



Universitat Autònoma  
de Barcelona

**Aplicación de Métodos Quimiométricos  
para la Caracterización y Control de Calidad  
de Plantas Medicinales.**

Juan Ricardo Lucio Gutiérrez

TESIS DOCTORAL

Programa de Doctorado de Química

Director: Dr. Jordi Coello Bonilla

Departamento de Química  
Facultad de Ciencias

2012



---

Memòria presentada per aspirar al Grau de Doctor per Juan Ricardo Lucio Gutiérrez

**Juan Ricardo Lucio Gutiérrez**

Vist i plau

**Dr. Jordi Coello Bonilla**  
Catedràtic Química Analítica

Bellaterra, 11 de Juliol de 2012.

---



## Agradecimientos

---

A Dios quién me da fuerza para enfrentar los retos que surgen en el cumplimiento de mi Leyenda Personal y para seguir por el sendero de la vida cada día.

A mi familia por su apoyo y comprensión, en esta estancia tan lejana.

Quiero empezar por mi asesor, el Dr. Jordi Coello, por toda la ayuda que me ha brindado a lo largo de este proyecto; por su asistencia a través de los (tediosos) trámites administrativos y por su tiempo, esfuerzo y paciencia. Así mismo, por las positivas contribuciones que ha realizado a mi formación profesional y como persona. También por haberme ayudado tanto, a través de sus explicaciones y consejos, con el proceso de adaptación al nuevo entorno cultural; una situación siempre complicada para un extranjero cuando se encuentra recién llegado a un País. También quiero hacer mención del Dr. Santiago MasPOCH por las, siempre certeras, observaciones y contribuciones hechas durante este proyecto. Así mismo, agradezco a los catedráticos que integran el Grupo de Quimiometría Aplicada, en especial al Dr. Marcelo Blanco, por las facilidades prestadas para que fuesen posibles los trabajos realizados.

A mis compañeros del Grupo de Quimiometría Aplicada, a los más antiguos y a los que han llegado en fechas más recientes y también a las personas con las que he convivido durante estos años en Barcelona; todos me han dejado alguna experiencia de vida y por lo tanto he aprendido algo nuevo de todos y cada uno, que de un modo u otro me ha enriquecido como persona. Espero tener la sabiduría para aplicar lo aprendido. No me es posible listar a todas las personas en este apartado, no quiero que nadie se sienta menospreciado; no obstante, me gustaría hacer mención de David Zamora por su amistad (siempre dispuesto a ofrecer asistencia, consejo y consuelo ante las adversidades personales que me aquejaban); al Sr. "Don" Ing. M. Sc. Francesc Vila García (Rancez), por su valiosa asistencia en el campo de la informática (me ahorro algunos euros en paracetamol); al Dr. José Amigo por su buena disposición para explicar aspectos técnicos de Matlab y de la quimiometría y a Isabel Osorio por su apoyo moral.

A mis profesores y amigos de la Facultad de Medicina de la U.A.N.L., en especial a la Dra. Ma. de la Luz Salazar Cavazos y a la Dra. Noemí Waksman, pues gracias a su apoyo fueron posibles las gestiones necesarias para realizar la Tesis Doctoral en España. Así mismo, agradezco a la Dra. Aurora Garza por su colaboración en el trabajo sobre *Turnera diffusa*. A mi estimada amiga Anabel Torres Cirio que siempre tiene palabras para motivarme.

---

## Agradecimientos

---

Agradezco a la Licenciada en Farmacia María José Alonso Osorio, vocal de plantas medicinales y homeopatía del Colegio Oficial de Farmacéuticos de Barcelona, por la orientación recibida sobre el uso y consumo de plantas medicinales en España y a la Empresa Amorós Nature por la donación realizada de diversas muestras autenticadas de plantas medicinales.

A las instituciones que posibilitaron el desarrollo de este proyecto de formación en Barcelona (España) por medio de una beca: La fundación Alβan - Universidad Autónoma de Nuevo León; “Con el apoyo del Programa Alβan, Programa de Becas de Alto Nivel de la Unión Europea para América Latina, beca nº E07D401919MX” (octubre de 2007 - julio de 2010) y el Consejo Nacional de Ciencia y Tecnología (Conacyt), beca nº 308657 (octubre de 2010 - abril de 2012). Al Ministerio de Ciencia y Tecnología de España por la financiación de los materiales necesarios para el desarrollo de esta Tesis en el marco del proyecto CTQ2007-26528.

---

---

***La felicidad es un viaje, no un destino.***

Robin S. Sharma

(El monje que vendió su Ferrari)

---





# Índice

---

<b>Resumen.....</b>	<b>xv</b>
<b>Abstract.....</b>	<b>xvii</b>
<b>1. Introducción.....</b>	<b>1</b>
1.1. Medicina tradicional.....	3
1.2. Plantas Medicinales.....	4
1.3. Situación actual y tendencias del control de calidad de las plantas medicinales...	5
1.4. Desafíos en el control de calidad de las plantas medicinales.....	7
1.5. Procedimientos analíticos empleados en el control de calidad de las plantas medicinales.....	9
1.5.1. Procedimientos de extracción.....	9
1.5.2. Técnicas instrumentales empleadas con los productos fitoterapéuticos.....	10
1.5.3. Espectroscopia en el infrarrojo cercano.....	12
1.5.4. Cromatografía de líquidos de alta resolución (CLAR).....	14
1.5.4.1. Factores que afectan el establecimiento del perfil cromatográfico..	14
1.6. Quimiometría.....	16
1.6.1. Pretratamientos de los datos obtenidos con los métodos instrumentales...	17
1.6.1.1. Variable aleatoria normal tipificada (SNV, Standard Normal Variate).....	17
1.6.1.2. Derivadas.....	17
1.6.1.3. Normalización.....	18
1.6.1.4. Escalado (Scaling).....	18
1.6.1.5. Corrección de la línea base (Baseline correction).....	18
1.6.1.6. Alineación optimizada mediante correlación (COW, Correlation Optimized Warping).....	19
1.6.1.7. Centrado (Mean centering).....	20
1.6.2. Métodos de reconocimiento de pautas.....	20
1.6.2.1. Análisis en componentes principales (PCA, Principal Component Analysis).....	20
1.6.2.2. Modelado blando independiente de analogías de clases (SIMCA, Soft Independent Modeling of Class Analogy).....	21
1.6.2.3. Análisis discriminante (DA, Discriminat Analysis).....	23

---

# Índice

---

1.6.3. Métodos de regresión.....	23
1.6.3.1. Regresión por mínimos cuadrados parciales (PLSR, Partial Least Squares Regression).....	23
1.6.3.2. Análisis discriminante con mínimos cuadrados parciales (PLS-DA, Partial Least Squares - Discriminat Analysis).....	24
1.6.4. Resolución multivariable de curvas - mínimos cuadrados alternados (MCR-ALS, Multivariate Curve Resolution - Alternating Least Squares)...	25
1.6.5. Verificación de los modelos y selección del número de factores.....	26
Referencias del capítulo.....	29
<b>2. Objetivos.....</b>	<b>35</b>
2.1. Objetivo General.....	37
2.2. Objetivos Específicos.....	38
<b>3. <i>Eleutherococcus senticosus</i>.....</b>	<b>39</b>
3.1. Introducción.....	41
3.1.1. Descripción de la planta medicinal.....	41
3.1.2. Problemática con el control de calidad del <i>Eleutherococcus senticosus</i> .....	41
3.1.3. Descripción del estudio.....	42
3.2. Metodología específica.....	42
3.2.1. Descripción de las muestras.....	42
3.2.2. Análisis espectroscópico en el infrarrojo cercano.....	44
3.2.3. Análisis multivariante.....	45
3.3. Resultados y discusión.....	47
3.3.1. Características de los espectros de infrarrojo cercano.....	47
3.3.2. Estudio del agrupamiento natural de las muestras.....	49
3.3.3. Modelado blando independiente de analogías de clases (SIMCA).....	50
3.3.4. Análisis discriminante con mínimos cuadrados parciales (PLS-DA).....	54
3.3.5. Análisis discriminante (DA).....	57
3.4. Conclusiones del capítulo .....	59
Referencias del capítulo.....	61

