



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

**FACULTAD DE QUÍMICA**

**ANÁLISIS METABOLÓMICO EN ORINA DE JÓVENES  
MEXICANAS DESPUÉS DEL CONSUMO DE JUGO DE  
ARÁNDANO O MANZANA. ESTUDIO POR  
ESPECTROSCOPIA DE RESONANCIA MAGNÉTICA  
NUCLEAR.**

**TESIS**

**QUE PARA OBTENER EL TITULO DE**

**QUÍMICA DE ALIMENTOS**

**PRESENTA**

**CIRCE CASANDRA HERNANDEZ ESPINO**



**CIUDAD UNIVERSITARIA, CDMX**

**AÑO 2017**



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

## **JURADO ASIGNADO**

<b>PRESIDENTE:</b>	M. en C. Francisco Rojo Callejas
<b>VOCAL:</b>	M. en C. Lucía Cornejo Barrera
<b>SECRETARIO:</b>	Dra. Nuria Esturau Escofet
<b>1er SUPLENTE:</b>	Dr. Eduardo Rodríguez de San Miguel
<b>2do SUPLENTE:</b>	M. en C. Tania Gómez Sierra

### **SITIO DONDE SE DESARROLLÓ EL TEMA:**



#### **INSTITUTO DE QUÍMICA**

Laboratorio Universitario de Resonancia Magnética Nuclear y Laboratorio 2-4

#### **UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

##### ASESOR DEL TEMA

---

Dra. Nuria Esturau Escofet

##### SUPERVISOR TÉCNICO

---

Dra. Beatriz Quiroz García

##### SUSTENTANTE

---

Circe Casandra Hernández Espino

---

CONTENIDO	Página
1. INTRODUCCIÓN .....	1
2. OBJETIVOS E HIPÓTESIS .....	3
2.1. OBJETIVO GENERAL.....	3
2.2. OBJETIVOS PARTICULARES.....	3
2.3. HIPÓTESIS .....	4
3. ANTECEDENTES.....	5
3.1. METABOLÓMICA.....	5
3.1.1. METABOLÓMICA DE ALIMENTOS: FOODOMICS .....	7
3.2. ABORDAJE EXPERIMENTAL EN METABOLÓMICA.....	8
3.2.1. PLATAFORMAS ANALÍTICAS .....	8
3.2.1. PRERARACIÓN DE LAS MUESTRAS .....	15
3.2.2. PROCESADO DE LA SEÑAL DE RMN.....	16
3.2.3. ANÁLISIS ESTADÍSTICO MULTIVARIANTE .....	18
3.2.4. IDENTIFICACIÓN DE BIOMARCADORES .....	24
3.1. ARÁNDANOS Y SUS PROPIEDADES MEDICINALES .....	24
3.1.1. COMPUESTOS ACTIVOS DE LOS ARÁNDANOS .....	26
3.1.2. ESTUDIO METABOLÓMICO PREVIO .....	29
4. METODOLOGÍA .....	31
4.1. COHORTES .....	31
4.2. PREPARACIÓN DE LA MUESTRA.....	32
4.3. ADQUISICIÓN DE ESPECTROS DE RMN.....	33
4.4. PRETRATAMIENTO DE LOS ESPECTROS .....	33
4.5. ANÁLISIS QUIMIOMÉTRICO.....	34
4.6. IDENTIFICACIÓN DE BIOMARCADORES.....	35
4.7. ANÁLISIS UNIVARIANTE .....	36

---

5.	RESULTADOS Y DISCUSIÓN .....	37
5.1.	CARACTERÍSTICAS DE LAS PARTICIPANTES.....	37
5.2.	ESPECTROS DE PROTÓN .....	37
5.2.1.	ALINEACIÓN Y REDUCCIÓN DE LA DIMENSIONALIDAD.....	38
5.2.2.	NORMALIZACIÓN .....	41
5.3.	MODELOS QUIMIOMÉTRICOS.....	41
5.3.1.	ANÁLISIS Y EXCLUSIÓN DE ANÓMALOS.....	41
5.3.2.	SELECCIÓN DE PREPROCESAMIENTO .....	48
5.4.	ANÁLISIS DE LAS MUESTRAS.....	49
5.5.	IDENTIFICACIÓN DE LAS SEÑALES .....	57
5.6.	ANÁLISIS UNIVARIANTE .....	61
6.	CONCLUSIONES .....	65
7.	REFERENCIAS .....	66

## ÍNDICE DE ACRÓNIMOS Y SIMBOLOGÍA

<b>Acrónimo</b>	<b>Significado</b>
1D-NOESY	Espectroscopia Monodimensional de Efecto Nuclear Overhauser ( <i>One Dimensional Nuclear Overhauser Effect Spectroscopy</i> ).
$^1\text{H}$	Protón
$^1\text{H}$ -RMN	Resonancia Magnética Nuclear de Protón
CPMG	Carr Purcel Meighboom Gill
DOSY	Espectroscopia de Difusión ordenada por RMN ( <i>Diffusion-Ordered NMR Spectroscopy</i> )
EM	Espectrometría de Masas
FA	Análisis de Factores ( <i>Factor Analysis</i> )
FID	Caída libre de inducción ( <i>Free induction decay</i> )
CG-EM	Cromatografía de Gases acoplada a Espectrometría de Masas
HSQC	Espectroscopia de Coherencia Heteronuclear con Cuanto Sencillo ( <i>Heteronuclear Single Quantum Coherence Spectroscopy</i> )
CL-EM	Cromatografía de Líquidos acoplada a Espectrometría de Masas
OPLS	Proyección Ortogonal a Estructuras Latentes ( <i>Orthogonal Projection to Latent Structures</i> )
OPLS-DA	Análisis Discriminante de Proyección Ortogonal a Estructuras Latentes ( <i>Orthogonal Projection to Latent Structures Discriminant Assay</i> )
PC	Componente principal ( <i>Principal Component</i> )

---

PCA	Análisis de Componentes Principales ( <i>Principal Component Analysis</i> )
PLS	Mínimos Cuadrados Parciales ( <i>Partial Least Squares</i> )
PLS-DA	Análisis Discriminante por Mínimos Cuadrados Parciales ( <i>Partial Least Squares Discriminant Assay</i> )
RMN	Resonancia Magnética Nuclear
TOCSY	Espectroscopia de Correlación Total ( <i>Total Correlation Spectroscopy</i> )

## 1. INTRODUCCIÓN

Desde la década de los 80 en Japón se estableció en el área Bromatológica la definición de alimento funcional, que es un alimento que además de ser nutritivo y brindar energía, cuando se consume con frecuencia contribuye a mejorar el bienestar de las personas. Bajo esta definición se hicieron famosos los frutos pertenecientes a la familia de las bayas, por su alto contenido de compuestos como flavonoles y proantocianidinas, a los que se les atribuyen efectos antioxidantes.<sup>1</sup> Entre estos frutos, el más famoso es el arándano, que se encuentra comercialmente en su forma de fruto fresco y seco, formando parte de postres y como jugo o concentrado listo para beber. Entre las propiedades que se le atribuyen se encuentra la protección contra las infecciones en las vías urinarias,<sup>2</sup> debido a la presencia de compuestos denominados proantocianidinas tipo B, y tipo A, a cuyos dímeros y trímeros se les relaciona con una propiedad antiadherente de las bacterias en el tracto urinario.<sup>3</sup> El estudio de dichos alimentos, así como su interacción con los seres humanos a través de su metabolismo es un campo de reciente exploración.

Las tecnologías ómicas son un conjunto de tecnologías analíticas empleadas en diferentes disciplinas que permiten generar y agrupar información de un sistema biológico con el fin de obtener una visión lo más completa posible.<sup>4</sup> Entre ellas están la genómica, la proteómica y la metabolómica, que estudian los genes, las proteínas y los metabolitos, respectivamente.

La tecnología ómica de más reciente aparición es la metabolómica, que apunta a medir el conjunto de metabolitos, productos o intermediarios de los procesos

---

<sup>1</sup> Oszmianski, J. et. al. Effect of dried powder preparation process on polyphenolic content and antioxidant capacity of cranberry (*Vaccinium macrocarpon* L.). *Industrial Crops and Products*, **2015**, 77, 658-665.

<sup>2</sup> Blatherwick, N. The specific role of foods in relation to the composition of the urine. *Arch. Inter. Med.* **1914**, 14, 409-450.

<sup>3</sup> Foo, L. Y. et. al. A-type proanthocyanidin trimers from cranberry that inhibit adherence of uropathogenic P-Fimbriated *Escherichia coli*. *J. Nat. Prod.* **2000**, 63, 1225-1228.

<sup>4</sup> Horgan, M.P.; Kenny, L.C. 'Omics' technologies: genomics, transcriptomics, proteomics and metabolomics. *The Obstetritian and Gynaecologist*. **2011**, 13(3), 189-195.



químicos-enzimáticos resultados del metabolismo celular, llamado metaboloma. El metaboloma está comprendido por moléculas de pequeño peso molecular presentes en muestras biológicas (suero, plasma, líquido cefalorraquídeo, orina y tejidos).<sup>5</sup> La metabolómica es capaz de percibir no solo la información propia del organismo estudiado, sino también los impactos ambientales, de estrés y de alimentación. Dicha información se ve reflejada en la presencia, ausencia o cambio en las concentraciones de metabolitos específicos presentes en las muestras.

Actualmente, existen diferentes plataformas analíticas para realizar estudios de metabolómica, siendo las dos principales la Espectroscopia de Masas (EM) acoplada a un método de separación como Cromatografía de Gases, de Líquidos o Electroforesis Capilar y la Resonancia Magnética Nuclear (RMN). Las dos metodologías son complementarias y no intercambiables, y aunque la RMN es menos sensible (rango micromolar) en comparación con la EM, es no destructiva, y por lo que se puede analizar la misma muestra más de una vez. La RMN es muy práctica para la obtención de perfiles masivos de metabolitos en fluidos biológicos, ya que la muestra puede ser evaluada sin ningún pretratamiento.<sup>6,7</sup>

El obtener información significativa a partir de una colección de datos altamente compleja, como son los espectros de RMN, requiere de herramientas matemáticas versátiles y robustas, como los métodos de análisis multivariante, que permiten comparar toda la información colectada al mismo tiempo. Aunada a esta herramienta, se encuentran diversos programas que conjuntan la información proporcionada por el desplazamiento químico y las intensidades relativas de las señales del espectro de RMN y ayudar a la rápida identificación de los metabolitos presentes.

---

<sup>5</sup> 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.

<sup>6</sup> Duarte, I.F.; Diaz, S.O. and Gil, A.M. NMR metabolomics of human blood and urine in disease research. *J. Pharm. Biomed. Anal.* **2014**, 93, 17–26

<sup>7</sup> Scrivo R, Casadei L, Valerio M, et al. Metabolomics approach in allergic and rheumatic diseases. *Curr. Allergy Asthma. Rep.* **2014**, 14(6), 445

En 2015 el grupo de Gu en la universidad de Florida publicó una investigación acerca de los cambios metabólicos causados, en mujeres jóvenes americanas, por el consumo de jugo de arándano o jugo de manzana utilizando la plataforma metabolómica basada en RMN.<sup>8</sup> En dicho estudio encontraron un incremento en la excreción urinaria de ácido hipúrico y en suero hallaron un aumento en el nivel de ácido cítrico y decremento en el de lactato y D-glucosa. También encontraron diferencias en la concentración de dos metabolitos no identificados, uno en plasma y otro en orina. En el presente trabajo de tesis, se realizó un estudio similar en orina de mujeres jóvenes mexicanas con la intención de: i) implementar la plataforma metabolómica por RMN, que no tiene precedente en nuestro país y ii) evaluar la diferencia entre muestras de orina tras el consumo de jugo de arándano o jugo de manzana, buscando reproducir parcialmente el estudio realizado por Gu y colaboradores e indagar similitudes o diferencias entre las poblaciones.

## 2. OBJETIVOS E HIPÓTESIS

### 2.1. OBJETIVO GENERAL

Implementar la plataforma metabolómica basada en la espectroscopia de <sup>1</sup>H RMN, para establecer las diferencias metabólicas entre las muestras de orina de dos grupos de mujeres jóvenes mexicanas con dos regímenes alimentarios diferentes: el consumo de jugos de arándano o de manzana.

### 2.2. OBJETIVOS PARTICULARES

1. Realizar la planificación del diseño experimental, seleccionar a las candidatas para el estudio, realizar las consideraciones de los procedimientos estandarizados y optimizados para la recolección y el almacenamiento de las muestras y garantizar la calidad de los experimentos.

---

<sup>8</sup> Liu, H.; Tayyari, F.; Khoo, C.; Gu, L. A. <sup>1</sup>H NMR-based approach to investigate metabolomic differences in the plasma and urine of young women after cranberry juice or apple juice consumption. *J. Funct. Foods*, **2015**, *14*, 76-86.

2. Obtener el perfil metabólico de las muestras de orina de las participantes mediante  $^1\text{H}$  RMN.
  - 2.1. Realizar la preparación de las muestras para su análisis por RMN.
  - 2.2. Adquirir con la secuencia adecuada (1D NOESY) los espectros en un equipo Bruker de 700 MHz.
  - 2.3. Realizar diferentes tipos de procesado en los espectros, haciendo uso de alineación y reducción de dimensionalidad por construcción de cajas (*binning*), con dos tipos de normalización del espectro, para determinar posteriormente, por medio de métodos quimiométricos, cuál es el mejor procesado con base en su valor de predicción  $Q^2$ .
3. Estratificar a las participantes con base en su perfil metabólico, utilizando métodos quimiométricos para la generación de modelos de agrupamiento y discriminación
4. Realizar el análisis de biomarcadores
  - 4.1. Identificar los metabolitos responsables de la segregación entre los grupos de estudio.
  - 4.2. Comparar los resultados con lo reportado por el equipo de Gu en Florida.
  - 4.3. Identificar el compuesto “desconocido 4” reportado por Gu y colaboradores.

### 2.3. HIPÓTESIS

Al igual que lo reportado por el equipo de trabajo de Gu en Florida, se espera que el grupo de mujeres que consuma jugo de arándano, tenga un alto contenido de ácido hipúrico y la presencia del compuesto *desconocido 4* (2.11 ppm) en orina, señalando a estos compuestos como posibles biomarcadores de consumo de jugo de arándano en ambas poblaciones de mujeres, americana y mexicana.

### 3. ANTECEDENTES

En este capítulo se describen las aplicaciones actuales de la metabolómica con las revisiones más importantes y recientes en las áreas de biología de sistemas y biomedicina, dando especial énfasis al impacto de la dieta. Se detalla cómo se aborda experimentalmente un proyecto mediante la plataforma metabolómica por RMN. También se indican las propiedades del arándano, un alimento funcional al que se le atribuyen propiedades benéficas contra algunos padecimientos.

#### 3.1. METABOLÓMICA

En biomedicina, el término *ómica* se usa para englobar todas aquellas disciplinas, tecnologías y áreas de investigación que estudian el conjunto o la totalidad de un sistema biológico.<sup>9</sup> El ejemplo más conocido y utilizado es la *genómica*, que se encarga de estudiar al genoma, o conjunto de genes. Así, una investigación en el área genómica, puede incluir la secuenciación del genoma, la regulación de éste a nivel funcional, el estudio de mutaciones y modificaciones genéticas, etc. Otras ómicas de interés son la *transcriptómica* (estudio de la totalidad de genes transcritos en un organismo), la *proteómica* (estudio del conjunto de proteínas existentes en un sistema biológico) y la que hasta hace poco tiempo se hizo presente: la *metabolómica*, que se encarga del estudio del metabolismo de un organismo vivo a partir del análisis de sus metabolitos.

La metabolómica permite identificar cambios sutiles en los perfiles metabólicos de sistemas biológicos en diferentes estados fisiológicos o patológicos. En este sentido, a diferencia de la genómica y la proteómica, la metabolómica permite la

---

<sup>9</sup> Medicina molecular. Glosario. <http://medmol.es/glosario/omicas/> (accesado el 31 de enero del 2016)

medición del impacto ambiental, la alimentación<sup>10</sup>, el ejercicio<sup>11</sup> o el estrés<sup>12</sup> sobre el sistema de un organismo. Así, el metaboloma se puede entender como el producto final de la expresión génica, teniendo a su vez un posible impacto en el fenotipo de una célula, tejido y organismo.<sup>13</sup>

Los metabolitos estudiados son moléculas de bajo peso molecular (<1500 Dalton) que intervienen en los procesos celulares y revelan cómo está funcionando el metabolismo de un órgano determinado o de un ser vivo. Generalmente se hace referencia a hidratos de carbono, aminoácidos, lípidos, etc., como metabolitos, aunque no son los únicos. El número de metabolitos depende del sistema biológico estudiado. En RMN, típicamente en células, tejidos o muestras de sangre se detectan entre 20 y 60 metabolitos.<sup>14</sup>

Dependiendo del campo de estudio y el enfoque utilizado, la metabolómica puede hacer uso de diferentes estrategias, entre las cuales las más importantes son las siguientes:

- Perfil metabólico (*Metabolic profiling*), se centra en el análisis cuantitativo de un conjunto de metabolitos predefinidos, relacionados con una determinada ruta metabólica.
- Huella metabólica (*Metabolic fingerprinting*), análisis rápido, global de las muestras permitiendo su clasificación, habitualmente sin cuantificación ni identificación.

---

<sup>10</sup> Barron, J.; Bermingham, K.; et. al. Twin metabolomics: the key to unlocking complex phenotypes in nutrition research. *Nutr. Res.* **2016**, 36, 291-304.

<sup>11</sup> Verástegui-Escolano, C. Efectos de la acupuntura sobre la fatiga inducida por ejercicios físicos exhaustivos: una investigación metabolómica. *Revista Internacional de Acupuntura.* **2015**, 9(4), 131-133.

<sup>12</sup> Zhou, X., et. al. Metabolomics identifies perturbations in amino acid metabolism in the prefrontal cortex of the learned helplessness rat model of depression. *Neuroscience.* **2017**, 343, 1-9.

<sup>13</sup> Fiehn, O. Metabolomics-the link between genotypes and phenotypes. *Plant Mol. Biol.* **2002**. 48. 155–171.

<sup>14</sup> Nagana-Gowda, G.A.; Raftery, D. Can NMR solve some significant challenges in metabolomics? *J. Magn. Reson.* **2015**, 260, 144-160.

- Análisis de uno o varios metabolitos (*Metabolite target analysis*), análisis cualitativo y cuantitativo de uno o varios metabolitos relacionados con una reacción metabólica específica.
- Exometabolómica (*Metabolite footprinting, Exometabolomics*), estudio de metabolitos en fluidos extracelulares.
- Metalometabolómica, estudio de metabolitos que presentan un metal o metaloide en su estructura (ej. Selenometabolómica, Ferrometabolómica).

El objetivo de dichos estudios es analizar un número de metabolitos y seleccionar los que verdaderamente aportan información útil. Debido a la enorme complejidad y dinamismo de los metabolomas, su investigación requiere de técnicas analíticas reproducibles, robustas, sensibles y que no requieran un pretratamiento complejo. Entre las técnicas más utilizadas se encuentran: espectroscopia de infrarrojo con transformada de Fourier<sup>15</sup>, EM acoplada a sistemas cromatográficos<sup>16</sup> y la RMN.

### 3.1.1. METABOLÓMICA DE ALIMENTOS: *FOODOMICS*

La metabolómica de los alimentos (*foodomics*) puede ser definida como la aplicación de la metabolómica a sistemas alimentarios, incluyendo recursos alimenticios, procesamiento de alimentos y dietas humanas. El estudio de la metabolómica en los alimentos ha incrementado gradualmente en años recientes debido a que los sistemas alimentarios están directamente relacionados a la nutrición y la salud humana.<sup>17</sup> Entre los objetivos principales está el definir, mediante la aplicación de los enfoques ómicos, la suma de los diferentes efectos del procesamiento, digestión y absorción de los alimentos, sobre la nutrición y la salud humana (Figura 1).<sup>18</sup>

---

<sup>15</sup> Kuligowski, J.; Pérez-Guaita, D.; et. al. Infrared biospectroscopy for a fast qualitative evaluation of sample preparation in metabolomics. *Talanta*. **2014**, 127, 181-190.

<sup>16</sup> Lu, X.; et. al. LC-MS-based metabolomics analysis. *J. Chromatogr.* **2008**, 866, 64-76.

<sup>17</sup> Kim, S.; Kim, J.; Ju, E.; Kim, K.H. Food metabolomics: from farm to human. *Curr. Opin. Biotechnol.* **2015**, 37, 16-23.

<sup>18</sup> Bordoni, A.; Capozzi, F. The foodomics approach for discovering biomarkers of food consumption in nutrition studies. *Curr. Opin. Food Sci.* **2015**, 4, 124-128.

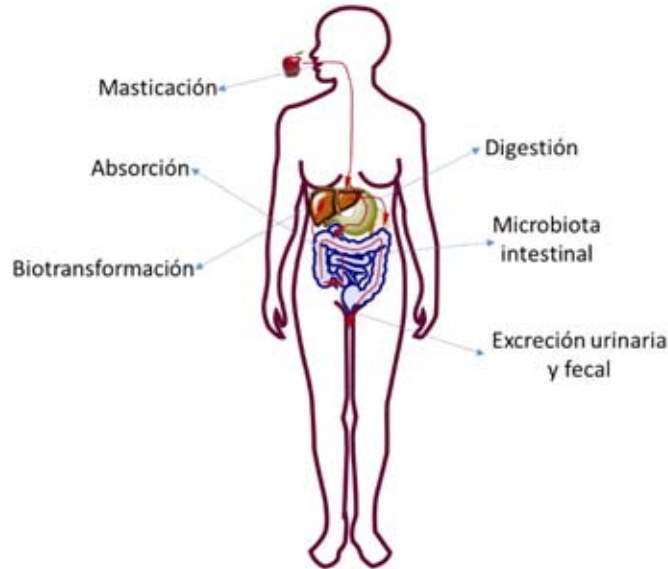


Figura 1. Proceso de absorción de los nutrientes contenidos en el alimento.

