.; van der Werf, M.J. Centering, scaling, and transformations: Improving the biological information content of metabolomics data. *BMC Genom.* **2006**, *7*, 142. [\[CrossRef\]](#)
41. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37–52. [\[CrossRef\]](#)
42. Trygg, J.; Wold, S. Orthogonal projections to latent structures (O-PLS). *J. Chemom.* **2002**, *16*, 119–128. [\[CrossRef\]](#)
43. Trygg, J.; Holmes, E.; Lundstedt, T. Chemometrics in metabolomics. *J. Proteom. Res.* **2007**, *6*, 469–479. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Simundic, A.-M. Diagnostic accuracy—Part 1. Basic concepts: Sensitivity and specificity, ROC analysis, STARD statement. *Point Care* **2012**, *11*, 6–8. [\[CrossRef\]](#)
45. Ekelund, S. ROC curves—What are they and how are they used? *Point Care* **2012**, *11*, 16–21. [\[CrossRef\]](#)
46. Ni, Y.; Su, M.; Lin, J.; Wang, X.; Qiu, Y.; Zhao, A.; Chen, T.; Jia, W. Metabolic profiling reveals disorder of amino acid metabolism in four brain regions from a rat model of chronic unpredictable mild stress. *FEBS Lett.* **2008**, *582*, 2627–2636. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Sun, H.Z.; Wang, D.M.; Wang, B.; Wang, J.K.; Liu, H.Y.; Guan, L.L.; Liu, J.X. Metabolomics of four biofluids from dairy cows: Potential biomarkers for milk production and quality. *J. Proteom. Res.* **2015**, *14*, 1287–1298. [\[CrossRef\]](#)
48. Thornalley, P.J. Modification of the glyoxalase system in human red blood cells by glucose in vitro. *Biochem. J.* **1988**, *254*, 751–755. [\[CrossRef\]](#) [\[PubMed\]](#)

49. Omole, O.O.; Nappert, G.; Naylor, J.M.; Zello, G.A. Both L- and D-lactate contribute to metabolic acidosis in diarrheic calves. *J. Nutr.* **2001**, *131*, 2128–2131. [\[CrossRef\]](#)
50. Sako, T.; Urabe, S.; Kusaba, A.; Kimura, N.; Yoshimura, I.; Tazaki, H.; Imai, S.; Ono, K.; Arai, T. Comparison of plasma metabolite concentrations and lactate dehydrogenase activity in dogs, cats, horses, cattle and sheep. *Vet. Res. Commun.* **2007**, *31*, 413–417. [\[CrossRef\]](#)
51. Krauze, K.M.; Oetzel, G.R. Understanding and preventing subacute ruminal acidosis in dairy herds: A review. *Anim. Feed Sci. Technol.* **2006**, *126*, 215–236. [\[CrossRef\]](#)
52. Asanuma, N.; Hino, T. Regulation of fermentation in a ruminal bacterium, *Streptococcus bovis*, with special reference to rumen acidosis. *Anim. Sci. J.* **2002**, *73*, 313–325. [\[CrossRef\]](#)
53. Overton, T.R.; Waldron, M.R. Nutritional management of transition dairy cows: Strategies to optimize metabolic health. *J. Dairy Sci.* **2004**, *87*, E105–E119. [\[CrossRef\]](#)
54. Aschenbach, J.R.; Kristensen, N.B.; Donkin, S.S.; Hammon, H.M.; Penner, B.G. Gluconeogenesis in dairy cows: The secret of making sweet milk from sour dough. *IUBMB Life* **2010**, *62*, 869–877. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Gonz  les, F.D.; Mu  o, R.; Pereira, V.; Campos, R.; Bendito, J.L. Relationship among blood indicators of lipomobilization and hepatic function during early lactation in high-yielding dairy cows. *J. Vet. Sci.* **2011**, *12*, 251–255. [\[CrossRef\]](#)
56. Vercoe, J.E. Urinary allantoin excretion and digestible dry-matter intake in cattle and buffalo. *J. Agric. Sci.* **1976**, *86*, 613–615. [\[CrossRef\]](#)
57. Koch, F.; Lamp, O.; Eslamizad, M.; Weitzel, J.; Kuhla, B. Metabolic response to heat stress in late-pregnant and early lactation dairy cows: Implications to liver-muscle crosstalk. *PLoS ONE* **2016**, *11*, e0160912. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Bell, A.W. Regulation of organic nutrient metabolism during transition from late pregnancy to early lactation. *J. Anim. Sci.* **1995**, *73*, 2804–2819. [\[CrossRef\]](#)
59. Hayirli, A.; Grummer, R.R.; Nordheim, E.V.; Crump, P.M. Animal and dietary factors affecting feed intake during the prefresh transition period in Holsteins. *J. Dairy Sci.* **2002**, *85*, 3430–3443. [\[CrossRef\]](#)
60. Grant, R.J.; Albright, J.L. Effect of animal grouping on feeding behavior and intake of dairy cattle. *J. Dairy Sci.* **2001**, *85*, E156–E163. [\[CrossRef\]](#)
61. Meikle, A.; de Brun, V.; Carriquiry, M.; Soca, P.; Sosa, C.; Adrien, M.L.; Chilbroste, P.; Abecia, J.A. Influences of nutrition and metabolism on reproduction of the female ruminant. *Anim. Reprod.* **2018**, *15*, 899–911. [\[CrossRef\]](#)
62. Kim, H.S.; Kim, E.D.T.; Eom, J.S.; Choi, Y.Y.; Lee, S.J.; Lee, S.S.; Chung, C.D.; Lee, S.S. Exploration of metabolite profiles in the biofluids of dairy cows by proton nuclear magnetic resonance analysis. *PLoS ONE* **2021**, *16*, e0246290.