---

# Índice

---

<b>4. <i>Panax ginseng</i>.....</b>	<b>63</b>
4.1. Introducción.....	65
4.1.1. Descripción de la planta medicinal.....	65
4.1.2. Problemática con el control de calidad del <i>Panax ginseng</i> .....	65
4.1.3. Descripción del estudio.....	66
4.2. Metodología específica.....	67
4.2.1. Descripción de las muestras.....	67
4.2.2. Análisis espectroscópico en el infrarrojo cercano.....	69
4.2.3. Análisis multivariante.....	70
4.3. Resultados y discusión.....	72
4.3.1. Características de los espectros de infrarrojo cercano.....	72
4.3.2. Estudio de la similitud de las muestras.....	73
4.3.3. SIMCA (Soft Independent Modeling of Class Analogy).....	75
4.3.4. PLS-DA (partial least squares - discriminant analysis).....	76
4.3.5. DA (discriminant analysis).....	77
4.3.6. Análisis semicuantitativo de mezclas.....	78
4.4. Conclusiones del capítulo.....	80
Referencias del capítulo.....	83
<b>5. <i>Valeriana officinalis</i>.....</b>	<b>85</b>
5.1. Introducción.....	87
5.1.1. Descripción de la planta medicinal.....	87
5.1.2. Problemática con el control de calidad de <i>Valeriana officinalis</i> .....	87
5.1.3. Antecedentes.....	88
5.1.4. Descripción del estudio.....	89
5.2. Metodología específica.....	90
5.2.1. Descripción de las muestras.....	90
5.2.2. Reactivos.....	91
5.2.3. Preparación de las muestras y de las mezclas de laboratorio.....	92
5.2.4. Instrumentación y procedimiento cromatográfico.....	92
5.2.5. Análisis multivariante.....	93
5.2.6. Construcción de las huellas dactilares con múltiples longitudes de onda...	95
5.2.6.1. Reducción de variables.....	95
5.2.6.2. Ajuste de la línea de base y normalización de los datos.....	96

---

# Índice

---

5.2.6.3. Desdoblamiento de la matriz (unfolding).....	96
5.2.7. Optimización de los parámetros para COW (correlation optimized warping).....	97
5.3. Resultados y discusión.....	97
5.3.1. Optimización y validación de la metodología.....	97
5.3.2. Características de los cromatogramas y de las huellas dactilares aumentadas.....	99
5.3.3. COW (correlation optimized warping) dirigida mediante PCA (principal component analysis).....	104
5.3.4. Comparación de los resultados de los modelos de clasificación.....	107
5.4. Conclusiones del capítulo.....	109
Referencias del capítulo.....	111
<b>6. Turnera diffusa.....</b>	<b>113</b>
6.1. Introducción.....	115
6.1.1. Descripción de la planta medicinal.....	115
6.1.2. Problemática con el control de calidad de <i>Turnera diffusa</i> .....	115
6.1.3. Antecedentes.....	116
6.1.4. Descripción del estudio.....	117
6.2. Metodología específica.....	117
6.2.1. Descripción de las muestras.....	117
6.2.2. Reactivos.....	117
6.2.3. Preparación de los extractos.....	118
6.2.4. Instrumentación y procedimientos analíticos.....	118
6.2.4.1. Procedimiento por CLAR-DAD.....	118
6.2.4.2. Determinación de la actividad captadora del radical libre DPPH mediante espectrofotometría.....	119
6.2.5. Análisis multivariante.....	120
6.2.6. Construcción de las huellas dactilares con múltiples longitudes de onda...	121
6.2.7. Alineación de las huellas dactilares aumentadas mediante COW.....	121
6.3. Resultados y discusión.....	122
6.3.1. Resultados del ensayo con DPPH y del método CLAR.....	122
6.3.2. Construcción de las huellas dactilares aumentadas.....	122
6.3.2.1. Selección de las longitudes de onda de trabajo.....	123

---

## Índice

---

6.3.2.2. Efecto de los pretratamientos.....	125
6.3.2.3. Alineación.....	126
6.3.3. Predicción de la actividad antioxidante (EC <sub>50</sub> ) por PLSR.....	128
6.4. Conclusiones del capítulo.....	131
Referencias del capítulo.....	133
<b>7. Conclusiones generales.....</b>	<b>135</b>
<b>8. Perspectivas.....</b>	<b>139</b>
<b>Anexo.....</b>	<b>143</b>

---



## Resumen

---

La medicina tradicional herbolaria y sus preparaciones han sido ampliamente empleadas por miles de años; sin embargo en muchos de los países occidentales aún no ha sido oficialmente aceptada. Esto es debido a que no hay datos suficientes de eficacia y seguridad que cumplan con los criterios requeridos para que su uso sea universal. No obstante, la Organización Mundial de la Salud ha confirmado que el 80 % de la población mundial, especialmente millones de personas de las áreas rurales de los países en desarrollo, hacen uso de las medicinas herbales para sus necesidades de salud. La seguridad en el consumo de un producto herbal debe ser otorgada por un proceso completo de control de calidad; este debe ser un proceso estandarizado, detallado, claro, conciso y reproducible. De no existir este proceso pueden surgir confusiones entre los productos que pueden resultar peligrosas para la salud del usuario. La falta de datos de investigación se debe, no sólo a políticas de salud pública, sino también a la falta de control de calidad de estos productos; debido principalmente a que no se dispone de metodologías adecuadas y aceptadas para la evaluación de los mismos. Uno de los motivos para la falta de métodos aceptables podría ser que, a diferencia de los compuestos de síntesis de alta pureza, una fórmula herbal puede consistir en cientos de compuestos. Con ayuda de la quimiometría se puede definitivamente tener una mayor oportunidad de afrontar las dificultades que surgen al emplear técnicas cromatográficas o espectroscópicas para el análisis de calidad de las plantas medicinales, como son las señales no resueltas y la gran cantidad de componentes en las plantas.

En esta memoria, se presenta la aplicación de diversos métodos quimiométricos para realizar el control de calidad de cuatro plantas medicinales de amplio uso comercial, para las cuales existen informes de confusiones, adulteraciones o presentación de reacciones adversas en los consumidores: *Panax ginseng*, *Eleutherococcus senticosus*, *Valeriana officinalis* y *Turnera diffusa*. Los capítulos guardan relación entre sí, pero han sido estructurados para ser autónomos, presentando la descripción y problemática de cada planta, metodología, resultados y bibliografía particulares.

## Resumen

---

El primer capítulo de la tesis consiste en una introducción general sobre el uso de la medicina tradicional, específicamente de las plantas medicinales, y sobre las diferentes estrategias que han sido utilizadas para su control de calidad, incluyendo una revisión de los aspectos analíticos y finalizando con la descripción de los métodos quimiométricos usados en los trabajos presentados en los capítulos. El capítulo dos consiste en los objetivos que persigue esta Tesis Doctoral.

Los capítulos tres y cuatro describen el uso de la espectroscopia de infrarrojo cercano como técnica analítica en el control de calidad del *Eleutherococcus senticosus* y el *Panax ginseng*, respectivamente, junto con varios métodos de reconocimiento de pautas (análisis en componentes principales, modelado blando independiente de analogías de clases y análisis discriminante) y el método de regresión de análisis discriminante con mínimos cuadrados parciales. Para el *Panax ginseng* (capítulo 4) se describe también el uso de la resolución multivariable de curvas - mínimos cuadrados alternados con el fin de realizar la cuantificación de mezclas.

En el capítulo cinco se describe un tratamiento de la información que implica el uso de cromatogramas a varias longitudes de onda para la construcción de huellas dactilares aumentadas de *Valeriana officinalis*, con la correspondiente evaluación de los resultados obtenidos de la clasificación efectuada con los métodos de reconocimiento de pautas.

El capítulo seis trata sobre la aplicación de la estrategia de huellas dactilares aumentadas, junto con la regresión por mínimos cuadrados parciales, para mejorar la predicción de la actividad antioxidante de *Turnera diffusa*.

El capítulo siete presenta las conclusiones globales obtenidas de los métodos de control de calidad propuestos para las cuatro plantas medicinales en estudio y el capítulo ocho describe las perspectivas que abre esta Tesis Doctoral, proponiendo diversas líneas de trabajo. Finalmente, en el anexo se lista las publicaciones resultantes de los trabajos realizados.



## Abstract

---

Traditional herbal medicine and its preparations have been widely used for thousands of years, but in many Western countries it has not yet been officially accepted. The reason is a lack of data about their efficacy and security that are required for a universal use. However, the World Health Organization has confirmed that 80% of the people in the world, especially millions who live in the rural areas in developing countries, make use of herbal medicines for their needs of health. Safety in the use of an herbal product should be granted for a complete process of quality control that should be a standardized, detailed, clear, concise and reproducible process. Its absence may arise confusion between products that may be hazardous to the health of the user. The lack of research data is due not only to public health policy but also to the deficient quality control of these products, mainly due to the unavailability of suitable methodologies for the evaluation thereof. One reason for so little acceptable methods may be that, unlike the synthesis of compounds of high purity, an herbal formulation may consist of hundreds of compounds. With the aid of chemometrics there is a greater opportunity to face the difficulties that arise when using chromatographic or spectroscopic techniques for the analysis of quality of the medicinal plants, such as unresolved signals and the large number of components in plants.