## 3.2. ABORDAJE EXPERIMENTAL EN METABOLÓMICA

### 3.2.1. PLATAFORMAS ANALÍTICAS

La EM, generalmente acoplada a técnicas cromatográficas GC y LC, permite analizar una gran cantidad de metabolitos a sensibilidades muy grandes, y eligiendo el método de separación adecuado, se consigue una gran selectividad. Sin embargo, existen ciertas limitaciones en la falta de estandarización de los métodos y un tratamiento de muestra tedioso, que hacen los experimentos poco reproducibles, sin mencionar lo destructivo de la técnica.<sup>19</sup>

La RMN es una herramienta más versátil y robusta que permite la medida de los metabolitos de forma fiable y repetitiva, con un nivel de automatización muy alto. No es un tipo de análisis destructivo y requiere un pretratamiento mínimo o nulo de las

<sup>19</sup> Dunn, W.B., Ellis, D.I. Metabolomics: Current analytical platforms and methodologies. *Trends Anal. Chem.* **2005**, 24(4), 285-294.

muestras. El campo magnético en los equipos de RMN utilizados en metabolómica es en general muy alto, llegando a ser superior a los 10 Teslas. Hoy en día, la aparición de las sondas criogénicas ha permitido llegar a niveles de sensibilidad de mg/mL. Entre sus desventajas se encuentra la superposición de señales provenientes de diferentes metabolitos, y una baja sensibilidad con respecto a la EM.

### 3.2.1.1. RESONANCIA MAGNÉTICA NUCLEAR

La RMN es una técnica de análisis instrumental no destructiva que se basa en la propiedad de absorción de energía electromagnética en la región de las radiofrecuencias por parte de algunos núcleos atómicos, siendo un requisito que el número de spin de dichos núcleos sea diferente de cero.

Estas absorciones son discretas y dependen del tipo de núcleo y del ambiente químico en el que se encuentran, obteniéndose dicho efecto solamente bajo el influjo de un campo magnético intenso.

En circunstancias normales (fuera del influjo de un campo magnético) el momento magnético de los núcleos está orientado en el espacio al azar. Pero, al colocar un campo magnético  $B_0$ , los núcleos se alinean con dicho campo, ya sea a favor o en contra de éste, siguiendo la distribución de Boltzmann, describiendo un movimiento de precesión a una frecuencia proporcional al campo magnético, conocida como frecuencia de Larmor. Esta frecuencia depende de la constante giromagnética es una propiedad intrínseca del núcleo.

Para poder hacer que los núcleos entren en resonancia, se les aplica una radiofrecuencia, similar a la frecuencia en que precesan, perpendicular a la dirección de la magnetización, sacándolos de su equilibrio y observando la señal que generan al regresar a su estado original.

El sacar del equilibrio a los núcleos implica, en el marco del modelo vectorial, que la magnetización deja de estar orientada sobre el eje +z (estado de equilibrio), para



desplazarse al plano  $xy$ . El tiempo que tarda la magnetización en regresar al eje  $+z$  es conocido como *relajación*, distinguiéndose dos tipos de relajación:

- La relajación sobre el *eje z* hasta alcanzar el equilibrio es denominado *relajación longitudinal* ( $T_1$ ) o espín-red, e involucra intercambios energéticos entre los espines nucleares y su entorno molecular (*red*).
- La relajación sobre el *plano xy* es llamada *relajación transversal* ( $T_2$ ) o espín-espín, que involucra intercambios energéticos entre espines a través de los enlaces.

El núcleo del hidrógeno,  $^1\text{H}$ , es uno de los núcleos más ampliamente estudiados en RMN, debido a que posee una constante giromagnética muy grande y es muy abundante en la naturaleza.

El equipo con el que se detecta el fenómeno de RMN es un espectrómetro, como el mostrado en la Figura 2. Este consta de varias partes, entre las cuales, las más importantes son:

1. Un imán superconductor que genere un campo magnético intenso, estable y homogéneo.
2. Una sonda compuesta de bobinas para excitación y detección de los espines nucleares.
3. Un transmisor de pulsos de radiofrecuencia para generar la excitación de los núcleos.
4. Un amplificador para amentar la respuesta.
5. Un filtro digital para convertir la señal en información.
6. Una computadora que controle el sistema de procesamiento de la señal.



Figura 2. Espectrometro Bruker Avance III de 700 MHz con sistema Sample Jet integrado.

La transmisión de las radiofrecuencias se hace a partir de pulsos de radiofrecuencias. Un pulso consiste en la aplicación de una radiofrecuencia por un periodo de tiempo definido ( $t_p$ ).

El compuesto más abundante en estudios de metabolómica es el agua, que es el principal componente de los biofluidos y tejidos. El agua presenta una intensa señal en los espectros de RMN que impide la observación de las señales correspondientes a los metabolitos. Por lo tanto, la supresión de agua es crucial para poder observar las señales más pequeñas y obtener espectros interpretables. Hay numerosas secuencias de pulsos que se usan en RMN para minimizar esta intensa señal, entre ellas se encuentran: <sup>19</sup>

1. Presaturación (presat): es el procedimiento más simple para minimizar y suprimir la intensa resonancia del agua. Se basa en aplicar un pulso de baja potencia, relativamente largo, a la resonancia del disolvente (para suprimir

su señal), antes de aplicar el pulso duro (usualmente de  $90^\circ$ ), no selectivo con el fin de excitar las señales de todo el espectro (Figura 3).

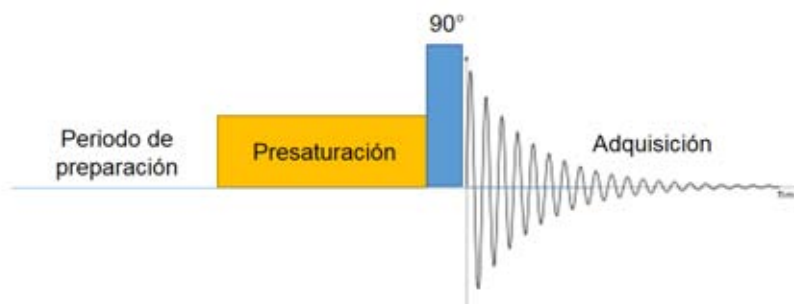


Figura 3. Secuencia de pulsos Presat

2. 1D NOESY-presat (*1 Dimension Nuclear Overhauser Effect Spectroscopy with Presaturation*)<sup>20</sup>: Esta secuencia es muy utilizada en el análisis de muestras de orina, presenta una supresión más eficiente de la señal de agua con mejor línea base que el experimento presat. Esta secuencia toma ventaja de la diferencia de tiempo de relajación longitudinal ( $T_1$ ) entre los núcleos del soluto y el solvente (usualmente el tiempo de relajación del agua es mayor que el de los metabolitos). La secuencia de pulsos se muestra en la Figura 4, esta comienza con la aplicación de un pulso de muy baja potencia sobre la intensa señal del agua para suprimirla. La señal residual del agua es posteriormente suprimida en una secuencia NOESY, donde los primeros dos pulsos de  $90^\circ$  de 1D NOESY invierten la magnetización (de los metabolitos y el solvente) al eje  $-z$ . Durante el tiempo de mezclado ( $t_m$ ) los metabolitos se relajan al eje  $+z$  (estado de equilibrio). El tiempo de mezclado está optimizado para que la magnetización residual del agua sea nula al momento del último pulso de  $90^\circ$ . Su uso se ha popularizado en metabolómica debido a una baja necesidad de optimización, aunado a que se evita la pérdida de protones

<sup>20</sup> McKay, R.T. How the 1D-NOESY suppresses solvent signal in metabolomics NMR spectroscopy: An examination of the pulse sequence. *Concepts Magn. Reson.* **2011**, 38A (5), 197-220.

intercambiables.<sup>21</sup> En un instrumento Bruker, la secuencia es llamada “*noesypr1d*”.

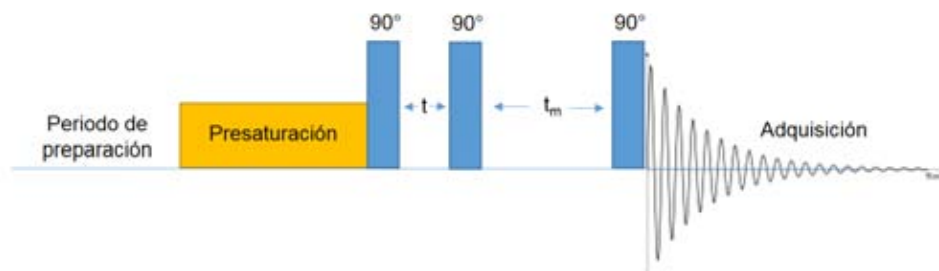


Figura 4. Secuencia de pulsos 1D NOESY.

3. CPMG (“*Car Purcel Meiboom Gill*”) con presaturación de agua:<sup>22, 23</sup> Esta secuencia es muy utilizada en el análisis de muestras de plasma y suero, ya que elimina las señales anchas producidas por las proteínas y los lípidos que distorsionan parcialmente la línea base y ocultan las señales de los metabolitos. CPMG es una secuencia que hace uso de las diferencias en los  $T_2$  entre las proteínas y los metabolitos. En la Figura 5 se muestra la secuencia de pulsos que comienza con la aplicación de un pulso de baja potencia para suprimir la señal del agua, seguida de un pulso de  $90^\circ$  que lleva la magnetización del eje  $+z$  al plano  $xy$ , un tiempo  $t$  donde los espines se desfasan, y un segundo pulso de  $180^\circ$  para reenfocar la magnetización. Este reenfoque *eco de espín* se repite  $n$  veces. Durante el periodo del eco la magnetización reenfocada decae debido al  $T_2$ , por lo que al final del periodo de reenfoque, la magnetización de los espines con un  $T_2$  corto es suprimida significativamente, mientras que la magnetización de los espines que poseen  $T_2$  más largos es detectada. ¡Error! Marcador no definido.

<sup>21</sup> Teng, Q. Chapter 9. NMR-based metabolomics. Structural biology: NMR Applications. 2<sup>nd</sup> ed. Springer. 2013. 337, 339-340.

<sup>22</sup> Car, H.Y.; Purcell, E.M. Effects of Diffusion on Free Precession In Nuclear Magnetic Resonance Experiments. *Phys. Rev.* **1954**, 94, 630-638.

<sup>23</sup> Meighboom, S.; Gill, D. Modified Spin Echo Method for Measuring Nuclear Relaxation Times. *Rev. Sci. Instrum.* **1958**, 9 (8), 688-691.

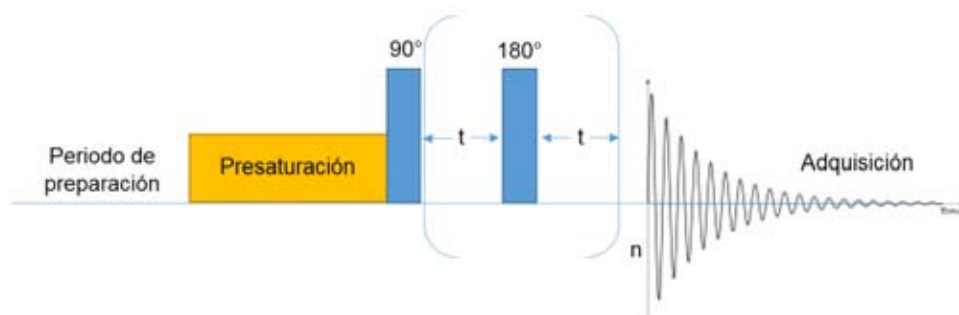


Figura 5. Secuencia de pulsos CPMG.

4. J-resuelta con presaturación de agua:<sup>24</sup> Se trata del experimento homonuclear 2D más sencillo que existe. El método se basa en la transformación del espectro 2D J-resuelto (que separa la información aportada por los acoplamientos escalares a lo largo de la dimensión indirecta F1) a un espectro 1D de protón homodesacoplado en los que la multiplicidad de las señales del espectro se simplifica a singuletes, aumentando la probabilidad de que un metabolito específico aparezca como un pico bien resuelto e identificable. La secuencia de pulsos comienza con la aplicación de un pulso de baja potencia para suprimir la señal del agua, seguido de un eco de espín, donde la duración del eco es modulada como tiempo de evolución (Figura 6).

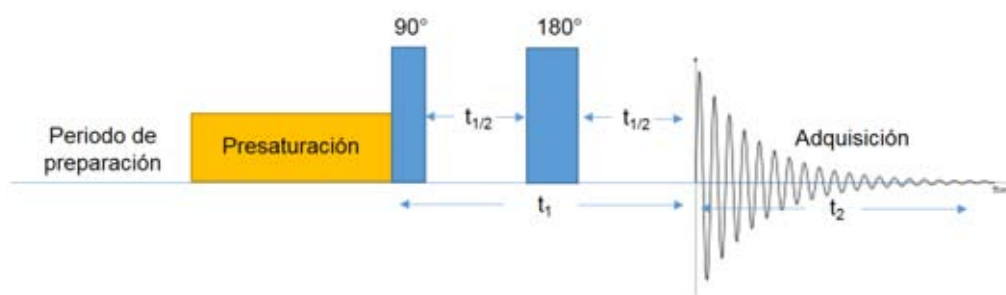


Figura 6. Secuencia de pulsos homonuclear 2D J-resuelta.

<sup>24</sup> Nagayama, K. et. al. Two-dimensional J-Resolved <sup>1</sup>H NMR Spectroscopy for Studies of Biological Molecules. *Biochem. Biol. Res. Commun.* **1977**, 8 (1), 99-105.

5. Difusión editada con presaturación de agua: este tipo de experimento atenúa las señales pertenecientes a moléculas de menor tamaño molecular, debido a que las moléculas más pequeñas se atenúan más rápido que las pertenecientes a moléculas más grandes. Hay varias secuencias para realizar estos experimentos de difusión editada, pero el bloque base en la secuencia es el eco de espín con gradientes, que se muestra en la Figura 7. Comienza con la aplicación de un pulso de baja potencia para suprimir la señal del agua. La función del primer gradiente es la de desfazar la magnetización y la del segundo es reenfoclarla. Estos gradientes de desfase y de reenfoque están separados por el tiempo de difusión ( $\Delta$ ). Si durante este tiempo  $\Delta$  los espines se mueven (difunden) a lo largo del eje z no serán reenfocados por el segundo gradiente y disminuirá la señal en el espectro de RMN. Esta secuencia es ampliamente usada cuando se desea estudiar proteínas o lípidos.

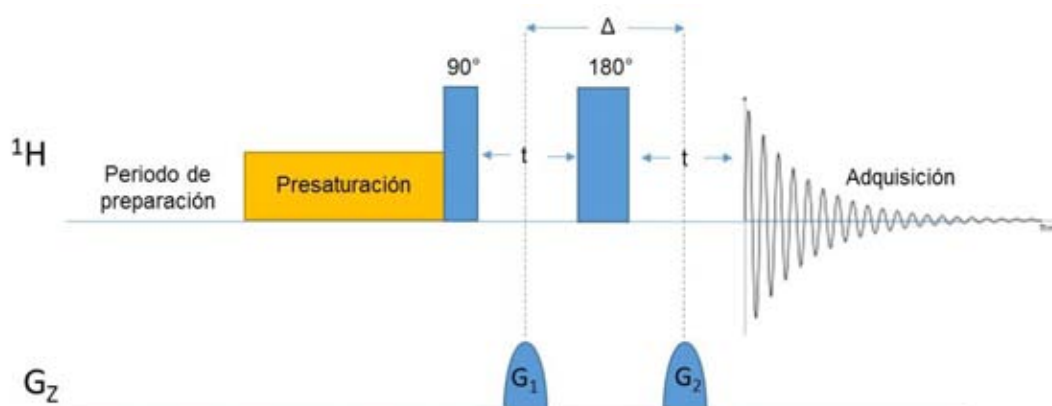
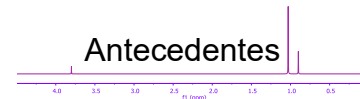


Figura 7. Secuencia Eco de Espín con gradientes.

### 3.2.1. PRERARACIÓN DE LAS MUESTRAS

Hoy en día, con el grado actual de automatización, el mayor porcentaje de errores generados en muchos análisis están en la parte pre analítica, correspondiente a la



preparación de la muestra.<sup>25</sup> En nuestro caso, muchas de las muestras biológicas son líquidas, y pueden ser analizadas directamente en RMN. Este análisis directo minimiza las pérdidas de metabolitos y reduce la preparación al mínimo.<sup>17</sup>

En el caso del estudio de metabolómica de orina por RMN, la señal intensa de agua en aproximadamente 4.8 ppm es fácilmente eliminada usando alguno de los métodos adecuado de supresión de solvente.<sup>26</sup> Con el fin de homogeneizar lo más posible las muestras, se agrega una solución tampón de fosfatos en agua deuterada a las muestras antes de obtener los espectros, para evitar que el cambio en el pH y en la atmosfera iónica de las muestras genere diferencias importantes en los desplazamientos químicos de moléculas con grupos funcionales como ácidos carboxílicos, aminoácidos, y otros metabolitos con propiedades ácido-base.<sup>27</sup> Dicho tampón contiene también un estándar interno, que para el caso de muestras acuosas, es la sal de sodio del ácido 3-trimetilsililpropionico (TSP).<sup>18</sup>

Beckonert y colaboradores,<sup>18</sup> sugieren un método de preparación para muestras de orina de rata y humanos, que consiste en mezclar 400  $\mu$ L de muestra con 200  $\mu$ L de buffer de fosfatos en tubos Eppendorf, se centrifugan a 12000 g durante 5 minutos a 4°C y se transfieren 550  $\mu$ L de muestra a un tubo de RMN.

### 3.2.2. PROCESADO DE LA SEÑAL DE RMN

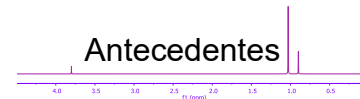
La señal de RMN que se obtiene se procesa para poder aplicar el análisis estadístico de datos multivariados y producir resultados fiables e interpretables. Este procesado del espectro de RMN debe de realizarse de forma muy cuidadosa, debido a que los errores en esta etapa se pueden extender a la interpretación estadística, y por tanto a la interpretación biológica. El procesado en RMN permite transformar

---

<sup>25</sup> Alvaréz-Sánchez, B.; Priego-Capote, F.; Luque-de-Castro, M.D. Metabolomic analysis II. Preparation of biological samples prior to detection. *Trends Anal. Chem.* **2010**, 29(2), 120, 123.

<sup>26</sup> Beckonert, O. et. al. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nature Publishing Group.* **2007**, 2(11), 2692-2703.

<sup>27</sup> Mannina, L.; Sobolev, A.P.; Viel, S. Liquid state <sup>1</sup>H NMR in food analysis. *Prog. Nucl. Magn. Reson. Spectrosc.* **2012**, 66, 1-39.



la señal FID (*Free Induction Decay*) en un espectro, mediante tratamientos matemáticos, como son: la transformada de Fourier (FT), la apodización, la corrección de la fase, el ajuste de la línea base y el ajustar los desplazamientos químicos respecto a la señal de un compuesto de referencia. La apodización consiste en multiplicar la señal FID por funciones envolventes, mejorando la relación señal/ruido.

Algunas opciones extra de preprocesamiento pueden ser usadas con el fin de descartar información que se sabe de antemano sólo genera error en los modelos estadísticos a usar, como la supresión de las señales de determinadas moléculas. Entre estas señales, se suele suprimir la señal del agua (4.6-4.9 ppm), o si es una muestra liposoluble, cualquier solvente que se haya agregado. Otra de las señales que se elimina con recurrencia en las muestras de orina es la señal de la urea (5.4-6.0 ppm).

Para optimizar el tratamiento estadístico se recurre a la normalización de las áreas bajo de la curva de los picos, y en años más recientes, la alineación de los desplazamientos a partir de algoritmos matemáticos para corregir el error intermuestras.<sup>28</sup>

Otro proceso importante es la reducción de la dimensionalidad por generación de cajas (*binning*) con tamaños homogéneos o heterogéneos, que consiste en la subdivisión del espectro, en pequeñas regiones. Este proceso permite obtener la integración de la intensidad de la señal dentro de cada caja para producir un conjunto más pequeño de variables. El tamaño de las regiones tiene que ver con el rango del desplazamiento químico, siendo el tamaño más común el de 0.04 ppm.<sup>29</sup>

---

<sup>28</sup> Cuperlovic-Culf, M. 4. *Metabolomics NMR Data Preprocessing – analysis of individual spectrum. NMR Metabolomics in Cancer Research*. Woodhead Publishing Limited. Cambridge, **2013**; 215-231.

<sup>29</sup> McKenzie, J.S. et. al. Analysis of complex mixtures using high-resolution nuclear magnetic resonance spectroscopy and chemometrics. *Prog. Nucl. Magn. Reson. Spectrosc.* **2011**, 59, 336-359.