In this thesis, it is presented the application of different chemometric methods for quality control of four medicinal plants in widespread commercial use for which there are reports of confusion, adulteration or presentation of adverse reactions in consumers: *Panax ginseng*, *Eleutherococcus senticosus*, *Valeriana officinalis* and *Turnera diffusa*. The chapters are interrelated, but have been structured to be self-describing; presenting description, problems, methodology, results and literature for each plant.

The first chapter of the thesis is a general introduction to the use of traditional medicine, particularly medicinal herbs, and about the different strategies that have been used for its quality control, including a review of analytical issues and finally, the description of the chemometric methods used in the studies presented in each chapter. Chapter two presents the objectives of this thesis.

## Abstract

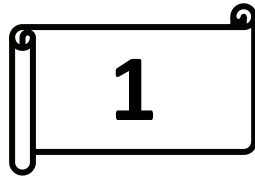
---

Chapters three and four describe the use of near infrared spectroscopy as an analytical technique in the quality control of *Eleutherococcus senticosus* and *Panax ginseng*, respectively, along with various methods of pattern recognition (Principal Component Analysis, Soft Independent Modeling of Class Analogy and Discriminant Analysis) and the regression method of Partial Least Squares - Discriminant Analysis. For *Panax ginseng* (Chapter 4) is also described the quantification of mixtures using Multivariate Curve Resolution - Alternating Least Squares.

Chapter five describes an information processing that involves the use of chromatograms at various wavelengths for the construction of enhanced fingerprints of *Valeriana officinalis*, with the corresponding evaluation of the results from the classification conducted by pattern recognition methods.

Chapter six deals with the implementation of the enhanced fingerprint strategy together with Partial Least Squares Regression, in order to improve the prediction of the antioxidant activity of *Turnera diffusa*.

Chapter seven presents the overall conclusions drawn from the quality control methods proposed for the four medicinal plants studied and Chapter eight describes the prospects offered by this thesis, proposing several possibilities for further research. Finally, the appendix lists the publications resulting from the work carried out.



# Introducción.

---



### **1.1. Medicina tradicional.**

Se puede definir la medicina tradicional como la suma total de los conocimientos, habilidades y prácticas basadas en las teorías, creencias y experiencias indígenas de distintas culturas (sean o no explicables), que se utilizan en la preservación de la salud; así como en la prevención, el diagnóstico, la mejora o el tratamiento de enfermedades físicas y mentales. En algunos países, donde la medicina tradicional no está incorporada al sistema predominante para el cuidado de la salud, los términos “medicina complementaria”, “alternativa” o “no convencional” son usados de manera indistinta [1]. En las últimas décadas, el uso de la medicina tradicional se ha expandido a nivel mundial debido a que ha ganado popularidad y, en consecuencia, no solamente está siendo utilizada para la atención primaria de la salud de los pobres en los países en desarrollo, sino que también se está utilizando en los países desarrollados donde la medicina convencional es predominante. Considerando la importante expansión mundial en el uso de la medicina tradicional, la seguridad, la eficacia así como el control de calidad de las medicinas tradicionales y los procedimientos basados en terapias tradicionales, se han convertido en tópicos relevantes tanto para las autoridades sanitarias como para los usuarios de estos remedios [2].

Se considera que la seguridad y la eficacia de la medicina tradicional ha sido demostrada a través del uso histórico; es decir, a través de la experiencia transmitida de generación en generación. No obstante, se reconoce que es necesaria la investigación científica para proporcionar evidencia adicional de su seguridad y eficacia [3]. Sin embargo, diversas prácticas de la medicina tradicional no han sido oficialmente reconocidas en la mayoría de los países occidentales; debido a esto, no existe un desarrollo paralelo de normas internacionales y métodos apropiados para su evaluación en diferentes culturas y regiones. Consecuentemente, la cantidad y calidad de los datos de seguridad y eficacia de la medicina tradicional están lejos de ser suficientes para cumplir con los criterios necesarios para apoyar su uso en todo el mundo. De esta manera, el reto actual es asegurar que la medicina tradicional es

empleada correctamente y determinar cómo debe de llevarse a cabo la investigación y evaluación en este campo [4].

## **1.2. Plantas Medicinales.**

La medicación tradicional normalmente implica el uso de plantas medicinales, partes de animales y minerales; siendo las plantas medicinales el recurso terapéutico más ampliamente utilizado [5]. Así, de las diversas prácticas de la medicina tradicional, el presente trabajo se focaliza en el control de calidad de las plantas medicinales.

Históricamente las plantas han sido utilizadas como medicamentos. Durante mucho tiempo los remedios naturales y sobre todo las plantas medicinales, fueron el principal e incluso el único recurso del que disponían los médicos para el tratamiento de enfermedades, por lo que las plantas medicinales forman parte esencial de las estrategias utilizadas por la población de países en desarrollo para enfrentar sus enfermedades cotidianas. Se estima que han sido identificadas de 250.000 a 350.000 especies de plantas en el mundo y que cerca de 35.000 son ampliamente usadas con propósitos medicinales. El uso de plantas medicinales se da no solamente en el medio indígena y rural, sino en las zonas urbanas y suburbanas, como resultado de la necesidad de recursos accesibles para enfrentar muy diversos padecimientos [6]. La Organización Mundial de la Salud (OMS) ha confirmado que el 80 % de la población mundial hace uso de la fitoterapia para atender sus necesidades de salud [3].

Por otro lado, las plantas medicinales y sus preparaciones han sido ampliamente usadas en muchos países orientales y su popularidad se ha extendido en las últimas décadas a países occidentales, aunque sin ser aún aceptadas oficialmente en muchos de ellos; esto es debido principalmente a que no hay datos suficientes de eficacia y seguridad. Las razones para la falta de estos datos incluye la carencia de políticas de salud pública y de metodologías adecuadas y aceptadas para la evaluación de las plantas medicinales. La escasez de metodologías puede ser debida a que, a diferencia de los compuestos de síntesis de alta pureza, un medicamento tradicional a base de plantas puede consistir en cientos de compuestos [7].

### **1.3. Situación actual y tendencias del control de calidad de las plantas medicinales.**

El creciente interés en el uso de plantas medicinales está dando lugar a un mercado de rápido crecimiento para los productos fitoterapéuticos, suplementos dietéticos, nutracéuticos y alimentos funcionales [8]. Esta importante expansión en el uso de plantas medicinales plantea muchas inquietudes acerca de su control de calidad, ya que podría dar lugar a la recolección masiva y sin supervisión, sin ninguna consideración de la calidad de las plantas usadas como materias primas [9]. La seguridad en el consumo de un producto fitoterapéutico debe ser garantizada por un proceso completo de control de calidad. Este debe ser un proceso estandarizado, detallado, conciso y reproducible que pueda ser utilizado para evaluar a otros productos de la misma naturaleza. El aseguramiento de calidad de productos fitoterapéuticos, es un prerrequisito para realizar ensayos clínicos con resultados reproducibles y es necesario con el fin de identificar adulteraciones, falsificaciones o confusiones entre plantas [3,6,10].

Una gran parte de los productos fitoterapéuticos son clasificados como suplementos dietéticos, por lo que no están sujetos a los mismos procesos de monitorización de los medicamentos; es por ello que en Estados Unidos, la FDA (Food and Drug Administration) emitió la guía DSHEA (Dietary Supplement Health and Education Act of 1994) para la regulación de este tipo de productos. La guía permite a las compañías hacer afirmaciones sobre los beneficios de sus productos en la conservación de la salud. Sin embargo, no revisa que se cumplan los estándares de control de calidad en cuanto a seguridad, eficacia y pureza [11].

La OMS, en su manual de métodos de control de calidad para plantas medicinales, la FDA y las Farmacopeas de diversos países presentan una colección de procedimientos recomendados para evaluar la identidad, pureza y contenido de los materiales de plantas medicinales; con estos procedimientos intentan proporcionar a los laboratorios una herramienta para llevar a cabo el control de calidad de los productos fitoterapéuticos. Los parámetros que incluyen estas referencias para describir la calidad de los productos a base de plantas usados como medicamentos

son: a) pruebas de identidad, donde consideran las características macroscópicas, microscópicas y organolépticas, así como el perfil obtenido por cromatografía en capa fina; b) pruebas de pureza para determinar humedad, cenizas, materia extraña, disolventes, la presencia de contaminantes microbianos, aflatoxinas, metales pesados y residuos de plaguicidas; y c) la valoración de principios activos y/o marcadores [1,6,10,12].

La mayoría de las plantas medicinales han sido autenticadas tradicionalmente por medios morfológicos e histológicos. Sin embargo, este enfoque puede no ser fiable porque las plantas de la misma familia son morfológicamente similares y los productos comerciales elaborados con plantas se preparan de diversas maneras que hacen más difícil su identificación a través de una inspección sensorial [13]. Actualmente, es una práctica común entre los analistas de productos naturales la selección de uno o más compuestos como marcadores para la identificación y evaluación de la calidad [14]. Sin embargo, es reconocido que este tipo de determinaciones no proporciona por sí una idea completa de un producto farmacéutico a base de plantas, ya que son múltiples los constituyentes responsables de sus efectos terapéuticos y a veces la selección de marcadores adecuados es difícil y subjetiva [15,16]. Estos constituyentes pueden trabajar sinérgicamente y pueden variar dependiendo no únicamente de la especie, sino también de las condiciones de crecimiento, de la época de cosecha, del origen, de los métodos de proceso y secado y del tiempo de almacenaje, entre otros factores [17,18]. Por otra parte, algunos productores sin escrúpulos continuamente están tratando de desarrollar maneras de hacer el perfil químico de sus productos similar al perfil auténtico presentado por la planta medicinal en cuestión. En estas circunstancias específicas, el uso de compuestos marcadores es incapaz de confirmar la identidad de una planta. Es por estas razones que se hace necesario determinar la mayoría de los constituyentes de las plantas medicinales con el propósito de asegurar la confiabilidad y precisión de las investigaciones farmacológicas y clínicas, para conocer su bioactividad y los posibles efectos colaterales de los compuestos activos, así como mejorar el control de calidad de sus productos [5].



El establecimiento de huellas dactilares para las plantas medicinales es un enfoque relativamente reciente adoptado por los investigadores para hacer frente a la complejidad química de los productos fitoterapéuticos y los problemas relacionados con el uso de pocos compuestos marcadores. La huella dactilar de una planta medicinal podría ser un espectro o un cromatograma obtenido por un procedimiento definido para caracterizar la composición química de una muestra; se supone que es una característica única para la identificación de las plantas, extractos de plantas o preparados en cuestión [16,19]. Así pues, el perfil cromatográfico o el perfil espectroscópico de la planta medicinal, parece ser la herramienta más conveniente para llevar a cabo el control de calidad de los productos de uso farmacéutico a base de plantas [20-24]. Normativas actualizadas sobre plantas medicinales publicadas por la Organización Mundial de la Salud (OMS), la Agencia de Medicinas y Alimentos de los Estados Unidos (Food and Drug Administration, FDA) y la Agencia Europea del Medicamento (EMA) se pronuncian por el uso de las huellas dactilares; indican que la identificación de la planta medicinal debe ser una de las primeras pruebas aplicadas para garantizar la calidad y para discriminar entre especies relacionadas o muestras adulteradas [3,12,25,26].

#### **1.4. Desafíos en el control de calidad de las plantas medicinales.**

La metabolómica de las plantas puede ser empleada para obtener huellas dactilares de las especies con propósitos bioquímicos o taxonómicos, ya que el metaboloma es la expresión final de los genes y define el fenotipo bioquímico de una célula o tejido; así, la determinación cualitativa y cuantitativa de los metabolitos celulares proveen una visión del estatus bioquímico de un organismo dado [27]. El concepto de fitoequivalencia hace uso del fenotipo bioquímico y fue desarrollado en Alemania para asegurar la congruencia entre la composición de distintos lotes de medicamentos tradicionales basados en plantas. Según este concepto, para ser considerada fitoequivalente, una planta medicinal requiere presentar un perfil químico que debe ser comparable con el perfil de un producto ya conocido; con este fin se pueden aplicar diversas técnicas cromatográficas y espectroscópicas, considerando como componente activo la planta medicinal en su totalidad [28-32].

La huella dactilar debe proporcionar las atribuciones fundamentales de integridad y semejanzas o diferencias para representar químicamente el producto fitoterapéutico investigado, permitiendo autenticar el material vegetal. Las huellas dactilares han de ser capaces de demostrar las semejanzas y diferencias entre varias muestras de una planta medicinal en particular, incluso si el número y/o la concentración de los componentes químicamente característicos no son exactamente iguales entre ellas, de tal manera que, al analizar la huella dactilar de una planta medicinal, se consideran los múltiples constituyentes y no solamente uno o dos marcadores durante el proceso de evaluación de la calidad de estos productos [33-36]. Sin embargo, en cualquier producto herbal existen cientos de componentes desconocidos y muchos de ellos se encuentran en baja concentración. Además, normalmente existen variabilidades entre muestras de la misma planta medicinal; en consecuencia, obtener huellas dactilares confiables que representen componentes químicamente característicos de un producto fitoterapéutico no es tarea fácil. Por ello, en el desarrollo de métodos para el análisis de plantas medicinales aparecen una serie de desafíos:

- El análisis de marcadores o componentes activos en una matriz compleja y algunas veces desconocida.
- Los analitos buscados pueden tener problemas en su extracción con disolventes o ser térmicamente lábiles.
- La falta de sustancias químicas de referencia y/o materiales de referencia certificados.
- La selección adecuada del método de extracción.
- La variación en la composición de la planta obtenida entre lote y lote [37,38].

Además, la seguridad, eficacia y calidad de productos finales a base de plantas medicinales dependen de la calidad de las materias primas y de cómo se manejan a través de procesos de producción. Considerando lo anterior, la presente memoria se centra en el control de calidad de las plantas medicinales en forma de materias primas.

### **1.5. Procedimientos analíticos empleados en el control de calidad de las plantas medicinales.**

En general, los métodos de control de calidad de las plantas medicinales deberían incluir, además de la inspección sensorial, ensayos analíticos tales como la cromatografía en capa fina, la cromatografía de gases, la cromatografía de líquidos de alta resolución, espectros de infrarrojo medio y análisis mediante otras espectrometrías diversas, como el infrarrojo cercano o resonancia magnética nuclear. En la presente memoria se han utilizado las técnicas de la espectroscopia de infrarrojo cercano y de la cromatografía de líquidos de alta resolución y se hace énfasis en sus aplicaciones para el control de calidad de las plantas medicinales.

#### **1.5.1. Procedimientos de extracción.**

El método de extracción y preparación de muestra es muy importante para obtener buenos resultados, ya que si se quiere tener una buena información del perfil de una planta medicinal es necesario contar con un método que extraiga la mayoría de los componentes activos de las plantas, que garantice la integridad de la actividad farmacológica y, a la vez, que sea un método preciso y ofrezca buenas recuperaciones [39]. En los productos fitoterapéuticos, los compuestos de interés suelen presentar grandes diferencias en cuanto a polaridad y termoestabilidad. Por ello, aún con una misma técnica de extracción las condiciones de trabajo (solvente, temperatura, etc.) pueden variar de un material vegetal a otro, lo que hace indispensable evaluar cuidadosamente la eficiencia del método empleado. El calentamiento con reflujo y la extracción soxhlet se encuentran entre los métodos más empleados para productos naturales. Sin embargo, este tipo de métodos son lentos, requieren grandes cantidades de solventes y pueden tener eficiencias de extracción bajas [40].

En la última década, se han desarrollado técnicas como la microextracción en fase sólida (SPME) o la microextracción en fase líquida (LPME) cuyas principales ventajas son la disminución en el consumo de disolventes y en los tiempos de análisis, su relativo bajo costo, su simplicidad y en muchos casos la posibilidad de obtener

factores de enriquecimiento mayores [41]. La SPME fue introducida en los años 90 y se ha convertido en una técnica muy utilizada para la preparación de muestras para cromatografía de gases y en menor medida para cromatografía líquida. En el análisis de productos fitoterapéuticos esta técnica ha sido empleada principalmente para caracterizar compuestos volátiles [41-46]. Se han publicado algunas aplicaciones de la SPME en modo headspace para la obtención de la huella dactilar de compuestos volátiles en plantas medicinales chinas y en otros en vegetales [47,48].

### **1.5.2. Técnicas instrumentales empleadas con los productos fitoterapéuticos.**

Con respecto a las técnicas de análisis de las plantas medicinales, las cromatográficas tienen una muy poderosa capacidad de separación para los complejos componentes químicos de los extractos, siendo posible separarlos en muchas subfracciones relativamente simples. La cromatografía de capa fina (CCF) ha sido la técnica de elección para el análisis de plantas medicinales y es aún usada con mucha frecuencia [6,10]. Con las limitaciones habituales de la técnica, la CCF es usada como un método de rastreo y semicuantitativo que provee perfiles cromatográficos característicos y permite identificar adulteraciones en los productos. Las principales ventajas de esta técnica es que es sencilla, rápida, versátil y permite analizar muchas muestras a la vez [16,49]. No obstante, su aplicación regularmente se limita a la identificación de compuestos marcadores, con las desventajas que ello conlleva.