### 3.2.3. ANÁLISIS ESTADÍSTICO MULTIVARIANTE

En el análisis de los datos en metabolómica, la variabilidad entre las diferentes muestras es muy difícil de detectar mediante la observación directa de los espectros de RMN y es necesario el uso de técnicas estadísticas multivariantes que permitan evaluar esas diferencias y reconocer patrones químicos en dichos espectros. Estos métodos analíticos reciben el nombre de métodos quimiométricos.

El término *quimiometría* fue introducido en 1972 por Svante Wold y Bruce R. Kowalski, y se define como la disciplina de la química que usa matemáticas, estadística y otros métodos para proveer el máximo de información relevante del análisis de datos químicos.<sup>30</sup>

La adaptación de la quimiometría a la espectroscopia de RMN fue probada en los inicios de la década de 1980 y demostró posteriormente tener un tremendo potencial en la metabolómica. La fusión entre la quimiometría y la bioinformática ha tenido como resultado un impulso muy grande en los últimos años, sobre todo con el desarrollo de bases de datos biológicos enormemente correlacionados con datos metabolómicos, obtenidos por medio de análisis químicos.<sup>31</sup>

Dependiendo del objetivo que persiga el uso de los métodos quimiométricos, se puede dividir a los problemas en descriptivos y predictivos. En los primeros, las propiedades de los sistemas químicos se modelan con la intención de aprender las relaciones subyacentes y la estructura del sistema, mientras que, en los segundos, las propiedades de los sistemas químicos se modelan con la intención de predecir nuevas propiedades o comportamientos de interés.

Entre las herramientas de estadística multivariante o de reconocimiento de patrones, generalmente en primer lugar se emplean los métodos de análisis no supervisados (que permiten evaluar la variabilidad de los datos sin tener en cuenta

---

<sup>30</sup> D.L. Massart, B. G. M. V.; S.N. Deming, Y. Michotte, L. Kaufman, *Chemometrics: A Textbook*. B.G.M. Vandeginste, L. K. **2003**; 5.

<sup>31</sup> Brereton, R. *Applied Chemometrics for Scientists*; Wiley: United Kingdom, **2007**.

su clasificación dentro de ningún grupo), seguido de los modelos supervisados (que utilizan la información sobre la clasificación de las muestras para maximizar la separación entre grupos). Las categorías principales de análisis estadístico multivariado son:

1. El análisis exploratorio de datos (métodos de análisis no supervisados) que se puede utilizar para determinar si existe alguna agrupación intrínseca dentro de un conjunto de datos, sin un conocimiento *a priori* de la clase de las muestras, mediante la reducción de la información a una forma más comprensible para exponer posibles valores atípicos. Generalmente se realiza el análisis de componentes principales (PCA, por sus siglas en inglés, *Principal Component Analysis*).
2. La regresión se utiliza para desarrollar modelos de calibración que correlacionen la información de un conjunto de mediciones o variables regresoras, con la propiedad de interés o variable respuesta. Dos de los algoritmos más comúnmente utilizados para la regresión son Cuadrados Parciales Medios (PLS, por sus siglas en inglés, *Partial Least Squares*) y la regresión de componentes principales (PCR, por su sigla en inglés, *Principal Components Regression*), ambos diseñados para evitar los problemas relacionados con el *ruido* de los datos.
3. Los modelos de clasificación (métodos de análisis supervisados) se aplican en situaciones donde se requiere clasificar a las muestras en categorías previamente definidas. Se asigna la clase a una muestra mediante la comparación de ésta con un conjunto de datos analizados anteriormente, por los que las categorías son ya conocidas. Algunos de los más empleados son el Análisis Discriminante por Mínimos Cuadrados Parciales (PLS-DA, por sus siglas en inglés, *Partial Least Squares Discriminant Analysis*), el Análisis Discriminante por Proyección Ortogonal a Estructuras Latentes (OPLS-DA, por sus siglas en inglés, *Orthogonal to Projection Latent Structures*) y el Análisis Discriminante Lineal (LDA, por sus siglas en inglés, *Linear Discriminant Analysis*).

### 3.2.3.1. ANÁLISIS DE COMPONENTES PRINCIPALES

El PCA es una técnica exploratoria no supervisada para la identificación de patrones y/o tendencias dentro de los grupos de muestras. El PCA descompone la matriz  $\mathbf{X}$ , que contiene el conjunto de datos originales (las filas, que contienen las observaciones  $N$ , y las columnas, que contienen las variables  $K$ ), en dos matrices más pequeñas, llamadas matrices de carga  $\mathbf{P}$  y de puntuación  $\mathbf{T}$ , además de la matriz del error  $\mathbf{E}$  (Figura 8). La matriz de carga  $\mathbf{P}$  contiene información sobre las variables  $K$ , que está compuesta por varios vectores: los Componentes Principales (PC, por sus siglas en inglés *Principal Component*). Los gráficos resultantes de puntuaciones y cargas permiten ver fácilmente la información de las observaciones y las variables.

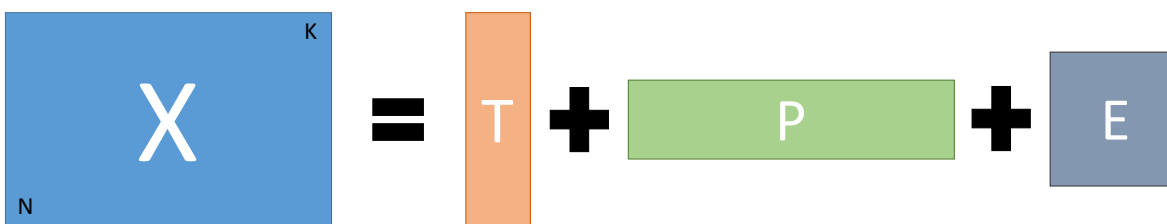


Figura 8. Descomposición matemática de la matriz de datos original ( $X$ ) a las matrices de puntuaciones ( $T$ ), carga ( $P$ ) y error ( $E$ ).

En el espacio  $K$ -dimensional, cada variable es un eje con un largo determinado por el tipo de escalado. En esta clase de técnicas es un requerimiento común el estandarizar los datos hacia un comportamiento de distribución normal con media igual a cero y varianza unitaria. En la práctica se establece un centrado restando el valor de la media de cada medición, y un escalado dividiendo los valores no constantes entre la desviación estándar. Los escalados más comunes usados en metabolómica son:

1. Centrado: resta el valor de la media al valor de la variable.
2. Varianza Unitaria (UV por sus siglas en inglés, *Unit Variance*): resta el valor de la media al valor de la variable, y la divide entre su desviación estándar.

3. Pareto (Par): resta el valor de la media al valor de la variable, y la divide entre la raíz cuadrada de su desviación estándar. Es un tipo de escalado intermedio entre el centrado y la varianza unitaria, y aumenta la importancia de las señales con intensidad relativa media, sin aumentar la aportación del ruido, por lo que es la opción preferida en estudios metabolómicos.

Un PCA encuentra las correlaciones entre las variables  $\mathbf{K}$ , y a través de combinaciones lineales obtiene nuevas variables denominadas *variables latentes* que son independientes entre sí. Cada variable latente es un PC, y estos PC definen un nuevo sistema de coordenadas para el sistema, que describen la variación entre los datos. El primer componente principal (PC1) describe la mayor cantidad de varianza explicable, el segundo (PC2) la siguiente mayor cantidad de varianza explicable, y así sucesivamente.

Geoméricamente hablando, el PCA introduce un componente principal y después un segundo componente principal ortogonal al primero en el espacio K-dimensional original. Cada componente principal pasa por el origen. Finalmente, los puntos son proyectados a un plano con coordenadas PC1, PC2. Su interpretación es que los objetos lejanos en esta proyección son diferentes entre sí, y los objetos cercanos exhiben variaciones similares.

### 3.2.3.2. REGRESIÓN POR MÍNIMOS CUADRADOS PARCIALES

A pesar de que el PCA es un método explorativo muy poderoso, no es garantía de que la variación esté reflejada por la hipótesis propuesta. La medición de metabolitos en biofluidos provee datos con múltiples tipos de variación, relevante e irrelevante, como el género, los hábitos alimenticios, la edad, la genética, la medicación, factores ambientales y el estilo de vida.<sup>32</sup> Estas variaciones pueden ser,

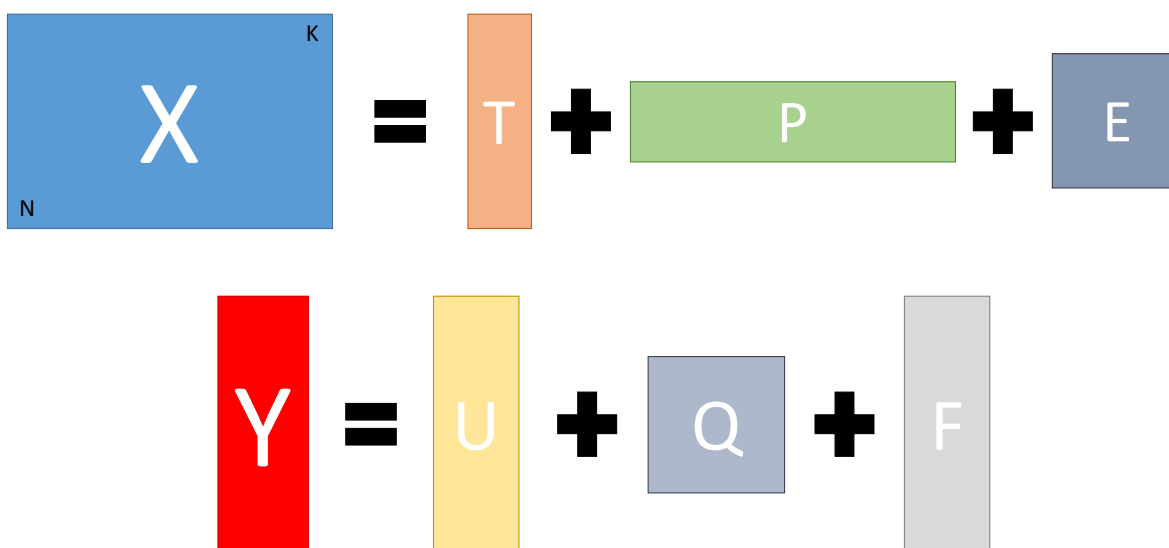
---

<sup>32</sup> Rasmussen, L. G. et. Al. Standarization of factors that influence human urine metabolomics. *Metabolomics*. 2011, 7(1), 71-83.

dependiendo del diseño del estudio, una mayor o menor parte de la variación total, y esto será reflejado en el modelo PCA.

El método PLS es una técnica supervisada que puede ser usada para mejorar la separación entre los grupos de observaciones, rotando los componentes del PCA hasta que se obtenga el máximo grado de separación con base en la información previamente conocida acerca de las muestras.<sup>33</sup> Se basa en la descomposición simultánea de dos matrices: una que contiene las señales analíticas ( $\mathbf{X}$ ) y otra que posee una correspondencia química o descriptiva de las muestras ( $\mathbf{Y}$ ) (Figura 9).

El modelo PLS se ocupa típicamente en situaciones de calibración, donde la información de la matriz  $\mathbf{Y}$  se ocupa para obtener una calibración multivariable, sin embargo, éste uso se puede extender a la generación de un modelo de calibración entre 2 o más tipos de muestras, o sea, una discriminación. La discriminación más sencilla corresponde a un vector  $\bar{Y}$  con unos y ceros describiendo la relación entre las clases.



*Figura 9. Descomposición matemática de las matrices  $\mathbf{X}$  y  $\mathbf{Y}$ . La descomposición de la matriz  $\mathbf{X}$  posee la misma forma que en PCA, mientras que la descomposición de la matriz  $\mathbf{Y}$  es análoga a la de la matriz  $\mathbf{X}$ .*

<sup>33</sup> Breton, R.G. 10. Biological and Medical Applications of Chemometrics. Applied Chemometrics for Scientist. Wiley. 2010, 287-318.

Un aspecto muy importante del uso del PLS es la selección del número de componentes principales a usar y los valores contenidos en  $\mathbf{Y}$  como criterio alternativo. A su vez el modelo debe ser capaz de predecir razonablemente a qué grupo corresponde una muestra desconocida, conocido como la validación del modelo. La técnica involucra la selección del número de PC, la validación y usualmente un grupo de prueba independiente, que contiene muestras que no han sido utilizadas para la construcción del modelo vía validación cruzada.

Adicionalmente, este tipo de modelos generan también gráficos de calificaciones y cargas, fácilmente interpretables, como en el caso de PCA. Las cargas en este caso pueden ayudar a determinar que variables químicas son las que caracterizan cada clase, ayudando en el caso de la identificación de biomarcadores.<sup>33</sup>

El uso de PLS se complica cuando hay más de dos grupos en el análisis. La forma de la asignación de los grupos es crucial, y puede incluir una matriz con una sola columna, o N-1 columnas, donde N es el número de clases.

Además del uso del PLS, el modelo OPLS es una extensión matemática introducida por Trygg y Wold en 2002<sup>34</sup> que ha sido muy importante para la interpretación de datos en las ciencias ómicas. Consta de una nueva forma de descomponer la solución del PLS en componentes correlacionados a  $\mathbf{Y}$ , y componentes únicos en  $\mathbf{X}$ , pero correlacionados ortogonalmente con  $\mathbf{Y}$ . Por lo que es un método de regresión y predicción que encuentra información en la matriz  $\mathbf{X}$  que esta correlacionada a la información conocida en la variable  $\mathbf{Y}$ , y también será juzgado a partir de su capacidad predictiva. Se basa en una proyección, y al igual que PLS busca la mayor separación entre las clases, pero a diferencia de éste, lo hace a partir de la correlación ortogonal entre  $\mathbf{X}$  y  $\mathbf{Y}$ , y al mismo tiempo la búsqueda de componentes que no están correlacionadas con  $\mathbf{Y}$ , pero que varían sistemáticamente en  $\mathbf{X}$  (Figura 10), como problemas instrumentales, problemas de muestreo o diferencia en el estilo de vida, aunque no haya una correlación directa

---

<sup>34</sup> Trygg, J.; Wold, S. Orthogonal Projections to Latent Structures (O-PLS). *J. Chemom.* **2002**, 16, 119-128.

entre ambas. Su interpretación reduce la posibilidad de perder información, mejora la interpretación biológica (como el establecimiento de biomarcadores), y mejora los procedimientos experimentales a usar en un futuro.

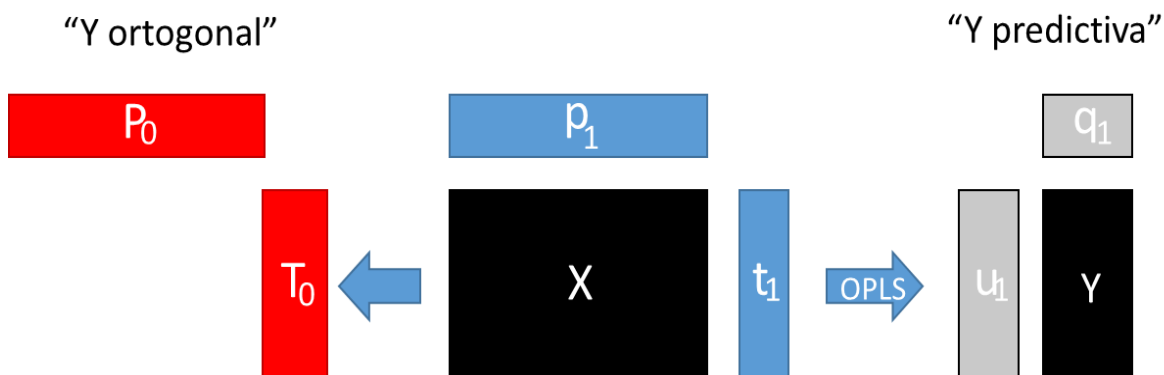


Figura 10. Forma del modelo OPLS. Por un lado realiza la correlación con la matriz  $Y$  y por el otro obtiene las variables latentes entre los datos de  $X$ .

### 3.2.4. IDENTIFICACIÓN DE BIOMARCADORES

Una vez que se tiene obtenidas las señales importantes, se compara con las referencias encontradas en las bases de datos de metabolitos. Las más conocidas son la *Human Metabolome Database* (hmdb.ca) y el *Biological Magnetic Resonance Data Bank* (bmrw.wisc.edu). Si la coincidencia es positiva, se habrá encontrado un biomarcador. En caso de duda, se recurre a la comprobación del analito a partir de la medida con estándares o con la realización de experimentos de RMN de correlación en 2 dimensiones, siendo los más usados el experimento TOCSY (por sus siglas en inglés *Total Correlation Spectroscopy*) que muestra las correlaciones  $^1\text{H}$ - $^1\text{H}$  a través de la red de espín, y el experimento HSQC (por sus siglas en inglés *Heteronuclear Single Quantum Spectroscopy*) que muestra correlación  $^1\text{H}$ - $^{13}\text{C}$  a un enlace de distancia.

## 3.1. ARÁNDANOS Y SUS PROPIEDADES MEDICINALES

El concepto de *alimento funcional* nació en Japón en la década de 1980, cuando las autoridades sanitarias se dieron cuenta de que ciertos alimentos mejoraban la salud y reducían el riesgo de contraer enfermedades.<sup>35, 36</sup> A pesar de que no hay una definición vigente aceptada por la Legislación Europea ni por las autoridades Sanitarias de los Estados Unidos, se entiende generalmente que un alimento funcional se consume dentro de la dieta regular y contiene componentes bioactivos.

Los arándanos (Figura 11), considerados como alimento funcional, son frutos pertenecientes al género *Vaccinium* de la familia de las ericáceas, con origen norteamericano. La planta es un arbusto de hojas perennes que alcanzan de 2 a 3 m de altura, mientras que el fruto posee forma de baya esférica, cubierta de pruina que se vuelve azulada, roja o negra con la madurez.<sup>37</sup>



Figura 11. Planta y bayas de arándano americano, *Vaccinium macrocarpon*

Dentro de sus usos en la industria alimenticia, los arándanos se emplean en fresco o en preparados como mermeladas, jugos y jaleas, teniendo mayor valor en crudo. Las bayas de arándanos son muy nutritivas, siendo ricas en fibra. Su aporte calórico es relativamente bajo, siendo de 30 calorías por cada 100 gramos. Además, cuentan

<sup>35</sup> Functional Foods Science in Europe. British Journal of Nutrition. **1998**, *80(1)*, S1-S193.

<sup>36</sup> Scientific Concepts of Functional Foods in Europe: Consensus Document. *Br J Nutr.* **1999**, *81(1)*, S1-S27.

<sup>37</sup> García-Rubio J.C.; Ciordia-Ara M.; García G. El cultivo del arándano. *Serida*, **2007**.



con un gran contenido de provitamina A, vitaminas C, E y magnesio. Junto a todo esto, se encuentran en las bayas diversas sustancias que les confieren propiedades antioxidantes, como son taninos catequicos y proantocianidinas oligoméricas, flavonoides, antocianos, ácidos fenólicos derivados del ácido cinámico y triterpenos, entre otros.<sup>8</sup>

Los arándanos han sido usados en la medicina tradicional para aliviar diversas afecciones, entre las que figuran trastornos del aparato digestivo, tratamiento de úlceras, problemas del sistema circulatorio, anemia, afecciones oculares e infecciones del tracto urinario. Sin embargo, es importante señalar que no hay pruebas contundentes acerca del beneficio del consumo de arándano en la prevención de las infecciones del tracto urinario, llegando a presentarse resultados y conclusiones contradictorias en diversos estudios realizados.<sup>38</sup>

### 3.1.1. COMPUESTOS ACTIVOS DE LOS ARÁNDANOS

Originalmente se había pensado que el posible efecto antibacteriano producido por el consumo de arándanos era provocado por el aumento de acidez en la orina, derivado del ácido quínico presente en el fruto.<sup>39</sup> Sin embargo, estudios posteriores demostraron que el nivel de acidez no era suficiente para ejercer un efecto bactericida por sí solo.<sup>40</sup>

Después, se encontró en la baya una gran cantidad de proantocianidinas, compuestos oligoméricos o poliméricos conformados por unidades de flavonoles, unidas por enlaces interflavánicos. Los dímeros tipo B tienen sus unidades flavan-3-ol unidas por un enlace interflavánico C4→C8 o C4→C6, mientras que los enlaces interflavánicos del dímero tipo A contienen un enlace tipo éter adicional entre C2 y

---

<sup>38</sup> Nicolle, L.E. Cranberry for Prevention of Urinary Tract Infection? *JAMA*. **2016**, 316(18), 1873-1874.

<sup>39</sup> Kinney, A.B.; Blount, M. Effect of cranberry juice on urinary pH. *Nurs. Res.* **1979**, 28, 287-290.

<sup>40</sup> Vasileiou, J. et. al. Current clinical status on the preventive effects of cranberry consumption against urinary tract infections. *Nutr. Res.* **2013**, 33, 595-607.

el hidroxilo de C7 (Figura 12).<sup>41</sup> Algunos estudios *in vitro* han demostrado que los dímeros y trímeros de las proantocianidinas tipo A presentan una actividad antiadherente que impide la unión de los microorganismos a las células endoteliales del tracto urinario, mientras que los dímeros de las proantocianidinas tipo B no presentan tal actividad.<sup>42</sup> Otros experimentos sugieren una reducción en la adherencia de la bacteria *E. coli* a las células endoteliales cuando las bacterias son pre incubadas con 5 a 75 mg de proantocianidinas tipo A.<sup>43</sup> Se ha postulado que tales proantocianidinas, *in vitro*, modifican los patrones de unión a las células endoteliales.<sup>44</sup>

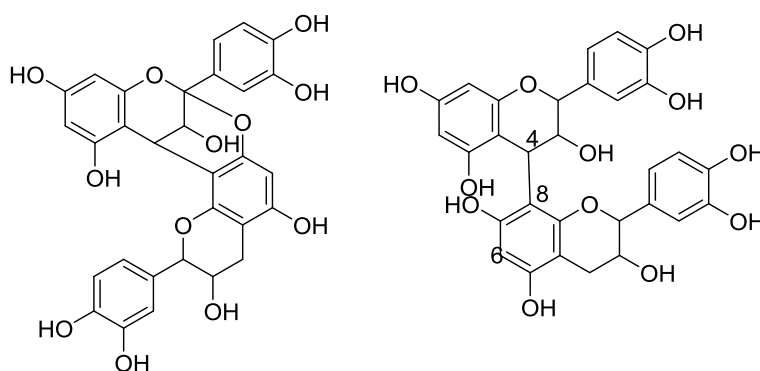


Figura 12. Proantocianidinas tipo A (izquierda) y B (derecha).