Es bien conocido que muchos de los principios farmacológicamente activos en las plantas medicinales son compuestos volátiles, por lo que su estudio es de gran importancia en el análisis de medicinas tradicionales a base de plantas. La cromatografía de gases (CG) ha sido empleada en el análisis de compuestos volátiles de las plantas medicinales con propósitos de establecer perfiles de identificación de las mismas, ya que la composición y concentración relativa de los compuestos orgánicos volátiles (aceites) son característicos de algunas plantas. Cambios en la composición de los aceites volátiles, pueden indicar adulteraciones o indicar cambios enzimáticos de oxidación o fermentación [5,50,51]. El empleo del acoplamiento con espectrometría de masas (CG-EM) para establecer perfiles cromatográficos de plantas ha sido

empleado por varios autores [7,43,45,47,52]. El establecimiento de estos perfiles se realiza mediante comparación de los tiempos de retención y los espectros de masas de los componentes comunes en los especímenes. La CG-EM tiene la ventaja de que proporciona, además, información de identidad de los compuestos. La principal desventaja de la CG es que no puede ser empleada para análisis de compuestos polares y no volátiles, a menos que se realice una derivatización previa de la muestra [5].

La electroforesis capilar (EC) fue introducida a principios de los 80 como una poderosa técnica analítica de separación que ha sido desarrollada rápidamente. Los métodos electroforéticos, especialmente la EC, han sido empleados en el análisis de plantas medicinales en la última década [53]. La EC permite la separación y análisis de los ingredientes activos en las plantas medicinales; tiene la ventaja de que requiere poca cantidad de muestra y permite realizar análisis rápidamente con muy buena capacidad de separación [54]. También es una buena herramienta para la generación de huellas dactilares de las plantas medicinales. Estudios recientes de alcaloides y flavonoides en plantas medicinales han empleado la EC [55].

Si bien los métodos cromatográficos han sido ampliamente utilizados para establecer la huella dactilar en productos fitoterapéuticos, ha habido avances en el uso de la resonancia magnética nuclear (RMN) de alta resolución y la espectroscopia de infrarrojo. Las huellas dactilares obtenidas por  $^1\text{H}$ RMN han sido utilizadas, sobre todo, en el análisis de alimentos naturales como vino, aceites de oliva, cervezas, entre otros. El método es versátil, no depende de la naturaleza de la muestra y se pueden obtener huellas dactilares cuantitativas [56]. Permite detectar simultáneamente todos los componentes en la muestra que tengan núcleos de hidrógeno, lo cual cubre prácticamente todos los compuestos orgánicos como carbohidratos, aminoácidos, ácidos grasos, aminas, ésteres, éteres, fenoles y lípidos [57]. *Gingko biloba* es una de las plantas sobre la cual se han publicado mayor número de trabajos con esta técnica, hasta el momento, con sensibilidades semejantes a las obtenidas por CLAR, pero con menor tiempo de análisis y evitando el consumo de insumos caros (solventes y columnas) [58]. Además, actualmente se dispone de programas para el procesamiento

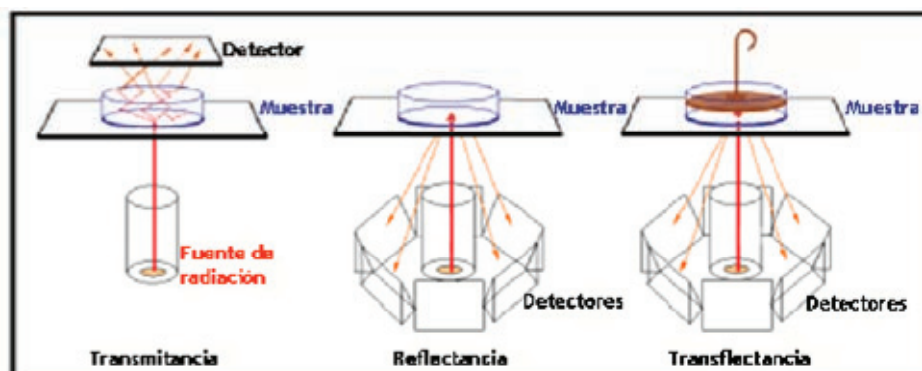
de los datos post-adquisición en forma *off-line*, lo cual permite mejorar la resolución, obtener una óptima relación señal/ruido y aumentar la sensibilidad [59-61].

### 1.5.3. Espectroscopia en el infrarrojo cercano.

La región del infrarrojo cercano (NIR), se extiende entre 800 y 2500 nm o, expresado en número de onda, de 12 500 a 4000  $\text{cm}^{-1}$ . Es la primera región espectral que muestra bandas de absorción relacionadas con las vibraciones moleculares y se caracteriza por las bandas armónicas y las bandas de combinación. Esta región se utiliza ampliamente para el análisis de la composición de diversos productos, incluyendo los alimenticios [62]. La espectroscopia NIR recoge la información de las estructuras químicas, a través de la valoración de los enlaces moleculares reflejados en el espectro obtenido (p. ej. C-H, N-H and O-H, que son los principales componentes estructurales de las moléculas orgánicas). La composición química de la muestra origina diferencias que quedan reflejadas en las vibraciones armónicas y de combinación, construyendo de esta manera un espectro característico que sirve como huella dactilar [63-66]. La relación entre la huella dactilar y la composición química de la muestra puede parecer poco evidente pero el hecho es que, aplicando el mismo procedimiento experimental, el espectro obtenido puede considerarse como una característica determinada únicamente por los componentes químicos presentes en la planta medicinal y en estas circunstancias es aplicable el concepto de fitoequivalencia, ya que este no exige el aislamiento e identificación de los componentes de la muestra [16,28].

Existen tres tipos de configuración instrumental para el registro de espectros NIR: transmitancia, reflectancia y transflectancia; en la Figura 1.1 se muestra una representación esquemática. Considérese que el haz de luz que emite la fuente de radiación ya ha atravesado el correspondiente sistema de selección de longitudes de onda. En las medidas por reflectancia (sólidos y semisólidos), el haz de luz es reflejado por la propia muestra y llega al detector. En modo transmitancia (gases, líquidos, semilíquidos y sólidos), el haz de luz atraviesa la muestra desde la fuente de radiación hasta el detector. Finalmente, un caso intermedio lo ocupa las medidas por

transflectancia (líquidos turbios y semilíquidos), donde el haz de luz atraviesa la muestra y se refleja en un reflector que está en contacto con la misma, hasta llegar de nuevo al detector. Para los trabajos descritos en la presente memoria se empleó el modo de reflectancia.



**Figura 1.1.** Esquema de las configuraciones para el registro de espectros NIR.

La espectroscopia NIR presenta las ventajas de ser un método rápido y no destructivo, que requiere una mínima (o ninguna) preparación de muestra; así mismo, hace posible la determinación simultánea del contenido químico y de algunas propiedades físicas de la muestra [67]. Como resultado, en los últimos años, ha habido una tendencia creciente en el uso de la espectroscopia NIR para el análisis cualitativo y cuantitativo de las plantas medicinales. Por mencionar algunos ejemplos, empleando el método NIR se ha determinado la autenticidad de *S. miltiorrhiza*, se han identificado las diferentes variedades geográficas de *Paeonia lactiflora*, *Phellodendri chinensis* y *P. amurensis*; así mismo, los algunos compuestos representativos de estas plantas se han cuantificado con PLS (partial least squares) [49]. A pesar de las ventajas en el empleo de los perfiles espectroscópicos, los cromatográficos siguen siendo los más utilizados para evaluar la calidad de las plantas medicinales [5]; probablemente debido a que la técnica no es muy sensible, así que por lo general sólo puede aplicarse a los componentes mayoritarios [68].

#### **1.5.4. Cromatografía de líquidos de alta resolución (CLAR).**

La CLAR es un método popular empleado en el análisis de plantas medicinales debido a que es una técnica relativamente sencilla y muy versátil, ya que permite analizar compuestos tanto volátiles como no volátiles, polares, no polares y hasta iónicos [14,69]. En general, la CLAR puede ser empleada para analizar la mayoría de los compuestos en las plantas medicinales, tal y como lo revelan las publicaciones a este respecto. Entre los métodos cromatográficos para la toma de huellas dactilares de las plantas medicinales, el más popular sigue siendo la cromatografía líquida de alta resolución acoplada a un detector de diodos (CLAR-DAD) [5,34,49]. El uso de detectores de diodos en línea proporciona información adicional a los perfiles cromatográficos obtenidos a longitudes de onda simples, ya que se pueden obtener además los espectros relacionados con cada uno de los picos, lo cual puede ser empleado como criterio de identificación y análisis de pureza de cada uno de los picos [15,20,70].

Otros acoplamientos, tales como: CLAR-EM, CLAR-RMN, también han sido utilizados con estos mismos propósitos; en particular, los acoplamientos CLAR-EM se han convertido en una herramienta ampliamente utilizada en el análisis de productos fitoterapéuticos ya que proporcionan también información espectral adicional, que es muy útil para el análisis cualitativo y para la elucidación estructural. Aunque tradicionalmente este tipo de acoplamientos han sido utilizados para la identificación de compuestos específicos, en la actualidad existe un interés creciente en su aplicación a la obtención de la huella dactilar del producto completo [20,21,71-73].

##### **1.5.4.1. Factores que afectan el establecimiento del perfil cromatográfico.**

La huella dactilar cromatográfica de una planta es una característica obtenida por un procedimiento definido, separando tantos compuestos como sea posible para construir un patrón de reconocimiento específico. La identidad de estos componentes químicos puede ser conocida o desconocida [14,34]. Puesto que todo el conjunto de compuestos caracteriza la composición química de la planta medicinal, la huella digital



cromatográfica representa una metodología cualitativa razonable, en la cual todo el cromatograma es evaluado durante el análisis de datos para discriminar entre diferentes especies de la misma familia de plantas [5,34,49]. Obtener un buen perfil cromatográfico que represente la fitoequivalencia de una planta medicinal depende de varios factores, tales como el procedimiento de extracción, el instrumento de medición y las condiciones de separación, entre otros [14,74-77]. La mayoría de los procesos de extracción para plantas medicinales, emplean mezclas hidroalcohólicas como solvente de extracción y los solventes más comúnmente utilizados son agua, metanol y etanol [2,5,16,49,78].

En cuanto a los aspectos instrumentales, se debe considerar que la huella dactilar puede ser afectada por factores como la degradación de la fase estacionaria; por ejemplo, debido a la baja estabilidad del sílice y de los soportes a base de sílice a valores altos de pH y también se debe contemplar la posibilidad del colapso de la fase unida C18 a causa de una fase móvil altamente polar (dewetting). Los pequeños cambios en la composición de la fase móvil también afectan a los perfiles obtenidos; causados por fluctuaciones de temperatura y presión, variaciones en la velocidad de flujo y en la dispersión del gradiente. Algunas variaciones pueden tener origen en el detector, por ejemplo, un desplazamiento de la longitud de onda en un espectrómetro UV-Visible o un mal alineamiento del monocromador. Hay que tener cuidado de no sobrecargar la columna, ya que algunos componentes pueden presentar una alta concentración. Finalmente, los perfiles cromatográficos se pueden ver afectados por la interacción entre analitos; este último factor, junto con los cambios instrumentales desconocidos, produce una gran incertidumbre en el establecimiento de los perfiles [5,38,79].

Una vez optimizadas las condiciones de tratamiento de muestra y separación, el siguiente reto es el análisis de los datos obtenidos. Como ya se ha mencionado es posible emplear las técnicas cromatográficas para obtener perfiles relativamente completos de las plantas medicinales que permitan establecer la fitoequivalencia. Sin embargo, la interpretación de estos perfiles se complica cuando el número de componentes es grande y más aún cuando se presentan señales no resueltas. La

quimiometría permite afrontar las dificultades en el análisis de calidad de las plantas medicinales. El empleo de técnicas quimiométricas facilita el análisis de datos, tanto cromatográficos como espectrales.

### **1.6. Quimiometría.**

Como se ha estado mencionando anteriormente, se pueden emplear muchas técnicas cromatográficas y espectroscópicas, incluyendo técnicas con acoplamientos, para el análisis de los productos fitoterapéuticos y para construir los perfiles típicos de éstos. El problema aquí es cómo se evalúa la información obtenida con estas técnicas.

La quimiometría es un campo interdisciplinario que involucra el análisis multivariable, los modelos matemáticos, la informática y la química analítica. Algunas de las áreas de aplicación más importantes de la quimiometría incluyen (1) la calibración, validación y pruebas de significación, (2) la optimización de las mediciones químicas y procedimientos experimentales y (3) la extracción del máximo de información química a partir de datos analíticos [80]. El perfil cromatográfico o espectroscópico de una planta medicinal es un sistema multivariable, tanto por la manera en que los instrumentos actuales realizan el registro de las señales analíticas, como porque se involucra en el perfil a la mayoría de los componentes fitoquímicos de la planta medicinal. Con ayuda de la quimiometría se puede afrontar, con mayores posibilidades de éxito, las dificultades que surgen al emplear técnicas cromatográficas o espectroscópicas para el control de calidad de las plantas medicinales, como son las señales no resueltas, la gran cantidad de componentes en las plantas y la complejidad de los espectros del infrarrojo cercano.

La quimiometría abarca diversos objetivos como la aplicación de pretratamientos a los datos experimentales para mejorar la calidad de la señal, la construcción de modelos para el reconocimiento de pautas y la determinación cuantitativa [81]. Son muchas las técnicas quimiométricas que se pueden utilizar para llevar a cabo estos objetivos; en esta sección se describen aquéllas que han sido utilizadas en esta memoria.

### **1.6.1. Pretratamientos de los datos obtenidos con los métodos instrumentales.**

De manera general, el pretratamiento es la modificación de datos realizada antes de construir un modelo matemático o antes de aplicar otra forma de analizar los datos. El propósito del pretratamiento es hacer lineal la respuesta de las variables y eliminar las fuentes extrañas de variación que no son de interés en el análisis. De no ser eliminada la varianza debida a los errores sistemáticos presentes en los experimentos, el modelo requerirá un trabajo más complejo para obtener la información de interés [78,82].

#### **1.6.1.1. Variable aleatoria normal tipificada (SNV, Standard Normal Variate).**

SNV es una transformación que se aplica habitualmente a los datos espectroscópicos para minimizar los efectos de la dispersión de la luz. Utiliza el centrado y escalado de cada espectro individual (es decir, estandariza cada espectro manipulando únicamente los datos del mismo). El resultado práctico del SNV es que minimiza las interferencias multiplicativas de la dispersión en los datos espectrales producidas por los distintos tamaños de partícula en la muestra; un efecto del SNV es que, en la escala vertical, cada espectro está centrado en cero [83].

#### **1.6.1.2. Derivadas.**

Las derivadas permiten aumentar las diferencias entre las bandas anchas y solapadas de los espectros; también se utilizan para corregir los efectos de línea base. La primera derivada elimina los desplazamientos de línea base paralelos al eje de las abscisas, mientras que la segunda derivada elimina los términos que varían linealmente con la longitud de onda. Los métodos de derivación más utilizados son el de Norris [84] y el de Savitzky- Golay [85], el cual incluye un suavizado simultáneo con la derivación.

### **1.6.1.3. Normalización.**

En muchos métodos analíticos, las variables medidas para una muestra dada son incrementadas o disminuidas de su valor auténtico por un factor multiplicativo. Los métodos de normalización intentan corregir este tipo de efectos mediante la identificación de algún aspecto de cada muestra que debería ser esencialmente constante de una muestra a otra, y realizar la corrección de la escala de todas las variables utilizando dicha característica. Cuando se construyen modelos de análisis discriminante tales como PLS-DA (partial least squares - discriminat analysis) o SIMCA (soft independent modeling of class analogy), la normalización se efectúa si la relación entre las variables y no la magnitud absoluta de la respuesta es el aspecto más importante de los datos para la identificación de una especie; por ejemplo, la concentración de un compuesto no es tan relevante, sólo el hecho de que se encuentre en una cantidad detectable. El uso de la normalización en estas condiciones debería ser considerado después de evaluar cómo son los cambios en la respuesta de las variables para las diferentes clases de interés [82,86].

### **1.6.1.4. Escalado (Scaling).**

El escalado de los datos de una matriz entre un valor mínimo y uno máximo, es un caso particular de normalización que puede ser aplicado antes de la construcción de los modelos matemáticos. Este pretratamiento puede ser útil para evitar la presencia de valores extremos en la escala de los datos en algunas muestras de origen natural y se prefiere su uso cuando se trata de aplicaciones cuantitativas [69].

### **1.6.1.5. Corrección de la línea base (Baseline correction).**

Existen diferentes maneras de hacer la corrección de la línea base; en la presente memoria se ha utilizado el método de mínimos cuadrados ponderados (WLS, Weighted Least Squares). Este método es comúnmente empleado en aplicaciones espectroscópicas (o cromatográficas) donde la señal de algunas variables es debida solamente a la señal de fondo. Estas variables sirven como referencia para determinar

cuanta señal de fondo debe ser eliminada de las variables cercanas. El algoritmo WLS emplea un enfoque automático para determinar qué puntos son los más probables para ser sólo línea base; esto lo hace ajustando de manera iterativa un línea base a cada espectro (o cromatograma) y determinando qué variables están claramente por arriba de la línea base (es decir, la señal) y cuáles están debajo de ésta. Se asume que los puntos debajo de la línea base son más significativos en el ajuste de la señal de fondo del espectro.