La sugerencia acerca de la dosis media recomendada a ingerir de 300 mL de jugo al día, con una concentración de 25 a 30%, es una cantidad con la que se logra reducir hasta en 50% la recurrencia de la bacteriuria.<sup>45</sup>

La tasa de absorción de las proantocianidinas es menor al 5%, por lo que estos compuestos llegan casi en su totalidad al colón, donde la microbiota intestinal los

<sup>41</sup> Dixon, R.A., Xie, D.Y., Sharma, S. B. Proanthocyanidins – a final frontier in flavonoid research? *New phytologist*. **2005**, 165(1), 9-28.

<sup>42</sup> Foo, L. Y.; Lu, Y.; Howell, A. B.; & Vorsa, N. A-type proanthocyanidin trimers from cranberry that inhibit adherence of uropathogenic P-Fimbriated Escherichia coli. *J. Nat. Prod*, **2000**, 63, 1225–1228.

<sup>43</sup> Gupta, K.; Chou, M.Y.; Howell, A.; et al. Cranberry products inhibit adherence of p-fimbriated Escherichia coli to primary cultured bladder and vaginal epithelial cells. *J. Urol*. **2007**, 177(6), p.p. 2357-60.

<sup>44</sup> Johnson, B.J.; Lin B.; Dinderman M.A. et al. Impact of cranberry on Escherichia coli cellular surface characteristics. *Biochem. Biophys. Res. Common*. **2008**, 19, 377(3), 992-4.

<sup>45</sup> Jepson, R.G.; Craig, J.C. Cranberries for preventing urinary tract infections. *Cochrane Database of Systematic Reviews*, **2008**.

degrada hasta obtener derivados de ácido benzoico. Una ruta sugerida por Stoupi et. al. es la mostrada en la Figura 13, en la que la proanticianidina tipo B (1) es degradada hacia su dimerio 2 y monomeros 3 y 4, que a su vez son degradados por la microbiota en 5-(3',4'-dihidroxifenil)- $\gamma$ -valerolactona (5), que entra en equilibrio con su forma abierta 5-(3'-metioxil-4'-dihidroxifenil)- $\gamma$ -valerolactona (6). Dicho compuesto sufre una dehidroxilación 7 por parte de la microbiota, seguida de descarboxilaciones tipo  $\beta$  8 y  $\alpha$  9, hasta terminar como derivados de ácido benzoico (10). El ácido benzoico se absorbe al torrente sanguíneo vía vena porta, y en hígado se conjuga con glicina, generando al ácido hipúrico.<sup>8</sup> Comunmente, el ácido benzoico es un producto del metabolismo microbiano de los aminoácidos fenilalanina y tirosina, aunque solo del 1 al 2% de estos aminoácidos se degradan de esta forma.<sup>46</sup>

---

<sup>46</sup> Bernard, K., Vulleumier, J.P., Burbacher, G. Zur Frage der Entstehung der Benzoesäure im Tierkörper. *Helv. Chim. Acta.* **1955**, 38, 1438.

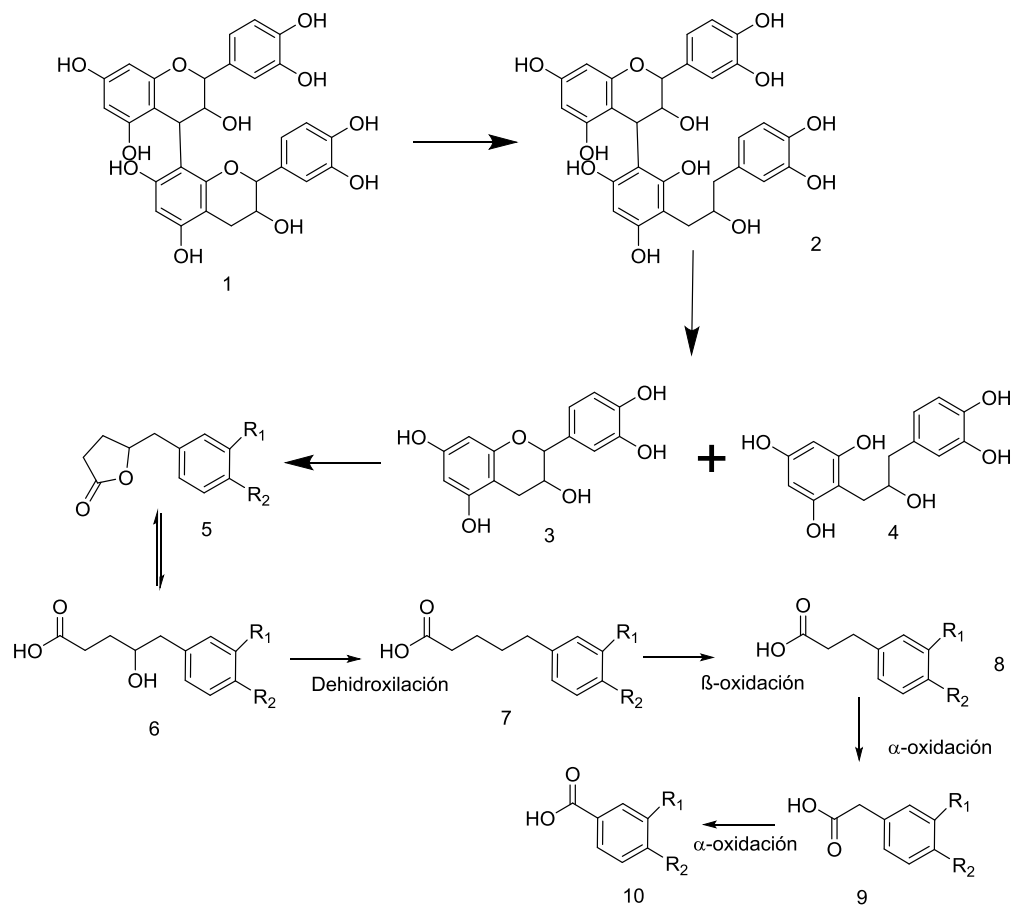
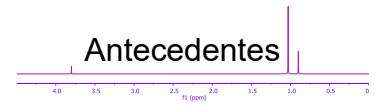


Figura 13. Vía metabólica para la degradación de proantocianidinas hacia derivados de ácido benzoico. Con  $R_1=H$  y  $R_2=H$ , el producto final de ácido benzoico.

### 3.1.2. ESTUDIO METABOLÓMICO PREVIO

En el año 2015 en la Universidad de Florida, se llevó a cabo una investigación acerca del cambio en el metaboloma de mujeres jóvenes americanas después de someterse a un régimen alimenticio con jugo de arándano o manzana. Se reclutaron a 18 jóvenes de la Universidad de Florida, con edades comprendidas entre los 21 y 29 años de edad, y con un Índice de Masa Corporal entre 18.5 y 21. A cada voluntaria se les dio instrucciones del régimen alimenticio, se les asignó de manera aleatorizada en un grupo de consumo, y se les pidió acudir a colecta de muestras de orina y sangre en los días indicados, para después repetir el ensayo de manera cruzada.



Las muestras, analizadas por RMN, revelaron que hay diferencia en el metaboloma del suero y de la orina después del consumo de los dos jugos. En suero se encontró que la concentración de lactato y D-glucosa disminuyeron tras el consumo de jugo de arándano, y se incrementó la concentración de citrato y tres compuestos desconocidos. Mientras que en la orina se encontró una elevación en la concentración de ácido hipúrico y de un compuesto desconocido.

## 4. METODOLOGÍA

El presente capítulo de metodología se encuentra dividido en 7 partes: las cohortes, el ensayo biológico, la adquisición y tratamiento de los espectros, el análisis multivariante, el análisis univariante y la identificación de los biomarcadores.

### 4.1. COHORTES

Se reclutaron 18 mujeres voluntarias estudiantes de la Facultad de Medicina de la UNAM, nivel licenciatura, con edades comprendidas en un rango de 18 a 22 años. Se les explicó el procedimiento para la ingesta de alimentos y la recolección de muestras (solo orina) durante los días del ensayo biológico, junto con un consentimiento informado de que no recibirían retribución económica por su participación. Previo al ensayo, se les realizaron mediciones de altura y peso para el cálculo de su Índice de Masa Corporal (IMC). En un principio se buscó la participación exclusivamente de mujeres sin sobrepeso, sin embargo, debido a la escasez de voluntarias, fue necesario incluir mujeres con ligero sobrepeso (hasta IMC=30).

El régimen alimentario y recolección de muestras se llevó a cabo de la siguiente manera:

1. Con el fin de *limpiar* el organismo, se le entregó a cada participante una lista de alimentos con niveles altos de proantocianidinas (fresas, uvas, manzanas, chocolates, bayas, etc.), los cuales no debían de ingerir durante el periodo denominado de depuración.
2. Terminando dicho periodo de 7 días, las voluntarias recolectaron la primera orina de la mañana, denominadas muestras blanco.
3. Se le entregó a cada voluntaria 6 botellas con 250 mL de jugo de manzana Sonrisa® o jugo de arándano Oceanspray®, con la finalidad de que se consumieran 250 mL en la mañana y 250 mL en la noche, en un periodo de 3 días. Las participantes fueron aleatorizadas balanceadamente a los grupos de consumo de jugo de arándano y manzana. Es importante mencionar que, en México, Oceanspray no comercializa jugo de manzana, y después de

comparar la composición de las marcas en el mercado, se encontró que el jugo de manzana de la marca Sonrisa® era la más parecida.

4. Al finalizar este periodo, las voluntarias volvieron a coleccionar y entregar la primera orina de la mañana.
5. Se les dio un periodo de reposo de 2 semanas para que volvieran a su dieta normal.
6. Finalizando este periodo, se realizó una repetición cruzada, con los mismos pasos desde el periodo de depuración por 7 días con entrega de la primera orina de la mañana como muestra blanco, seguidos del periodo de consumo por 3 días con entrega de la primera orina de la mañana, pero con el jugo contrario al que habían recibido en la primera ocasión.

Las muestras de orina fueron enfriadas a 4°C hasta el momento de su alicuotación. Las muestras se homogeneizaron en un vortex y se alicuotaron 600 µL en tubos Eppendorf de 1 mL. Dichas muestras se rotularon y congelaron a -80°C hasta el momento de su tratamiento.

La Tabla 1 muestra ejemplos de los códigos asignados a las muestras, en las que se tomó a la voluntaria 2 como referencia.

*Tabla 1. Significado de los códigos usados en la rotulación de las muestras*

<i>Clave</i>	<i>Significado</i>
<i>2B1</i>	<i>Voluntaria dos, Blanco, Primera colecta</i>
<i>2A2</i>	<i>Voluntaria dos, Arándano, Segunda colecta</i>
<i>2B3</i>	<i>Voluntaria dos, Blanco, Tercera colecta</i>
<i>2M4</i>	<i>Voluntaria dos, Manzana, Cuarta colecta</i>

## 4.2. PREPARACIÓN DE LA MUESTRA

Se prepararon 25 mL de una solución tampón de fosfatos 0.2 M en agua deuterada (pH=7). Las sales utilizadas fueron el fosfato monobásico de sodio (99.9%) y el fosfato dibásico de potasio (99.9%) de la marca comercial *Aldrich Chemical Company*, mientras que el agua deuterada (D<sub>2</sub>O 99.9%) fue obtenida con la casa

*Cambridge Isotope Laboratories*. Se adicionó 0.286 mg de sal TSP, como estándar interno, obteniendo una concentración de 67 $\mu$ M.

Las muestras almacenadas en congelación, fueron descongeladas a temperatura ambiente y homogeneizadas con vortex por 10 segundos.

De cada alícuota se tomaron 500  $\mu$ L y se colocaron en tubos Eppendorf nuevos. Se centrifugaron a 6900 RPM durante 10 minutos, para eliminar cuerpos formes de la orina, como cristales de sales o células epiteliales. Del sobrenadante se tomaron 400  $\mu$ L, se llevaron a otro tubo Eppendorf, y se adicionaron 200  $\mu$ L del buffer preparado. Las muestras se homogeneizaron en vortex por 5 segundos y se trasvasaron a tubos para RMN de 5 mm, para su adquisición en el espectrómetro de 700 MHz.

### 4.3. ADQUISICIÓN DE ESPECTROS DE RMN

Las muestras se adquirieron en un espectrómetro *Bruker Avance III HD* de 700 MHz (campo magnético de 16.44 T) con muestreador automático *Sample jet*. La secuencia de pulsos seleccionada fue el experimento 1D NOESY con presaturación de agua. La ganancia se calculó para cada muestra y se adquirieron un total de 128 repeticiones. La adquisición de los espectros se llevó a cabo de manera automática con el programa *IconNMR* de la compañía *Bruker*. El tiempo total de adquisición por espectro fueron 13 minutos aproximadamente.

Un total de 50 muestras, 44 muestras y 6 *pool* (mezcla de las muestras colectadas), fueron adquiridas el mismo día de la preparación.

### 4.4. PRETRATAMIENTO DE LOS ESPECTROS

El preprocesamiento de las muestras se realizó con el programa *Mestre Nova* (versión 8.1, *MestreLab*, España). Se cargaron los espectros en el programa, y se superpusieron. Posteriormente se llevó a cabo la corrección de fase diseñada para metabólica, que no toma en cuenta el pico del agua en 4.8 ppm. La corrección de línea base se realizó aplicando la alineación por polinomios de Bernstein.



Se realizaron 3 tratamientos a la serie de espectros: En uno se realizó alineación de señales (con intervalos seleccionados por el usuario, a la media de los espectros), en otro se le realizó reducción de dimensionalidad por generación de cajas homogéneas con un tamaño de 0.04 ppm (Sección 3.2.2), y en el último se realizó conjuntamente alineación y reducción de dimensionalidad. Para cada tratamiento se realizaron 2 tipos de normalizaciones: un normalizado a total del área y un normalizado con referencia al pico más intenso de la creatinina, ubicado en 3.06 ppm aproximadamente. A todos los tratamientos se les eliminó la región ubicada entre 4.70 y 6.00 ppm, zona donde se encuentran el agua y la urea.

Una vez realizado el pretratamiento, se extrajo la matriz y se guardaron los datos con el *script: NMR Matrix*. (.txt\*), que nos devuelve una matriz de KxN en un formato txt.

#### 4.5. ANÁLISIS QUIMIOMÉTRICO

Las matrices fueron exportadas al programa *SIMCA* (versión 14.1.0.247, MKS *Umetrics AB*, Umeå, Suiza). Se incluyó una columna X cualitativa que consta en la clasificación de las muestras dependiendo de su origen (blancos, provenientes del consumo de jugo de arándano y provenientes del consumo de jugo de manzana), y un identificador secundario numérico. Para todos los modelos realizados se utilizó un escalado tipo Pareto (par).

Como modelo inicial, se realizó el análisis PCA sobre el total de las muestras con el fin de observar tendencias internas en los grupos a partir del gráfico de puntuaciones. Posteriormente se realizaron los modelos PLS-DA y OPLS-DA, igualmente sobre el total de las muestras, esta vez para observar si había correlaciones con la matriz **Y** cualitativa, y si se hallaban presentes estructuras latentes conforme a esta. Además de analizar los gráficos de puntuaciones de estos modelos, se analizaron los gráficos de cargas y los *Hotelling's T2* con el fin de encontrar muestras anómalas.

También se construyeron modelos: PCA, PLS-DA y OPLS-DA, excluyendo los datos de las muestras los blancos para conocer sólo la variación entre las muestras

provenientes del consumo de jugo de arándano y de jugo de manzana, y al igual que con los primeros tres modelos se analizaron las puntuaciones, cargas y *Hotelling's T2*.

El análisis del gráfico *Hotelling's T2* dio a conocer muestras anómalas en los modelos. Se investigó la causa de la anomalía, y considerándose pertinente, las muestras anómalas se eliminaron del modelo original, hasta tener un *Hotelling's T2* sin anómalos.

El preprocesamiento seleccionado para el análisis definitivo fue aquel que presentó en el modelo PLS-DA el mayor valor de predicción  $Q^2$ .

Finalmente, para conocer las señales que causan la diferencia entre los grupos, se generó la Línea-S del modelo OPLS-DA de resultados provenientes del consumo de arándano y consumo de manzana. En vista de los altos valores de predicción alcanzados, se optó por tomar como límite un valor de correlación mayor a 0.7 al seleccionar las señales como posibles biomarcadores.

#### 4.6. IDENTIFICACIÓN DE BIOMARCADORES

La información proporcionada por el gráfico de Línea-S se comparó con lo obtenido por el grupo de Gu y colaboradores. Se encontraron otras señales no reportadas en el estudio y se buscaron en las bases de datos *Human Metabolome Data Base* (HMDB) y la *Biological Magnetic Data Bank* (BMDB), para señalar a los metabolitos responsables.

Se analizó la correspondencia para el ácido hipúrico entre las señales obtenidas y las señales en la literatura.

También se realizaron espectros bidimensionales  $^1\text{H}$  TOCSY y HSQC en una muestra donde las señales de los compuestos desconocidos se encontraban en mayor cantidad.

#### 4.7. ANÁLISIS UNIVARIANTE

De las señales que originaba la variación en los grupos, se seleccionaron aquellas que fueron las más intensas dentro del compuesto, y se sometieron a un test de Shapiro-Wilks para normalidad. Comprobada la normalidad de los datos, se realizaron pruebas de t-Student para muestras dependientes con un 95% de confianza usando el programa *Statistica* (versión 8.0, *Stat Soft Inc.*, EE.UU.AA.).

## 5. RESULTADOS Y DISCUSIÓN

En este capítulo se muestran los resultados y el análisis de los mismos. Se muestran las características de las participantes en las cohortes, los espectros de  $^1\text{H}$  RMN, los modelos quimiométricos y la identificación de los metabolitos a través de los espectros monodimensionales y bidimensionales.

### 5.1. CARACTERÍSTICAS DE LAS PARTICIPANTES

De las 18 voluntarias que se presentaron a la reunión informativa, una participante fue excluida por presentar obesidad. Ese día 17 voluntarias firmaron el consentimiento informado, pero solo 15 se presentaron para el reconocimiento físico, y 11 completaron todo el estudio.

El reconocimiento físico se llevó a cabo por el doctor Joel Nava Rangel, Médico del Deporte y Actividad Física, de la Facultad de Medicina de la UNAM. Los resultados de IMC de las que completaron todo el estudio se muestran en la Tabla 2.

Tabla 2. Índice de masa corporal de las participantes del estudio

Voluntaria	IMC (kg/m <sup>2</sup> )	Observaciones
V2	20.75	Sana
V3	24.62	Sana
V4	29.46	Sobrepeso
V5	22.85	Sana
V6	25.50	Sobrepeso
V8	26.91	Sobrepeso
V9	20.00	Sana
V10	27.34	Sobrepeso
V11	24.46	Sana
V12	24.47	Sana
V15	21.16	Sana

### 5.2. ESPECTROS DE PROTÓN

En la Figura 14 se muestra el espectro  $^1\text{H}$  RMN obtenido de una de las muestras *pool* de orina, donde se pueden apreciar una gran cantidad de señales bien resueltas que permiten la identificación de algunos metabolitos urinarios. Cada

señal o grupo de señales que se obtiene en el espectro es atribuible a un metabolito, y el área contenida bajo la señal es directamente proporcional a la concentración de dicho metabolito.

Un metabolito puede presentar varias señales en diferentes partes del espectro, por ejemplo, la creatinina (4.06 y 3.05 ppm) o una sola señal como la urea (5.80 ppm) o el agua (4.82 ppm). En la Figura 14 se identificaron los picos característicos de algunos de los metabolitos.