Este método también es llamado "asymmetric weighted least squares". El efecto práctico de su uso es una eliminación automática de la señal de fondo, evitando la creación de picos intensamente negativos [79,82].

#### **1.6.1.6. Alineación optimizada mediante correlación (COW, Correlation Optimized Warping).**

COW es un método para alinear las señales de datos de muestras que exhiben cambios en su posición a lo largo del eje  $x$ , empleando un vector de referencia para alinear el vector de la muestra. El vector usado como referencia debe ser representativo de las muestras que se quieren alinear. El método funciona mediante la división del vector de la muestra a ser alineada en segmentos de tamaño definido y permitiendo el aumento o la disminución de la longitud de estos segmentos para buscar la correlación óptima con los segmentos del vector de referencia [87].

Así, los parámetros que hay que definir para la aplicación de este método de alineación son: el vector a ser usado como referencia, la longitud del segmento y el grado de flexibilidad (slack); la selección adecuada de estos parámetros puede ser realizada aplicando algoritmos como los descritos por Skov [88] y Tomasi [89]. COW se aplica típicamente en datos cromatográficos donde la posición de los picos puede cambiar entre análisis, debido a pequeños cambios en las muestras o en las condiciones del cromatógrafo.

**1.6.1.7. Centrado (Mean centering).**

Este pretratamiento calcula el valor medio de cada columna de la matriz de datos y resta este valor de la columna, trasladando los ejes del sistema de coordenadas hacia el centroide de los datos y hace que cada muestra exhiba solo las diferencias que tiene con respecto a la muestra promedio de los datos originales [81].

**1.6.2. Métodos de reconocimiento de pautas.**

Aquí se mencionan las técnicas multivariadas utilizadas en esta memoria mediante las que se puede establecer agrupaciones de muestras en función de su similitud y métodos de clasificación de nuevas muestras.

**1.6.2.1. Análisis en componentes principales (PCA, Principal Component Analysis).**

El PCA es un método matemático bilineal que provee una interpretación general de la principal información contenida en una tabla multivariable, mediante la extracción y visualización de la variación sistemática de los datos en dicha tabla [35]. Los datos originales se transforman en una combinación lineal de variables no correlacionadas entre sí, llamadas componentes principales (PCs). No todos los componentes principales contienen la misma información; los primeros son los que describen la mayor variación en los datos, que se asocia a la información más relevante, mientras que los últimos describen variaciones en los datos que pueden ser debidas a ruido o error experimental, o a un sobreajuste del modelo y pueden ser descartados, con lo que se consigue una importante reducción del número de variables. Se aplica la siguiente ecuación:

$$X = \sum_{i=1}^a t_i p_i^T + E \quad (1.1)$$

donde  $X$  es la matriz de datos a tratar (en esta tesis serán huellas dactilares, datos espectrales o cromatográficos después de su pretratamiento),  $a$  es el número de

componentes principales que contienen la información deseada,  $t_i$  son los scores para cada componente principal y  $p_i$  los loadings y  $E$  es una matriz de error, es decir la variación residual de  $X$  que no es explicada por el modelo con  $a$  componentes principales. El superíndice  $T$  indica la matriz traspuesta. La ecuación 1.1 se puede expresar de la siguiente manera:

$$X = TP^T + E \quad (1.2)$$

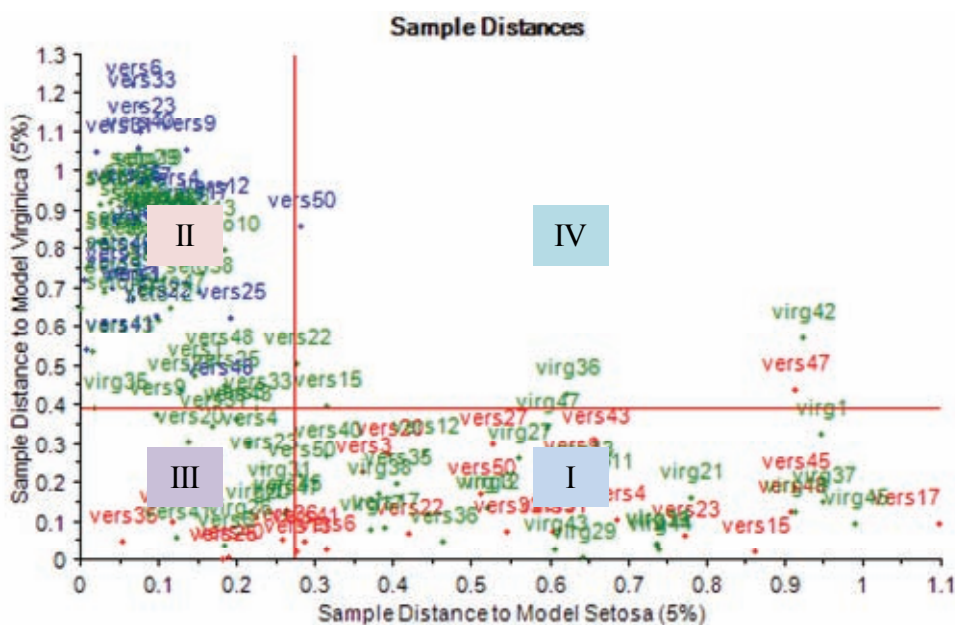
donde  $T$  es la matriz de scores, con tantas filas como la matriz original  $X$ , y que contiene la información sobre las filas (los objetos) de la matriz original;  $P^T$  corresponde a la traspuesta de la matriz de loadings, con tantas columnas como los datos originales, y que contiene información sobre las columnas (las variables) de la matriz original. El número de columnas de la matriz de scores corresponde al número calculado de componentes principales,  $a$ ; PC1 explica la máxima variabilidad de la matriz  $X$ , PC2 es ortogonal a la primera componente y abarca la mayor cantidad de la variación restante, y así sucesivamente, hasta que se explica el total de la variación útil de la matriz original  $X$ . Al representar gráficamente los valores de los scores, es posible detectar e interpretar las pautas presentadas por las muestras, agrupaciones, similitudes, diferencias y muestras anómalas.

#### **1.6.2.2. Modelado blando independiente de analogías de clases (SIMCA, Soft Independent Modeling of Class Analogy).**

SIMCA es un método de clasificación supervisado donde cada clase se modela usando un PCA de forma independiente, de forma que cada clase tiene un modelo específico que la describe con el número óptimo de PCs [80]. De esta manera, se puede construir un “espacio de la clase”, cuya volumen marca el límite entre las muestras que son descritas por el modelo como pertenecientes a la clase estudiada y las muestras que no pueden ser consideradas como pertenecientes a ella, para un nivel de confianza seleccionado [17]. Para saber si una nueva muestra se ajusta a la clase, la información que se calcula es dividida en dos, una parte explicada por el modelo de la clase y otra que permanece en los residuales. Si los residuales de la

muestra son significativamente mayores que los de la clase, la muestra es rechazada [74].

Dependiendo del programa utilizado en la construcción del modelo SIMCA, se pueden tener diferentes maneras de representar los resultados en la clasificación de las muestras desconocidas; por ejemplo, pueden ser visualizados empleando un gráfico de Cooman, que muestra la discriminación de dos clases a la vez (véase la Figura 1.2). La distancia desde el modelo para la clase 1 es representada frente a la distancia del modelo 2, las distancias críticas son indicadas en ambos ejes. Por lo tanto, se definen cuatro zonas en el gráfico: clase 1 (I), clase 2 (II), una zona de confusión en la cual se superponen las clases 1 y 2 (III), y la zona de valores atípicos (IV). Al representar las muestras del conjunto de validación en este gráfico, es más fácil evaluar qué tan certera es la clasificación.



**Figura 1.2.** Ejemplo de un gráfico de Cooman; las líneas rojas indican las distancias críticas. (Imagen extraída del software The Unscrambler X).

Otra forma de interpretar los resultados obtenidos, es determinar el espacio definido por una clase dada ( $T^2$ ) y su espacio residual ( $Q$ ), con la finalidad de evaluar si los valores de cada muestra predicha están mejor descritos por el espacio  $T^2$ ; de ser así, las muestras son clasificadas como miembros de la clase.