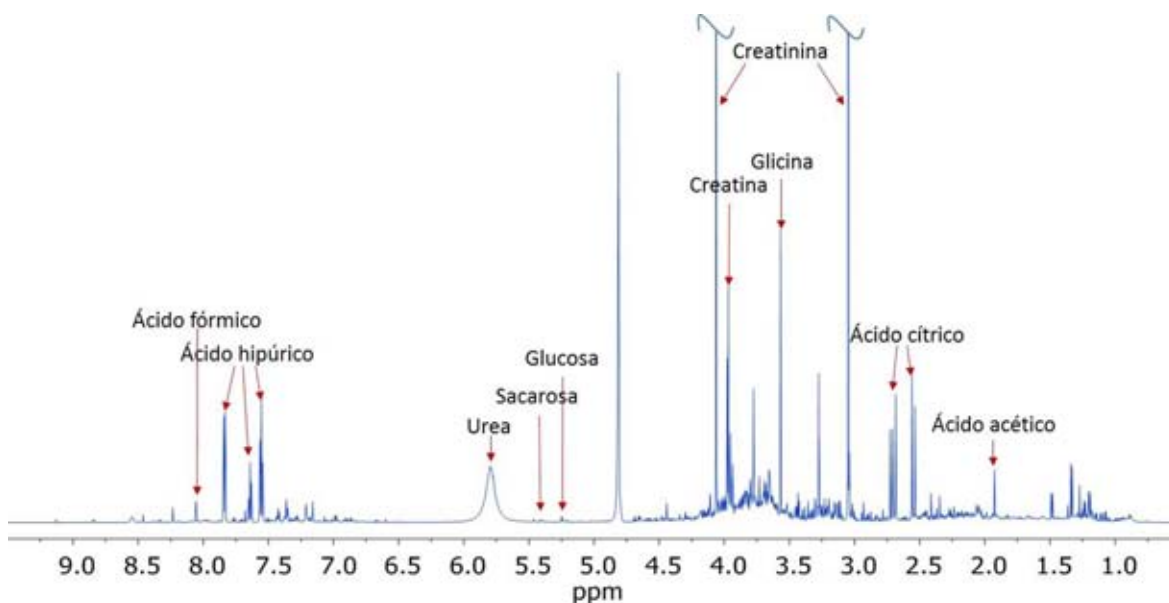


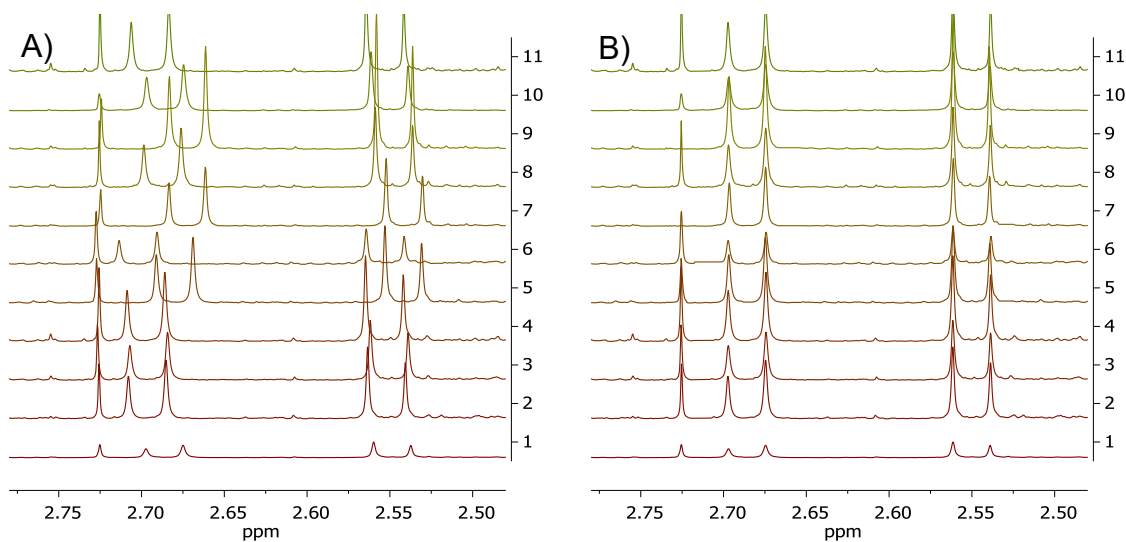
Figura 14. Espectro  $^1\text{H}$  RMN (700 MHz,  $\text{D}_2\text{O}/\text{H}_2\text{O}$  10/90, 25°C, 1D-NOESY-Prsat) de orina. Algunos de los metabolitos más abundantes son identificados en el espectro.

### 5.2.1. ALINEACIÓN Y REDUCCIÓN DE LA DIMENSIONALIDAD

Como se mencionó anteriormente, una característica común en los espectros de  $^1\text{H}$  RMN es que un cambio de pH resulta en ligeros cambios en los desplazamientos químicos de los protones con comportamiento ácido-base. Para hacer frente a este problema, se recurrió a dos estrategias igualmente válidas en el procesamiento de los espectros.

La primera opción fue el uso de algoritmos de alineación. En la Figura 15 se muestra la región del espectro donde aparecen las señales del ácido cítrico (2.53 y 2.65 ppm)

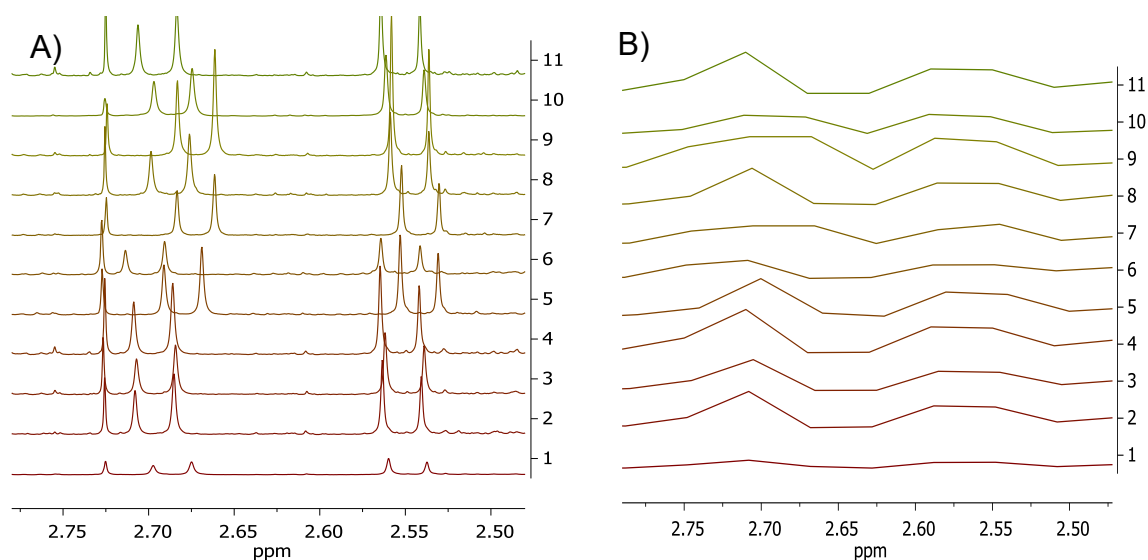
y de la dimetilamina (2.73 ppm). En la Figura 15A se puede observar que las señales presentan ligeras variaciones en su desplazamiento químico. Dichas diferencias se corrigieron aplicando un algoritmo de alineación, contenido en el programa Mestre Nova, dando como resultado los espectros mostrados en la Figura 15B, donde las señales se encuentran en el mismo desplazamiento químico, eliminando el error biológico aportado por la diferencia en el pH. Es muy importante hacer notar que en los espectros 7 y 9 de la Figura 15B la señal correspondiente a la dimetilamina se atenuó después de la alineación. Esto se debió a que el límite del rango de alineación está justo sobre el pico, y atenuó su señal en el proceso. Por esta razón, la selección de los intervalos de alineación debe realizarse con cuidado, a fin de no eliminar información relevante.



*Figura 15. Ampliación de la región de los espectros  $^1\text{H}$  RMN donde aparecen las señales de ácido cítrico y dimetilamina (11 espectros de 3 voluntarias). A) Antes de alinear y B) después de alinear.*

La otra herramienta utilizada fue la reducción dimensional. Normalmente, los tamaños más aceptados de cajas suelen ser desde 0.01 hasta 0.04 ppm, aunque

hay artículos que describen el uso de cajas de 0.1 ppm.<sup>47</sup> En la Figura 16A nuevamente se muestra la región del espectro entre 2.45 ppm y 2.80 ppm y en la Figura 16B se muestra la misma la región del espectro después de reducir la dimensión a cajas de tamaño de 0.04 ppm. En este espectro se puede observar como las señales de ácido cítrico y de la dimetilamina se encuentran agrupadas en estas cajas. Si las señales son lo suficientemente reproducibles y el tamaño de caja lo suficientemente grande, se contendrán las pequeñas variaciones provenientes de la variación de las muestras y ayudará a eliminar la contribución de los artefactos, sin generar problemas en el aporte al modelo estadístico.



*Figura 16. Ampliación de la región de los espectros  $^1\text{H}$  RMN donde aparecen las señales de ácido cítrico y dimetilamina (11 espectros de 3 voluntarias). A) Antes de la reducción dimensional y B) después de la reducción dimensional.*

En las investigaciones de metabolómica no hay un procedimiento comúnmente aceptado, por lo que muchos autores prueban diferentes métodos hasta encontrar el que mejores resultados arroje. En nuestro estudio se decidió probar alineación, reducción dimensional y el conjunto de ambos. En la sección 5.3.2 se muestran los resultados de los preprocesamientos.

<sup>47</sup> Watson, D.G. Application of Principal Component Analysis to 1H-NMR Data Obtained from Propolis Samples of Different Geographical Origins. *Phytochem. Anal.* 17. **2006**. 323-331

## 5.2.2. NORMALIZACIÓN

En las muestras biológicas hay dos tipos de variación: la variación biológica relacionada con el objetivo del experimento, como estados patológicos y fisiológicos determinados, y la variación no controlada, como ruido experimental o estados no controlados dentro del objetivo del experimento (estrés, medicaciones o alimentación). En el caso concreto de las muestras de orina, la cantidad de agua ingerida por el individuo es crítica en la concentración de los metabolitos excretados.

La normalización de los espectros es un paso necesario para controlar las posibles variaciones de la concentración de las muestras analizadas. Dicho procedimiento asegura que el resultado hallado en el modelado de los datos corresponda realmente a una diferencia en la concentración de los metabolitos, y no a errores no controlados. Los métodos más comunes de normalización para espectros de RMN son la normalización por el área total espectral, especialmente cuando las diferentes muestras contienen aproximadamente el mismo número de metabolitos, y la normalización al área de un pico de espectro. En este trabajo se usó la normalización al total de área, en la que se confía en que los espectros son lo suficiente similares entre sí como para controlar el error. El otro tipo de normalización usado en este trabajo fue la normalización a la intensidad relativa de la creatinina, que se seleccionó porque se considera que su concentración en orina es constante entre las muestras de personas que no tengan daño renal. Los resultados se muestran en la sección 5.3.2.

## 5.3. MODELOS QUIMIOMÉTRICOS

### 5.3.1. ANÁLISIS Y EXCLUSIÓN DE ANÓMALOS

Al analizar los gráficos *Hotelling's T2*, se notó que para todos los modelos quimiométricos generados, las muestras provenientes de la voluntaria cinco salían del intervalo de 95, e incluso, del 99% de confianza. En la figura 17 se muestra el gráfico *Hotelling's T2* correspondiente a un PCA con el total de las muestras, en el que la muestra 5M4 sale del 95% de confianza, mientras que en la figura 18 se



muestra el gráfico *Hotelling's T2* de un modelo PCA con las muestras de tratamiento con jugos, donde la muestra 5M4 sale del 95% de confianza.

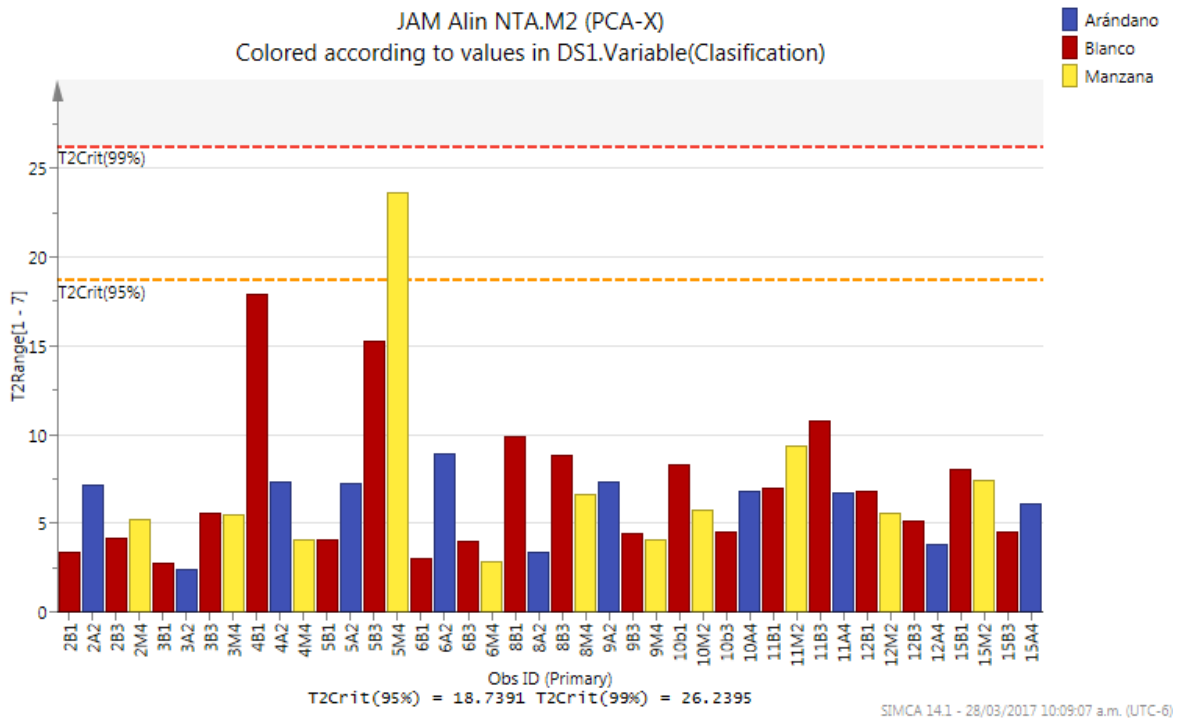


Figura 17. Gráfico *Hotelling's T2* del modelo PCA del total de las muestras con el preprocesamiento: Alineación y Normalización al Total de Área.

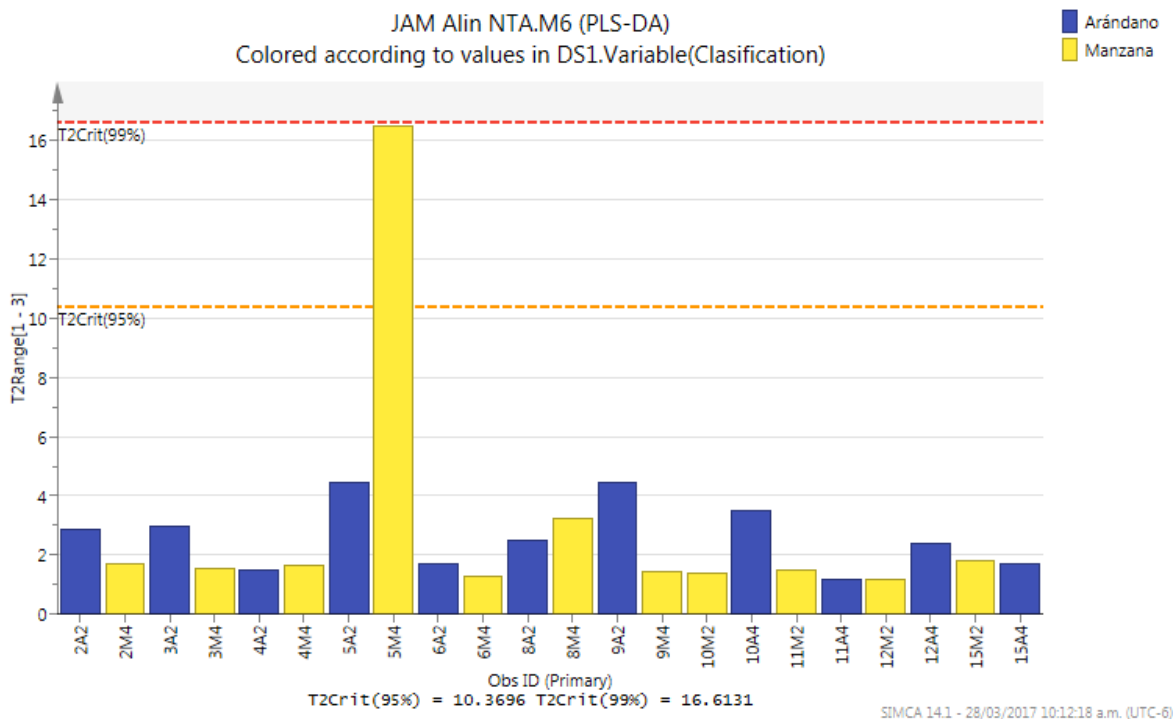


Figura 18. Gráfico Hotelling's  $T^2$  del modelo PCA de las muestras de consumo de jugos con el preprocesamiento Alineación con Normalización al Total de Área.

Al revisar los datos de intensidad relativa de los espectros  $^1\text{H}$  RMN de dicha participante, se observó que dichos datos correspondientes a las señales de ácido hipúrico en la zona de los aromáticos (8-7 ppm) se encontraban elevados. Como ejemplo, se muestra el comparativo de los espectros  $^1\text{H}$  RMN en la región de los aromáticos, de las muestras blanco de la primera colecta (Figura 19), en donde se observa que la señal de ácidos hipúrico de la voluntaria 5 estaba muy aumentado. A partir de esto se consideró el hecho de que esta participante pudo no haber seguido las recomendaciones del régimen alimenticio señalado, y se decidió eliminar sus muestras de los modelos quimiométricos.

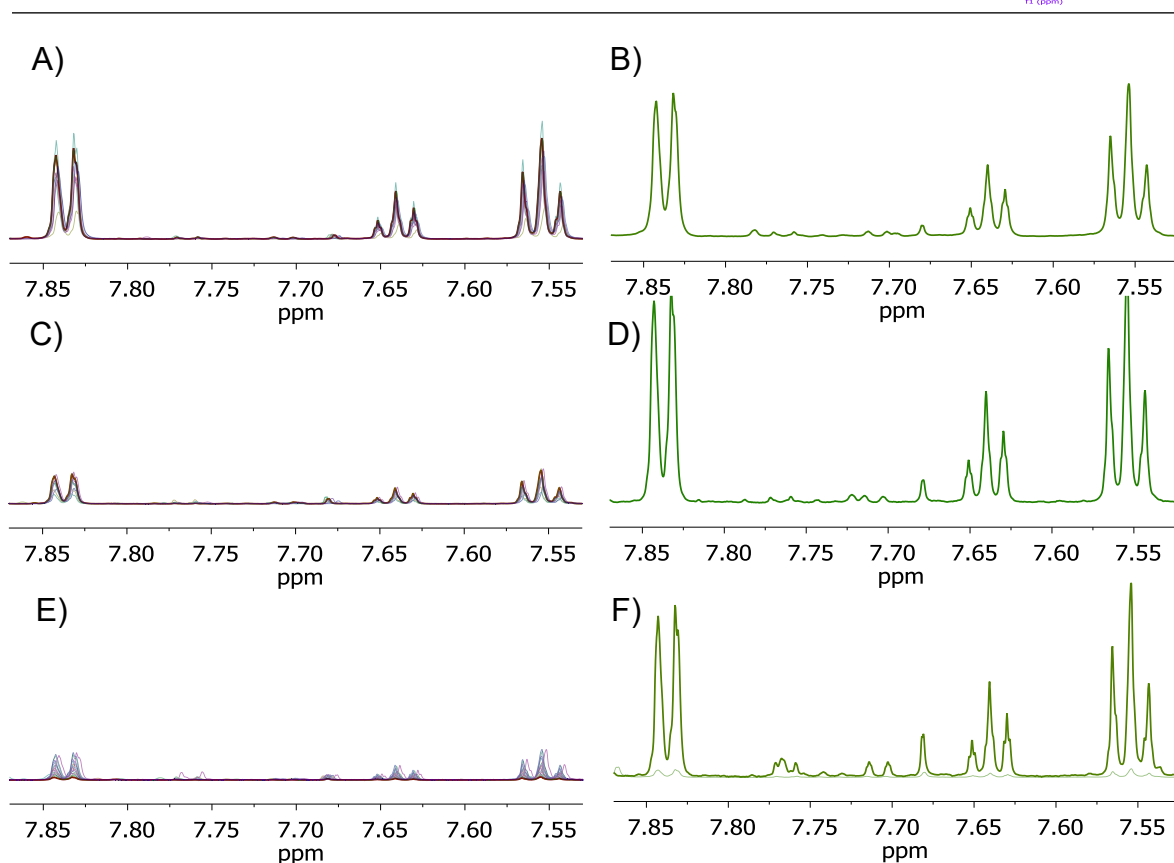


Figura 19. Ampliación de los espectros  $^1\text{H}$  RMN (700 MHz,  $\text{D}_2\text{O}/\text{H}_2\text{O}$  10/90, 25 °C, 1D-NOESY-Presat) en la zona de aromáticos. Tratamiento con: A) y B) jugo de arándano, C) y D) jugo de manzana, E) y F) blancos. B), D) y F) Voluntaria 5, A), C) y E) Resto de las voluntarias.

Los parámetros obtenidos de los modelos quimiométricos antes de la eliminación de los espectros de la voluntaria 5 se muestran en la tabla 3. A pesar de que se pueden encontrar dos modelos predictivos con un valor muy alto de  $Q^2$  (mayores a 0.90 para modelos de correlación con pre-procesamiento con *binning* y normalización al total de área), la mayor parte de los valores de predicción estimados para los modelos, independientemente del tipo de preprocesamiento rondan entre el 0.70 y 0.89.

Tabla 3. Parámetros de los modelos quimiométricos con todas las muestras.

Modelo/Parámetro	N	A	R <sup>2</sup> X	R <sup>2</sup> Y	Q <sup>2</sup>
<i>Preprocesamiento: alineación y normalización a creatinina</i>					
PCA	48	8	0.635	-	0.291
PCA	42	7	0.609	-	0.273
PLS-DA	42	3	0.405	0.812	0.601
OPLS-DA	42	2+1+0	0.405	0.804	0.549
PCA	22	3	0.579	-	0.266
PLS-DA	22	4	0.587	0.991	0.823
OPLS-DA	22	1+1+0	0.485	0.868	0.723
<i>Procesamiento: alineación y normalización al total de área</i>					
PCA	48	7	0.526	-	0.210
PCA	42	7	0.541	-	0.207
PLS-DA	42	2	0.260	0.886	0.762
PCA	22	4	0.553	-	0.154
PLS-DA	22	3	0.418	0.990	0.894
OPLS-DA	22	1+3+0	0.475	0.999	0.858
<i>Procesamiento: binning 0.04 ppm y normalización a creatinina</i>					
PCA	48	6	0.696	-	0.300
PCA	42	5	0.658	-	0.303
PLS-DA	42	3	0.491	0.770	0.530
OPLS-DA	42	2+2+0	0.559	0.805	0.579
PCA	22	4	0.742	-	0.353
PLS-DA	22	1	0.333	0.782	0.730
OPLS-DA	22	1+1+0	0.543	0.841	0.744
<i>Procesamiento: binning 0.04 ppm y normalización al total de área</i>					
PCA	48	9	0.751	-	0.208
PCA	42	7	0.700	-	0.257
PLS-DA	42	3	0.438	0.898	0.787
OPLS-DA	42	2+3+0	0.530	0.942	0.816
PCA	22	4	0.682	-	0.268
PLS-DA	22	4	0.600	0.989	0.915

<i>OPLS-DA</i>	22	1+4+0	0.669	0.996	0.919
<i>Procesamiento: alineación con binning 0.04 ppm y normalización a creatinina</i>					
<i>PCA</i>	48	8	0.763	-	0.259
<i>PCA</i>	42	7	0.737	-	0.271
<i>PLS-DA</i>	42	3	0.490	0.772	0.533
<i>OPLS-DA</i>	42	2+2+0	0.553	0.810	0.570
<i>PCA</i>	22	4	0.741	-	0.377
<i>PLS-DA</i>	22	1	0.333	0.783	0.730
<i>OPLS-DA</i>	22	1+1+0	0.546	0.841	0.743
<i>Procesamiento: alineación con binning 0.04 ppm y normalización a total de área</i>					
<i>PCA</i>	48	7	0.742	-	0.340
<i>PCA</i>	42	7	0.691	-	0.206
<i>PLS-DA</i>	42	3	0.434	0.901	0.788
<i>OPLS-DA</i>	42	2+3+0	0.531	0.945	0.823
<i>PCA</i>	22	4	0.678	-	0.275
<i>PLS-DA</i>	22	4	0.598	0.992	0.913
<i>OPLS-DA</i>	22	1+4+0	0.673	0.994	0.915

N: número de muestras usadas para generar el modelo; A: número de componentes principales generados;  $R^2X$ : varianza explicada por la matriz **X**;  $R^2Y$ : varianza explicada por la matriz **Y**;  $Q^2$ : predictibilidad.

Una vez eliminados los datos de la voluntaria 5 (tabla 4), el valor del parámetro  $Q^2$  de los modelos supervisados aumentó comparado con los modelos anteriores, exceptuando en los modelos de PLS realizados con el preprocesamiento de *binning* 0.04 ppm y normalización al total de área con alineación (0.913 a 0.91) y sin alineación (de 0.91 a 0.89). Ambos cambios pequeños y poco representativos.

Tabla 4. Parámetros de los modelos quimiométricos después de la eliminación de los datos de la voluntaria 5.

Modelo/Parámetro	N	A	R <sup>2</sup> X	R <sup>2</sup> Y	Q <sup>2</sup>
<i>Procesamiento: alineación y normalización a creatinina</i>					
PCA	38	6	0.558	--	0.275
PLS-DA	38	3	0.393	0.847	0.672
OPLS-DA	38	2+2+0	0.423	0.894	0.620
PCA	20	3	0.566	--	--
PLS-DA	20	2	0.469	0.916	0.829
OPLS-DA	20	1+1+0	0.469	0.916	0.822
<i>Procesamiento: alineación y normalización al total de área</i>					
PCA	38	7	0.555	--	0.191
PLS-DA	38	3	0.325	0.949	0.796
OPLS-DA	38	2+3+0	0.392	0.980	0.840
PCA	20	3	0.498	--	0.237
PLS-DA	20	2	0.379	0.983	0.900
OPLS-DA	20	1+3+0	0.379	0.983	0.897
<i>Procesamiento: binning 0.04 ppm y normalización a creatinina</i>					
PCA	38	6	0.706	--	0.355
PLS-DA	38	4	0.583	0.872	0.604
OPLS-DA	38	2+4+0	0.663	0.910	0.707
PCA	20	3	0.687	--	0.458
PLS-DA	20	2	0.583	0.782	0.859
OPLS-DA	20	1+1+0	0.583	0.841	0.857
<i>Procesamiento: binning 0.04 y normalización al total de área</i>					
PCA	38	3	0.493	--	0.269
PLS-DA	38	5	0.565	0.968	0.836
OPLS-DA	38	2+4+0	0.595	0.977	0.890
PCA	20	3	0.627	--	0.324
PLS-DA	20	2	0.492	0.973	0.895
<i>Procesamiento: alineación con binning 0.04 y normalización a creatinina</i>					
PCA	38	6	0.704	--	0.349
PLS-DA	38	3	0.501	0.821	0.585
OPLS-DA	38	2+4+1	0.660	0.551	0.532

PCA	20	3	0.683	--	0.450
PLS-DA	20	2	0.585	0.906	0.858
OPLS-DA	20	1+1+0	0.585	0.906	0.856
<i>Procesamiento: alineación con binning 0.04 y normalización a total de área</i>					
PCA	38	6	0.666	--	0.278
PLS-DA	38	5	0.526	0.970	0.846
OPLS-DA	38	2+6+0	0.667	0.987	0.908
PCA	20	1	0.437	--	0.351
PLS-DA	20	2	0.496	0.975	0.910
OPLS-DA	20	1+2+0	0.549	0.989	0.915

N: número de muestras usadas para generar el modelo; A: número de componentes principales generados;  $R^2X$ : varianza explicada por la matriz  $X$ ;  $R^2Y$ : varianza explicada por la matriz  $Y$ ;  $Q^2$ : predictibilidad.

### 5.3.2. SELECCIÓN DE PREPROCESAMIENTO

Los preprocesamientos probados para la corrección del error biológico fueron: i) alineación de señales, ii) reducción de la dimensionalidad con un tamaño de caja de 0.04 ppm y iii) una combinación de ambos métodos. Por otro lado, los preprocesamientos probados para la normalización fueron la normalización al total de área (considerando que las muestras tienen una gran similitud), y normalización al pico de la creatinina (pico más intenso en 3.06 ppm) cuya concentración urinaria en pacientes sanas es muy poco variable.

Para seleccionar el mejor procesamiento realizado sobre los datos, se decidió sólo tomar en cuenta la capacidad predictiva del modelo PLS-DA, debido a que: i) el modelo PCA solo modela la dirección de variación máxima, aunque está no concuerde con la variación que se busca en el estudio. ii) El modelo OPLS-DA, a pesar de ser una extensión del modelo PLS-DA, no se tomó en cuenta porque en un modelo no se obtuvo la generación de componentes ortogonales.

El tipo de preprocesamiento que se seleccionó fue el que obtuvo el mayor valor de predicción  $Q^2$ , que como se puede observar en la tabla 4 fue el que se realizó con alineación con *binning* y normalización al total de área, dando un valor de  $Q^2 = 0.846$

para el modelo con los tres tipos de muestras y  $Q^2 = 0.910$  para el modelo con las muestras de solo los regímenes alimentarios.

#### 5.4. ANÁLISIS DE LAS MUESTRAS

El grupo de investigación en la Universidad de Florida reportó que, al realizar un modelo considerando los tres tipos de muestras (las muestras blanco, y tras el tratamiento con jugo de arándano y de manzana), no se encontró agrupación alguna conforme al tipo de muestra en el PCA ni en el OPLS-DA. Sin embargo, en el presente trabajo, si se encontraron tendencias y agrupaciones en ambos modelos. Se puede notar en el gráfico de PCA (Figura 20) que las muestras del tratamiento con arándano se diferencian por completo de las muestras blanco y de las de tratamiento con manzana, aunque estas últimas se confunden entre sí en su agrupación. Esto refleja que la variación de las muestras por sí solas no es capaz de explicar una diferencia real entre las muestras de tratamiento con manzana y los blancos.

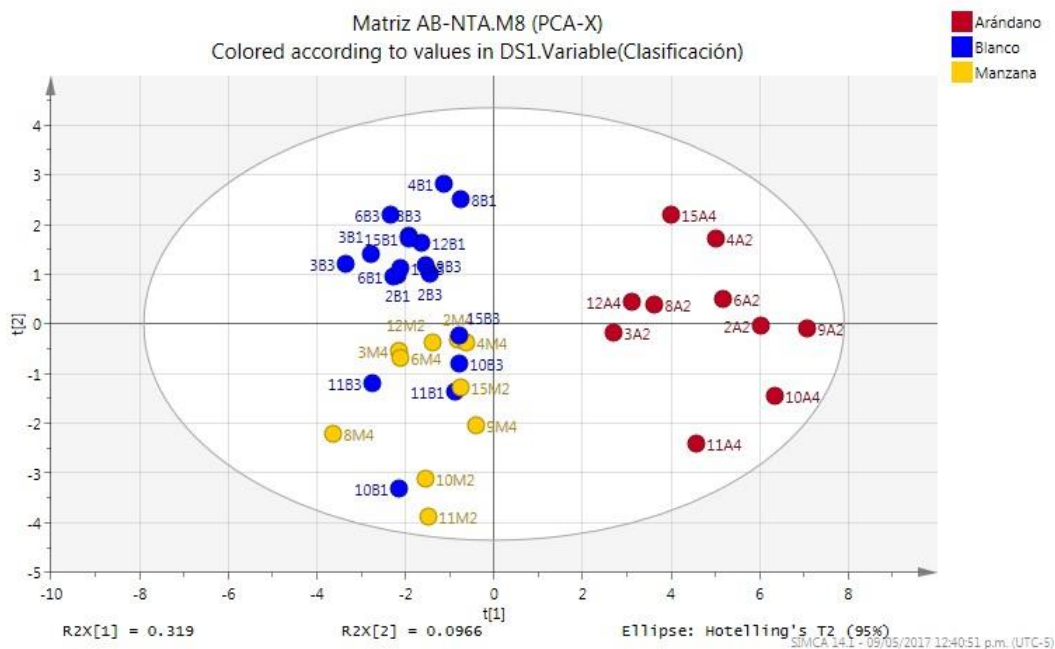


Figura 20. Gráfico de puntuaciones del modelo PCA con los tres tipos de muestras.



En el gráfico de PLS-DA (Figura 21) se encuentra una agrupación y diferenciación de las muestras, que se acentúa en el gráfico de OPLS-DA (Figura 22). Al realizar el modelo de regresión discriminativa, la variación sí es capaz de modelarse conforme al tipo de muestra, obteniendo un nivel de predicción bastante alto. Como el modelo OPLS-DA toma toda la variación no explicada y la agrupa en un vector ortogonal, extrayéndola de la proyección, explica el aumento en la aproximación de los puntos en el gráfico.

Tanto en el modelo PLS-DA como en el modelo OPLS-DA, el PC1 describe la variación entre las muestras de tratamiento con arándano, ubicadas a la izquierda del gráfico, y las muestras de tratamiento con manzana y blancos, ubicados a la derecha del gráfico. El PC2 por su parte, describe la variación entre las muestras de tratamiento con manzana, ubicadas en la parte superior, y las muestras blanco, ubicadas en la parte inferior.

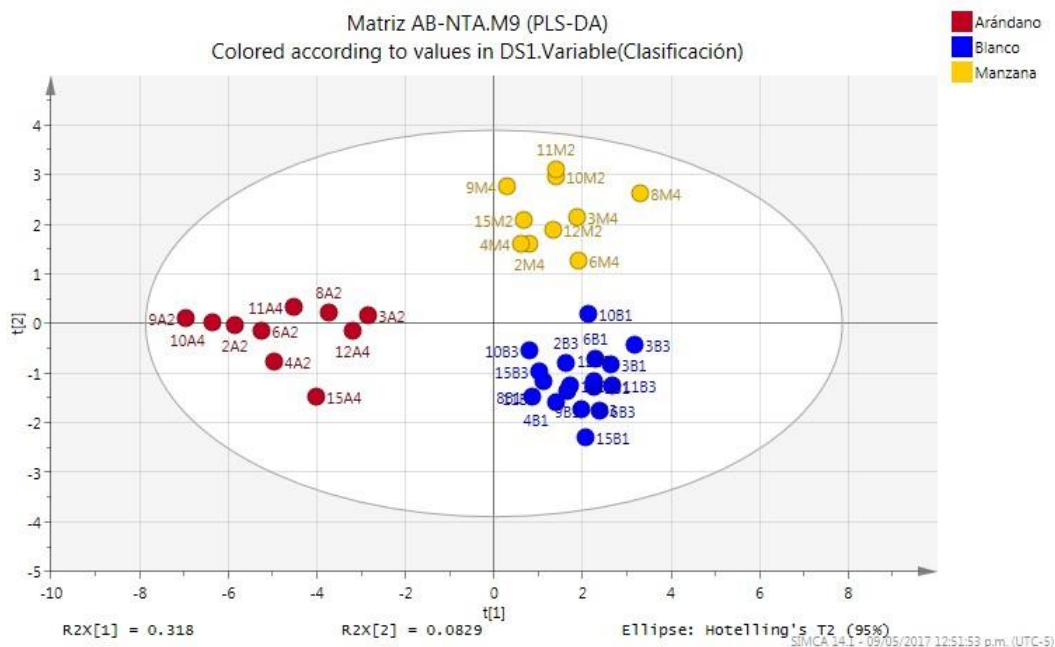


Figura 21. Gráfico de puntuaciones del modelo PLS-DA con los tres tipos de muestras.

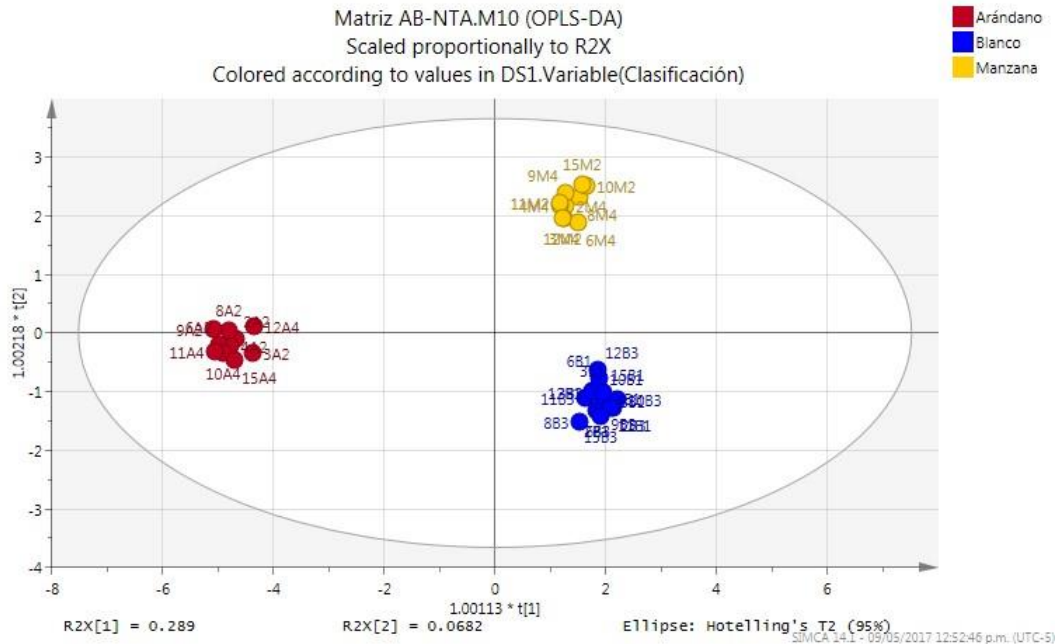


Figura 22. Gráfico de puntuaciones del modelo OPLS-DA con los tres tipos de muestras.

Al eliminar las muestras blanco como en el estudio de la Universidad de Florida y realizar nuevamente el PCA, se obtuvo el gráfico de puntuaciones mostrado en la Figura 23, cuyo modelo describe la varianza de las muestras con un solo PC, donde se separan en muestras de tratamiento con manzana (hacia abajo) y muestras de tratamiento con arándano (hacia arriba).

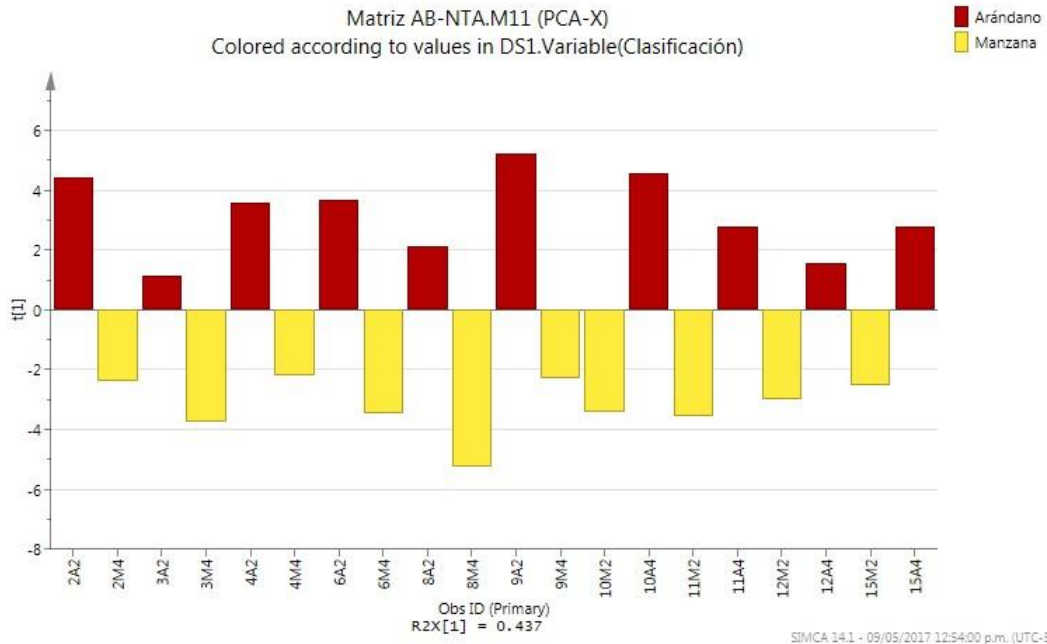


Figura 23. Gráfico de puntuaciones del modelo PCA con muestras de tratamiento con arándano y manzana.

Al indicar el grupo al que pertenecen las muestra en un modelo de discriminación, y generar el modelo PLS-DA (Figura 24), se observa que la proyección de las muestras ha dado un giro de 180° sobre el PC1, ubicando a las muestras de tratamiento con jugo manzana en el lado derecho y las muestras de tratamiento con jugo de arándano en el lado izquierdo, ambos en una agrupación más cercana.

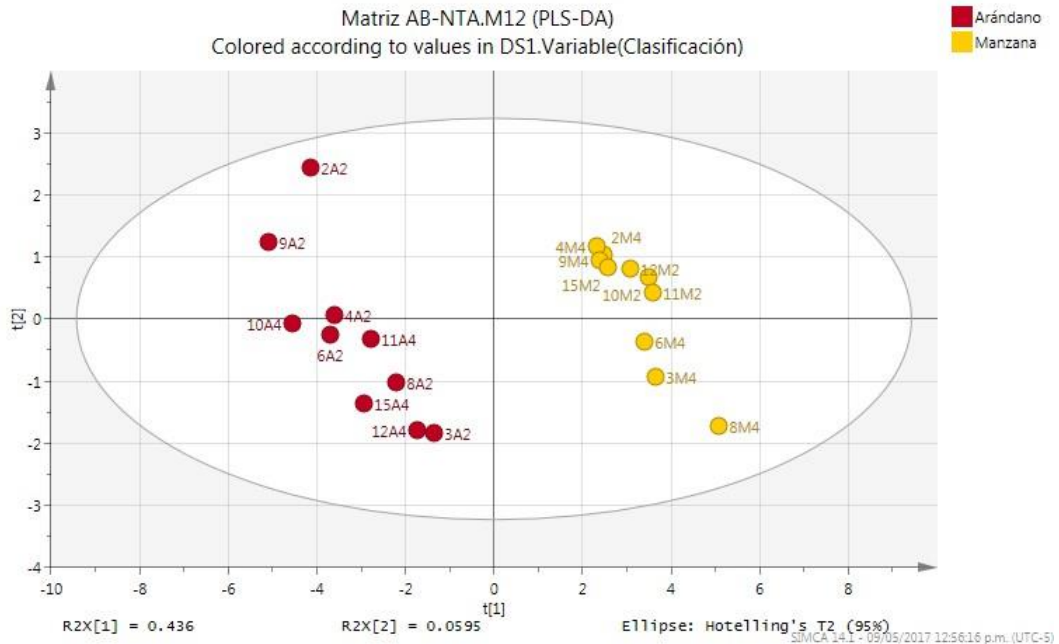


Figura 24. Gráfico de puntuaciones del modelo PLS-DA con las muestras del tratamiento con arándano y manzana.

Conforme al modelo OPLS-DA (Figura 25) se puede observar que la diferenciación se realiza por completo sobre el PC1, siendo la dispersión mínima. Es de suponer que en este modelo se encierra la variación biológica no controlada sobre el vector ortogonal, mejorando la cercanía entre las muestras descritas por su variación en el PC1.

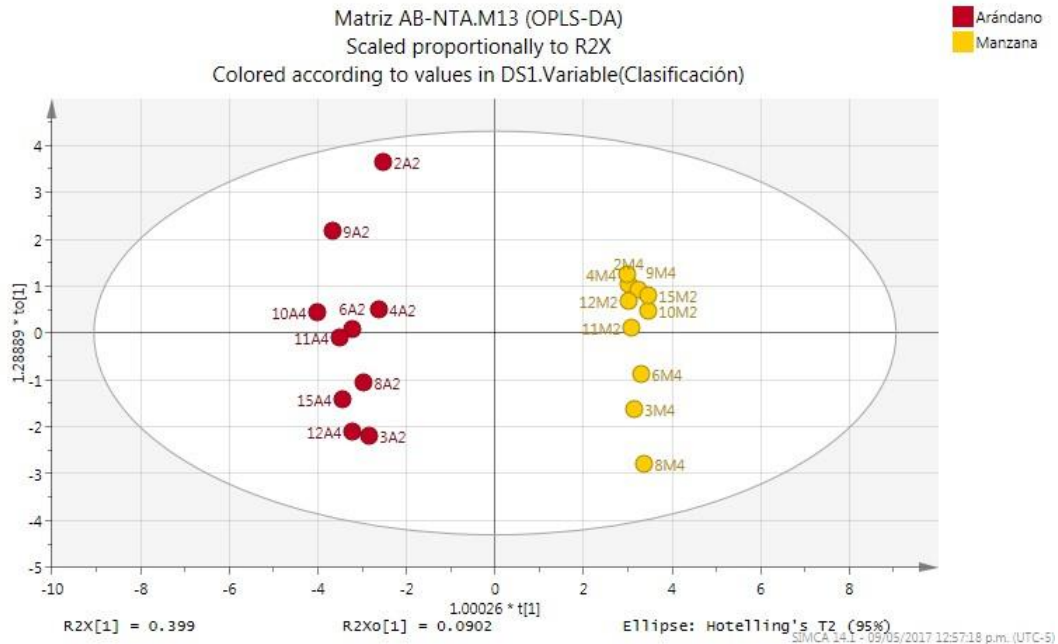


Figura 25. Gráfico de puntuaciones del modelo OPLS-DA con las muestras del tratamiento con arándano y manzana.

Para conocer la validez de los modelos PLS-DA y OPLS-DA, al igual que en el estudio ejecutado en la Universidad de Florida, se realizaron ensayos de permutaciones con  $n=100$  y 7 repeticiones (Figura 26), que en lo reportado por el grupo se seleccionó como límite de validez un valor menor a 0.05. En este trabajo, los valores de intersección para  $Q^2$  obtenidos para los modelos con el total de muestras fueron respectivamente -0.590 y -0.857 para PLS-DA y OPLS-DA, y para los valores obtenidos en los modelos con las muestras de los dos regímenes alimentarios fueron -0.083 y -0.679 para PLS-DA y OPLS-DA, respectivamente, por lo que todos los modelos son válidos.

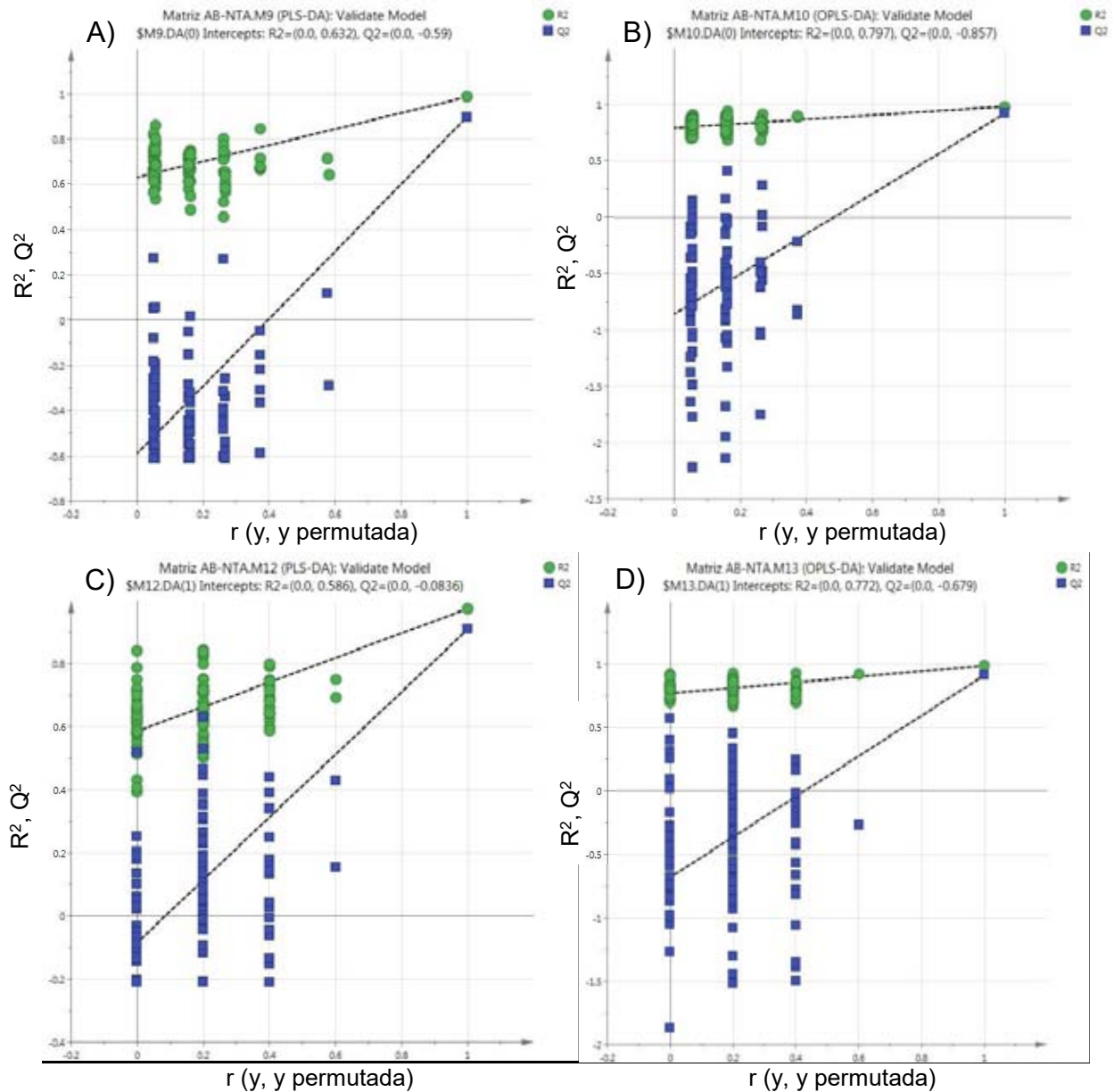


Figura 26. Gráficos del ensayo de permutaciones ( $n=100$ ) con 7 repeticiones. A) PLS-DA y B) OPLS-DA con los tres tipos de muestras. C) PLS-DA y D) OPLS-DA con los dos tratamientos experimentales

Analizando los gráficos de cargas se pueden conocer cuáles son las variables (ppm) que más peso tienen en el modelo. El gráfico de cargas del modelo PLS-DA (Figura 27), muestra que las variables que tienen un mayor impacto son las señales en 4.01, 7.56 y 7.69 ppm, y con un impacto menor la zona comprendida entre 3 y 4 ppm.

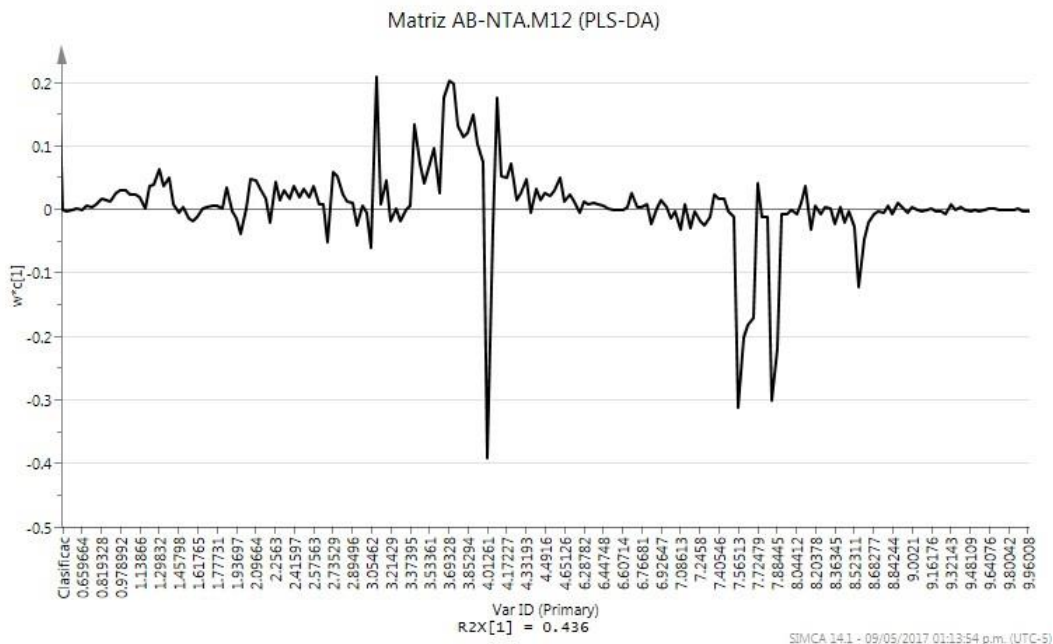


Figura 27. Gráfico de cargas del modelo PLS-DA.

También se utilizó la herramienta del modelo OPLS llamada Línea-S. Este gráfico muestra en colores el valor de correlación absoluta ( $p(\text{corr})$ ) de una variable con el modelo, y, por lo tanto, las variables que se acerquen más al color rojo, son variables que generan mayor diferencia entre los dos grupos de muestras. La Línea-S del modelo OPLS-DA se muestra en la Figura 28. De manera análoga al gráfico de cargas del modelo PLS-DA, se observa que las señales en 4.01, 7.56 y 7.69 ppm son las que más impactan al modelo y generan las diferencias entre las cohortes estudiadas. En la tabla 5 se enlistan los valores absolutos de correlación de las señales antes mencionadas en la *Línea-S*.

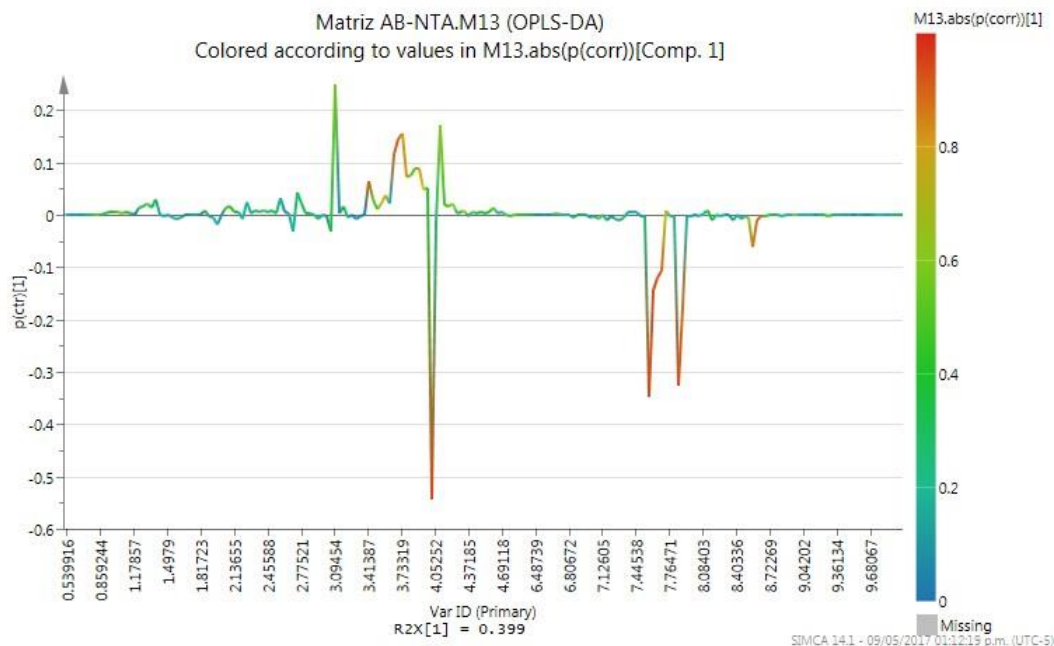


Figura 28. Línea-S del modelo OPLS-DA.

Tabla 5. Desplazamiento químico con valores de correlación mayor a 0.7 en la Línea-S.

Desplazamiento (ppm)	Correlación absoluta	Aumentado en el tratamiento con:
3.414	0.9113	Manzana
4.013	0.9413	Arándano
7.565	0.9360	Arándano
7.685	0.9031	Arándano
7.845	0.9394	Arándano
8.563	0.8654	Arándano

## 5.5. IDENTIFICACIÓN DE LAS SEÑALES

Los picos que obtuvieron el valor de correlación más alto se compararon con la literatura. Las señales que hacen la principal diferencia entre grupos se identificaron como señales del ácido hipúrico (3.97, 7.55, 7.64 y 7.83 ppm). Como ejemplo se muestran los espectros de las muestras de la voluntaria 2 (Figura 29), donde se puede observar la diferencia en las intensidades relativas de dichas señales.



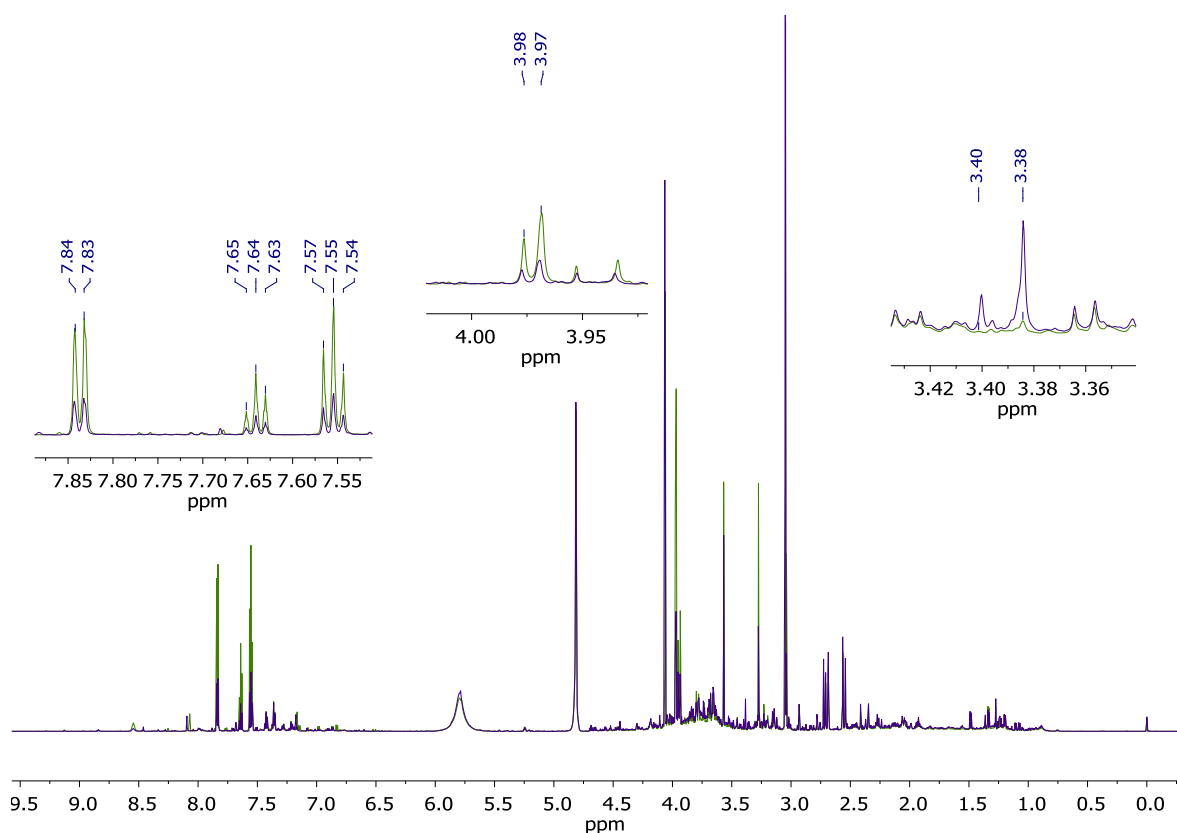


Figura 29. Espectros  $^1\text{H}$  RMN de la voluntaria 2 con tratamiento con jugo de arándanos (color verde) y con jugo de manzana (color morado). En la parte superior se muestran ampliaciones de las regiones donde aparecen con los picos causantes de la variación entre muestras obtenidos en la Línea-S.

En la *Human Metabolome Data Base*<sup>48</sup> se reporta el espectro del ácido hipúrico, cuya estructura química tiene como núcleo un anillo aromático monosustituido con un grupo amida, seguido de un metileno que conecta a un ácido carboxílico. En los espectros obtenidos se observan claramente las señales de este compuesto (coincidiendo el desplazamiento, la multiplicidad y la integración), con lo que se confirma que la concentración del ácido hipúrico cambia entre los tipos de muestra (Figura 30).

<sup>48</sup> Human Metabolome Data Base. Hippuric acid.  $^1\text{H}$  NMR Spectrum. [http://www.hmdb.ca/spectra/nmr\\_one\\_d/1494](http://www.hmdb.ca/spectra/nmr_one_d/1494). (Accesado el 8 de diciembre de 2016).

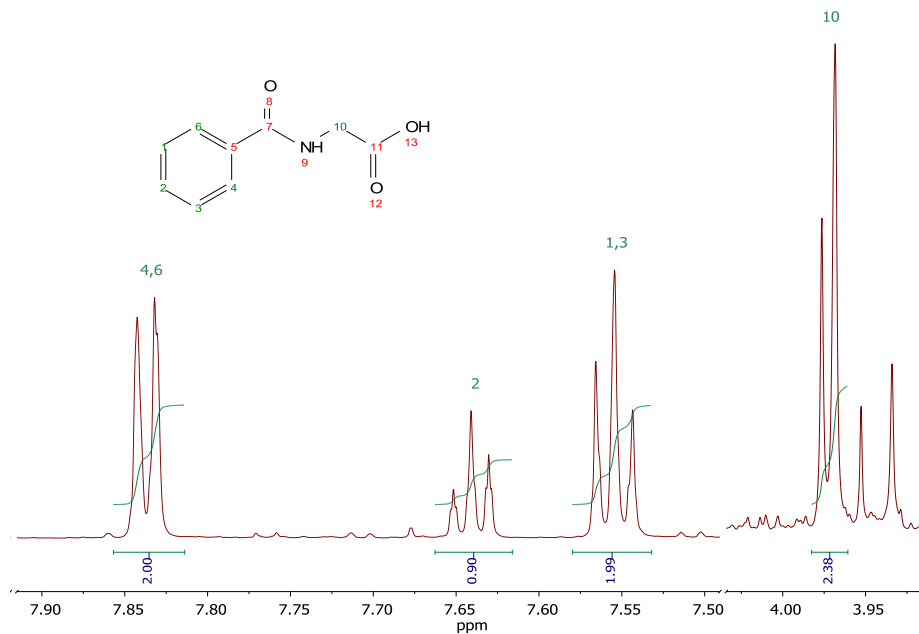


Figura 30. Expansión de las regiones de espectro 2A2 donde se encuentran las señales del ácido hipúrico. En la parte superior de muestra la estructura del ácido hipúrico

Al analizar en los espectros la señal obtenida en 3.41 ppm que en la línea-S mostró una correlación de 0.91 se encontró que estaba compuesta por dos señales en 3.38 y 3.40 ppm que aparecen en las muestras con tratamiento de jugo de manzana, pero no en las muestras blanco o con tratamiento de jugo de arándano. La diferencia en las intensidades relativas de ambos picos, indica que se trata de dos singuletes y no de un doblete (Figura 31). En dicha figura el espectro morado proviene de una muestra de tratamiento con manzana, mientras que el espectro verde proviene de una muestra de tratamiento con arándano.

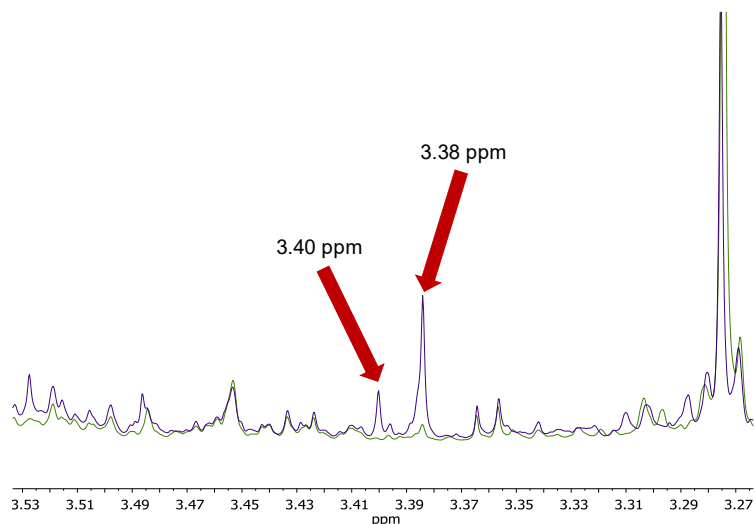


Figura 31. Ampliación del espectro 1H RMN de la voluntaria 2 con tratamiento con jugo de arándanos (color verde) y con jugo de manzana (color morado).

No se logró obtener más información de las bases de datos para ubicar a qué metabolito corresponden, por lo que se procedió a la adquisición de los espectros TOCSY y HSQC en una muestra de tratamiento de jugo de manzana. Al tratarse de dos singletes el espectro TOCSY de la muestra no reveló correlaciones con estos picos (espectro no mostrado), de manera que se confirmó que son singletes. Por otro lado, el espectro HSQC (Figura 32) muestra las correlaciones H-C a un enlace entre la señal a  $\delta_H=3.38$  con la señal a  $\delta_C=68.37$  y la señal a  $\delta_H=3.40$  con la señal a  $\delta_C=68.40$  ppm, que por desplazamiento químico, se propone que los singletes encontrados pertenecen a protones de grupos unidos a átomos electronegativos (como el oxígeno), y que podría tratarse de grupos metoxilos.

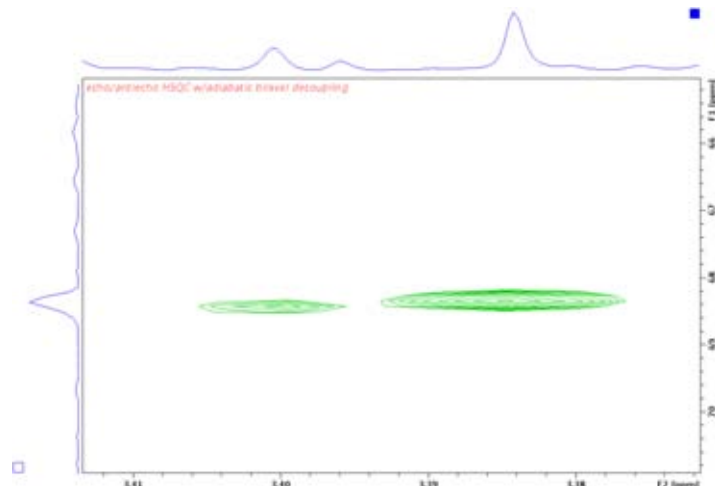


Figura 32. Expansión del espectro HSQC editado (700 MHz, H<sub>2</sub>O/D<sub>2</sub>O 90/10, 25°C, hsqcedetgsp). Ampliación en las señales de interés que al ser positivas indica se trata de CH o CH<sub>3</sub>

## 5.6. ANÁLISIS UNIVARIANTE

Comprobado que los picos que causan variación entre los grupos son correspondientes al ácido hipúrico, se seleccionó el pico representativo de entre el grupo de picos, 3.969 ppm, para llevar a cabo la prueba t de Student con un 95% de confianza.

Conforme al ácido hipúrico, se encontraron diferencias estadísticamente significativas ( $t_{(0.05, 10)} = 7.10$ ,  $P < 0.0001$ ) entre los tratamientos con jugo de arándano y manzana. La intensidad relativa en los picos de ácido hipúrico en los espectros de tratamiento con arándano es mayor que en los espectros de tratamiento con manzana. Como la intensidad relativa se encuentra relacionada con la concentración, la concentración de ácido hipúrico en las muestras de tratamiento con arándano es mayor que en las muestras de tratamiento con jugo de manzana. La gráfica de barras con los resultados se muestra en la Figura 33, señalando que la cantidad de ácido hipúrico en las muestras con tratamiento de jugo de arándano es mayor.

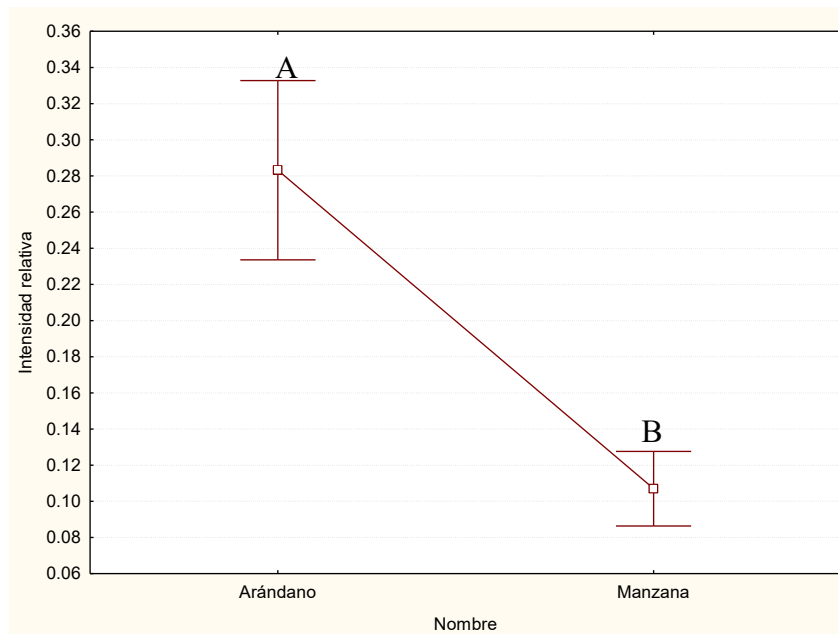


Figura 33. Resultados de la intensidad relativa del pico 3.969 ppm, correspondiente al ácido hipúrico.

Para los dos picos desconocidos se aplicó de igual manera la prueba t-Student a un 95% de confianza. Los resultados del Desconocido 1 se muestran en la Figura 34, donde se encontraron diferencias estadísticamente significativas entre los grupos de estudio ( $t_{(0.05, 10)} = 8.62$ ,  $P < 0.00001$ ). Los resultados del Desconocido 2 se muestran en la Figura 35, y al igual que con el compuesto Desconocido 1, se encontraron diferencias significativas ( $t_{(0.05, 10)} = 5.26$ ,  $P < 0.001$ ). Como se muestran en las figuras correspondientes, la intensidad relativa de las señales es mayor en las muestras de consumo de jugo de manzana, comparadas con las muestras de consumo de jugo de arándano.

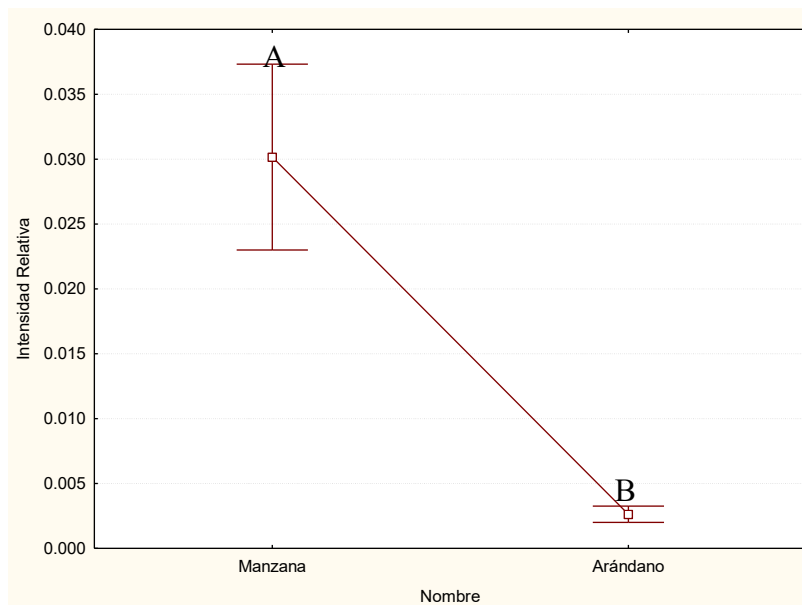


Figura 34. Resultados para la intensidad relativa del pico 3.38 ppm, correspondiente al compuesto Desconocido 1.

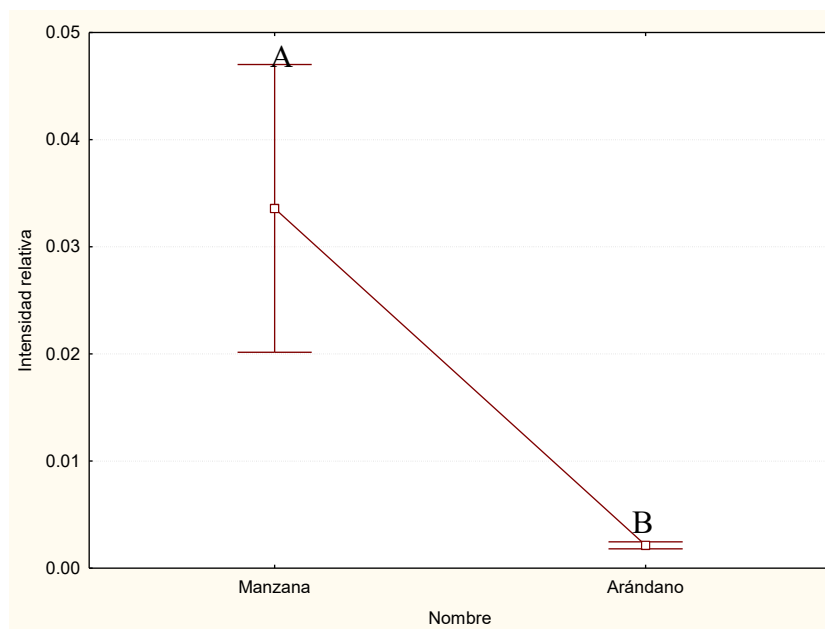


Figura 35. Resultados para la intensidad relativa del pico 3.40 ppm, correspondiente al compuesto Desconocido 2

Es importante resaltar que en el estudio de Florida se utilizó una sola marca de jugos (OceanSpray®), y este trabajo se utilizaron dos marcas de jugo. OceanSpray® para el de arándano, pero para el de manzana de utilizó la marca Sonrisa®, debido a que

no se comercializa el jugo puro de manzana de OceanSpray®. Puede que los compuestos encontrados en estos desplazamientos químicos se deban a la presencia de dichos compuestos o sus precursores en el jugo de manzana de la marca Sonrisa®.

La información hallada sobre los metabolitos que impactan el modelo según el consumo de jugo de arándano OceanSpray® o jugo de manzana Sonrisa® en este estudio se muestra de manera resumida en la tabla 6

**Tabla 6. Resumen de la información correspondiente a los metabolitos hallados.**

<i>Metabolitos</i>	<i>Desplazamiento químico y multiplicidad</i>	<i>Valor P</i>	<i>Grupo en que aumenta</i>
<i>Ácido hipúrico</i>	<i>3.97(d), 7.55 (t), 7.64 (t), 7.83 (d)</i>	<i>&lt;0.0001</i>	<i>Arándano</i>
<i>Desconocido 1</i>	<i>3.38 (s)</i>	<i>&lt;0.00001</i>	<i>Manzana</i>
<i>Desconocido 2</i>	<i>3.40 (s)</i>	<i>&lt;0.001</i>	<i>Manzana</i>



## 6. CONCLUSIONES

Se implementó la plataforma analítica de RMN y se logró la reproducción parcial del estudio llevado a cabo en la Universidad de Florida.

El preprocesamiento de las muestras con alineación, reducción dimensional y combinando ambas técnicas da como resultado diferentes parámetros de  $R^2X$ ,  $R^2Y$  y  $Q^2$ . El tipo de procesamiento que da una mayor capacidad predictiva es el uso de la alineación de señales con *binning* y una normalización al total de área. El valor de  $Q^2$  obtenido para el modelo PLS-DA con los tres tipos de muestras (blancos y regímenes alimentarios) fue de 0.85, mientras que con solo las muestras de los dos regímenes alimentarios fue de 0.91.

El análisis quimiométrico, en especial los modelos PLS-DA y OPLS-DA, mostraron variación entre las muestras según su procedencia y, a diferencia del artículo en el cual se basó esta investigación, se hallaron agrupaciones con coeficientes de predicción altos cuando se compararon los tres grupos de muestras (blancos, tratamiento de jugo de arándano y tratamiento de jugo de manzana).

Al igual que lo reportado para las mujeres jóvenes que participaron en el estudio de Florida, el principal compuesto responsable de la diferencia entre los regímenes alimentarios es el ácido hipúrico, que aumenta al consumir jugo de arándano. El análisis de resultados no evidenció la existencia del compuesto “desconocido 4”, reportado en el estudio de Florida, en ninguna muestra, lo cual se puede deber a una diferencia en la alimentación poblacional.

En nuestro estudio hay dos picos en 3.38 y 3.40 ppm que impactan al modelo, ausentes en el estudio de Florida, los cuales se sugiere pertenecen a protones de grupos unidos a átomos electronegativos, y se sospecha podrían provenir de una diferente alimentación, por ejemplo, el consumo de una marca de jugo diferente.



## 7. REFERENCIAS

1. Oszmianski, J. et. al. Effect of dried powder preparation process on polyphenolic content and antioxidant capacity of cranberry (*Vaccinium macrocarpon L.*). *Industrial Crops and Products*, **2015**, 77, 658-665.
2. Blatherwick, N. The specific role of foods in relation to the composition of the urine. *Arch. Inter. Med.* **1914**, 14, 409–450.
3. Foo, L. Y. et. al. A-type proanthocyanidin trimers from cranberry that inhibit adherence of uropathogenic P-Fimbriated *Escherichia coli*. *J. Nat. Prod.* **2000**, 63, 1225–1228.
4. Horgan, M.P.; Kenny, L.C. ‘Omics’ technologies: genomics, transcriptomics, proteomics and metabolomics. *The Obstetritian and Gynaecologist*. **2011**, 13(3), 189-195.
5. 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.
6. Duarte, I.F.; Diaz, S.O. and Gil, A.M. NMR metabolomics of human blood and urine in disease research. *J. Pharm. Biomed. Anal.* **2014**, 93, 17–26
7. Scrivo R, Casadei L, Valerio M, et al. Metabolomics approach in allergic and rheumatic diseases. *Curr. Allergy Asthma. Rep.* **2014**, 14(6), 445
8. Liu, H.; Tayyari, F.; Khoo, C.; Gu, L. A. <sup>1</sup>H NMR-based approach to investigate metabolomic differences in the plasma and urine of young women after cranberry juice or apple juice consumption. *J. Funct. Foods*, **2015**, 14, 76-86.
9. Medicina molecular. Glosario. <http://medmol.es/glosario/omicas/> (accesado el 31 de enero del 2016)
10. Barron, J.; Bermingham, K.; et. al. Twin metabolomics: the key to unlocking complex phenotypes in nutrition research. *Nutr. Res.* **2016**, 36, 291-304.
11. Verástegui-Escolano, C. Efectos de la acupuntura sobre la fatiga inducida por ejercicios físicos exhaustivos: una investigación metabolómica. *Revista Internacional de Acupuntura*. **2015**, 9(4), 131-133.
12. Zhou, X., et. al. Metabolomics identifies perturbations in amino acid metabolism in the prefrontal cortex of the learned helplessness rat model of depression. *Neuroscience*. **2017**, 343, 1-9.
13. Fiehn, O. Metabolomics-the link between genotypes and phenotypes. *Plant Mol. Biol.* **2002**. 48. 155–171.
14. Nagana-Gowda, G.A.; Raftery, D. Can NMR solve some significant challenges in metabolomics? *J. Magn. Reson.* **2015**, 260, 144-160.
15. Kuligowski, J.; Pérez-Guaita, D.; et. al. Infrared biospectroscopy for a fast qualitative evaluation of sample preparation in metabolomics. *Talanta*. **2014**, 127, 181-190.
16. Lu, X.; et. al. LC–MS-based metabonomics analysis. *J. Chromatogr.* **2008**, 866, 64–76.
17. Kim, S.; Kim, J.; Ju, E.; Kim, K.H. Food metabolomics: from farm to human. *Curr. Opin. Biotechnol.* **2015**, 37, 16-23.
18. Bordoni, A.; Capozzi, F. The foodomics approach for discovering biomarkers of food consumption in nutrition studies. *Curr. Opin. Food Sci*, **2015**, 4, 124-128.

19. Dunn, W.B., Ellis, D.I. Metabolomics: Current analytical platforms and methodologies. *Trends Anal. Chem.* **2005**, 24(4), 285-294.
20. McKay, R.T. How the 1D-NOESY suppresses solvent signal in metabolomics NMR spectroscopy: An examination of the pulse sequence. *Concepts Magn. Reson.* **2011**, 38A (5), 197-220.
21. Teng, Q. Chapter 9. NMR-based metabolomics. Structural biology: NMR Applications. 2<sup>nd</sup> ed. Springer. **2013**, 337, 339-340.
22. Car, H.Y.; Purcell, E.M. Effects of Diffusion on Free Precession In Nuclear Magnetic Resonance Experiments. *Phys. Rev.* **1954**, 94, 630-638.
23. Meighboom, S.; Gill, D. Modified Spin Echo Method for Measuring Nuclear Relaxation Times. *Rev. Sci. Instrum.* **1958**, 9 (8), 688-691.
24. Nagayama, K. et. al. Two-dimensional J-Resolved <sup>1</sup>H NMR Spectroscopy for Studies of Biological Molecules. *Biochem. Biol. Res. Commun.* **1977**, 8 (1), 99-105.
25. Alvaréz-Sánchez, B.; Priego-Capote, F.; Luque-de-Castro, M.D. Metabolomic analysis II. Preparation of biological samples prior to detection. *Trends Anal. Chem.* **2010**, 29(2), 120, 123.
26. Beckonert, O. et. al. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nature Publishing Group.* **2007**, 2(11), 2692-2703.
27. Mannina, L.; Sobolev, A.P.; Viel, S. Liquid state <sup>1</sup>H NMR in food analysis. *Prog. Nucl. Magn. Reson. Spectrosc.* **2012**, 66, 1-39.
28. Cuperlovic-Culf, M. 4. Metabolomics NMR Data Preprocessing – analysis of individual spectrum. NMR Metabolomics in Cancer Research. Woodhead Publishing Limits. Cambridge, **2013**; 215-231.
29. McKenzie, J.S. et. al. Analysis of complex mixtures using high-resolution nuclear magnetic resonance spectroscopy and chemometrics. *Prog. Nucl. Magn. Reson. Spectrosc.* **2011**, 59, 336-359.
30. D.L. Massart, B. G. M. V.; S.N. Deming, Y. Michotte, L. Kaufman, Chemometrics: A Textbook. B.G.M. Vandeginste, L. K. **2003**; 5.
31. Brereton, R. *Applied Chemometrics for Scientists*; Wiley: United Kingdom, **2007**.
32. Rasmussen, L. G. et. Al. Standarization of factors that influence human urine metabolomics. *Metabolomics.* **2011**, 7(1), 71-83.
33. Brereton, R.G. 10. Biological and Medical Applications of Chemometrics. Applied Chemometrics for Scientist. *Wiley.* **2010**, 287-318.
34. Trygg, J.; Wold, S. Orthogonal Projections to Latent Structures (O-PLS). *J. Chemom.* **2002**, 16, 119-128.
35. Functional Foods Science in Europe. British Journal of Nutrition. **1998**, 80(1), S1-S193.
36. Scientific Concepts of Functional Foods in Europe: Consensus Document. *Br J Nutr.* **1999**, 81(1), S1-S27.
37. García-Rubio J.C.; Ciordia-Ara M.; García G. El cultivo del arándano. *Serida*, **2007**.
38. Nicolle, L.E. Cranberry for Prevention of Urinary Tract Infection? *JAMA.* **2016**, 316(18), 1873-1874.
39. Kinney, A.B.; Blount, M. Effect of cranberry juice on urinary pH. *Nurs. Res.* **1979**, 28, 287-290.

40. Vasileiou, J. et al. Current clinical status on the preventive effects of cranberry consumption against urinary tract infections. *Nutr. Res.* **2013**, 33, 595-607.
41. Dixon, R.A., Xie, D.Y., Sharma, S. B. Proanthocyanidins – a final frontier in flavonoid research? *New phytologist.* **2005**, 165(1), 9-28.
42. Foo, L. Y.; Lu, Y.; Howell, A. B.; & Vorsa, N. A-type proanthocyanidin trimers from cranberry that inhibit adherence of uropathogenic P-Fimbriated Escherichia coli. *J. Nat. Prod.* **2000**, 63, 1225–1228.
43. Gupta, K.; Chou, M.Y.; Howell, A.; et al. Cranberry products inhibit adherence of p-fimbriated Escherichia coli to primary cultured bladder and vaginal epithelial cells. *J. Urol.* **2007**, 177(6), p.p. 2357-60.
44. Johnson, B.J.; Lin B.; Dinderman M.A. et al. Impact of cranberry on Escherichia coli cellular surface characteristics. *Biochem. Biophys. Res. Common.* **2008**, 19, 377(3), 992-4.
45. Jepson, R.G.; Craig, J.C. Cranberries for preventing urinary tract infections. *Cochrane Database of Systematic Reviews*, **2008**.
46. Bernard, K., Vulleumier, J.P., Burbacher, G. Zur Frage der Entstehung der Benzoesäure im Tierkörper. *Helv. Chim. Acta.* **1955**, 38, 1438.
47. Watson, D.G. Application of Principal Component Analysis to <sup>1</sup>H-NMR Data Obtained from Propolis Samples of Different Geographical Origins. *Phytochem. Anal.* 17. **2006**. 323-331.
48. Human Metabolome Data Base. Hippuric acid. <sup>1</sup>H NMR Spectrum. [http://www.hmdb.ca/spectra/nmr\\_one\\_d/1494](http://www.hmdb.ca/spectra/nmr_one_d/1494). (Accesado el 8 de diciembre de 2016).