### **1.6.2.3. Análisis discriminante (DA, Discriminat Analysis).**

El análisis discriminante es una técnica supervisada de clasificación, es decir, donde se han definido previamente el número de clases y de muestras que pertenecen a cada una de ellas [63]. Cuanto mayores son las diferencias entre dos clases dadas, mayor es la distancia de Mahalanobis entre ellas. La discriminación de las muestras se consigue mediante el cálculo de la distancia de Mahalanobis de cada una de ellas a los centros de los grupos considerados [95]. Una muestra desconocida se clasifica como perteneciente al grupo con el cual tenga una distancia más cercana al centro [80]. El DA puede considerarse similar al PCA sólo en el sentido de que ambos determinan un hiperplano con un menor número de variables en el cual se proyectan los datos de las muestras desde un plano con más variables. Sin embargo, el PCA selecciona la dirección que retiene la máxima estructura entre los datos, mientras que el DA selecciona la dirección en la que se consigue una separación máxima entre los grupos definidos [96]. En la construcción de este modelo se debe tener en consideración que se requiere un mayor número de muestras que de variables.

### **1.6.3. Métodos de regresión.**

A continuación se mencionan las técnicas multivariadas empleadas en esta memoria para buscar una relación entre la señal analítica y alguna propiedad de la muestra.

#### **1.6.3.1 Regresión por mínimos cuadrados parciales (PLSR, Partial Least Squares Regression).**

La regresión parcial por mínimos cuadrados (PLSR) se utiliza para encontrar las relaciones fundamentales entre las variables independientes (X) y las variables dependientes (Y), las cuales son modeladas simultáneamente teniendo en cuenta no sólo la varianza de X, sino la covarianza entre X e Y [90]. Entonces, X e Y se descomponen simultáneamente en un producto de otras dos matrices de scores y loadings; como es descrito por las siguientes ecuaciones:

$$X = TP^T + E \quad (1.3)$$

$$Y = UQ^T + F \quad (1.4)$$

donde  $TP^T$  se aproxima a los datos de la matriz  $X$  y  $UQ^T$  a los verdaderos valores de  $Y$ . La descomposición no es independiente y existe una relación lineal entre los scores  $T$  y  $U$ . Los términos  $E$  y  $F$  de las ecuaciones 1.3 y 1.4 son matrices de error y el superíndice  $T$  significa la transpuesta de la matriz. El algoritmo PLS trata de encontrar factores (llamados variables latentes) que maximizan la cantidad de variación explicada en  $X$  que es relevante para la predicción de  $Y$ , es decir, capturar varianza y conseguir correlación [91].

#### **1.6.3.2. Análisis discriminante con mínimos cuadrados parciales (PLS-DA, Partial Least Squares - Discriminat Analysis).**

Es un método supervisado de análisis que se basa en la regresión en mínimos cuadrados parciales [96]. Las variables independientes ( $X$ ) pueden ser el espectro o el cromatograma de cada muestra, mientras que la variable dependiente  $Y$  es una variable categórica, definida por el analista, que codifica cada clase de manera numérica [81]. Habitualmente, la matriz  $Y$  consta de números enteros [97].

El resultado de la regresión son valores de  $Y$  cercanos a los que indican cada categoría. Una muestra del conjunto de predicción se considera que está correctamente categorizada si cumple dos criterios: debe tener un perfil cromatográfico o espectral que no sea significativamente diferente de los perfiles de las muestras que forman el conjunto de calibración y el valor calculado de la variable categórica debe estar dentro de un intervalo de valores alrededor del número con el que se codifica la categoría (cut - off value).

#### 1.6.4. Resolución multivariable de curvas - mínimos cuadrados alternados (MCR-ALS, Multivariate Curve Resolution - Alternating Least Squares).

El objetivo del MCR es la descomposición bilineal de una respuesta instrumental compleja en las contribuciones vinculadas a cada uno de los  $n$  componentes puros en el sistema bajo estudio [80,92], como se describe con la siguiente ecuación:

$$D = CS^T + E \quad (1.5)$$

En el caso de la espectroscopia NIR, por ejemplo, las filas de la matriz  $D$  son los espectros de las mezclas que se midieron durante el experimento, los perfiles de las columnas de la matriz  $C$  y los perfiles de las filas de la matriz  $S^T$  están asociados con las contribuciones y los espectros puros de los componentes resueltos, respectivamente. El superíndice  $T$  significa la transpuesta de la matriz  $S$ , donde los espectros puros son los perfiles contenidos en las columnas.  $E$  es una matriz de error, es decir, la variación residual del conjunto de datos que no se explica por el producto  $CS^T$  y por lo tanto no debería estar relacionada con ninguna contribución química.

La ecuación 1.5 puede resolverse de forma iterativa utilizando ALS, donde las matrices  $C$  y  $S^T$  son refinadas, tanto como sea posible, para producir perfiles que se ajusten a los datos experimentales de la matriz  $D$ . El proceso de optimización se inicia empleando un número propuesto de factores y con las estimaciones iniciales de  $C$  o  $S^T$ ; estos parámetros de entrada pueden ser obtenidos con el conocimiento previo de los componentes implicados o mediante el uso de otros métodos matemáticos tales como PCA (Principal Component Analysis) o EFA (Evolving Factor Analysis) [93,94].

Durante la optimización con ALS, se puede reducir el número de posibles soluciones de la ecuación 1.5 aplicando una serie de limitaciones (constraints) tales como la no negatividad (todos las contribuciones o espectros tienen valores positivos), unimodalidad (permite la presencia de sólo un máximo por perfil) y closure (con esta limitación, la suma de las concentraciones de todas las especies

consideradas se ve obligada a ser igual a un valor constante, es decir, a la concentración total). La convergencia se logra cuando en dos ciclos de iteraciones consecutivas, la diferencia relativa en la desviación estándar de los residuales que hay entre los datos experimentales y los obtenidos con ALS son menores a un valor previamente seleccionado.

#### **1.6.5. Verificación de los modelos y selección del número de factores.**

La verificación (también llamada validación) de un modelo basado en datos empíricos significa revisar qué tan bien se desempeñará cuando sea utilizado en la predicción de nuevas muestras, del mismo tipo que las utilizadas en su construcción; es decir, estima la incertidumbre de las futuras predicciones. Si la incertidumbre es razonablemente baja, el modelo puede ser considerado válido [80,81].

La verificación de la bondad del modelo se puede ensayar sobre un nuevo conjunto de muestras, distintas de las utilizadas en la construcción del mismo (llamado conjunto de prueba, test set) o mediante un sistema de validación cruzada (cross-validation). Con el método de la validación cruzada, las mismas muestras se utilizan tanto para la estimación del modelo, como para su evaluación. Unas cuantas muestras quedan fuera del conjunto de calibración y se construye un modelo con las muestras restantes. Entonces, son predichas las muestras que han quedado fuera de la calibración y se calculan los residuales de predicción. El proceso se repite con otro subconjunto de muestras del conjunto de calibración, y así sucesivamente hasta que todos los objetos se ha dejado fuera una vez [80]. A partir de la validación cruzada se pueden calcular figuras de mérito para evaluar cómo se comporta el modelo cuando se aplica a nuevos datos y también se puede determinar el número óptimo de factores a incluir (componentes principales para un PCA o variables latentes en un PLS) [82]. Los nombres de las figuras de mérito obtenidas, en la validación cruzada y en la predicción de un conjunto de muestras independiente, varían entre los diferentes programas de quimiometría; en este trabajo, se han utilizado los siguientes:

- Raíz del error cuadrático medio de calibración (RMSEC, root mean square error of calibration), nos proporciona información sobre el ajuste del modelo a los datos de calibración; es una medida de lo bien que se ajusta el modelo a los datos.
- Raíz del error cuadrático medio de la validación cruzada (RMSECV, root mean square error of cross-validation), que es una medida de la capacidad de un modelo para predecir las muestras que no fueron utilizadas para construirlo durante el procedimiento de validación cruzada.
- Raíz del error cuadrático medio de predicción (RMSEP, root mean square error of prediction), nos da información sobre el ajuste del modelo a los datos de predicción. Son utilizadas las muestras de validación preparadas y medidas independientemente de las muestras de calibración; es decir, que no fueron empleadas en la construcción del modelo o en la validación cruzada.

El parámetro PRESS (Prediction Sum of Squares) es la suma del cuadrado de los residuales utilizando un modelo construido con un determinado número de componentes principales o variables latentes. El PRESS se representa gráficamente en función del número de factores y sirve como ayuda para la selección del número óptimo de componentes en el modelo. El error del modelo debería decrecer en la medida en que se incluyen factores que describen grandes cantidades de varianza sistemática, y debería aumentar al incluir factores que describen ruido; es decir, que sólo describen una pequeña cantidad de varianza. Se debe seleccionar el número de factores donde el error del modelo sea mínimo [81,82].

Hay otras maneras de seleccionar el número adecuado de factores para un modelo. Algunos programas proporcionan las curvas de varianza explicada (o las curvas de varianza residual), tanto para la calibración como para la predicción obtenida por la validación cruzada; el usuario debe comparar dichas curvas para elegir el número de factores.



**Referencias.**

- (1) Xiaorui, Z. General Guidelines for Methodologies on Research and Evaluation of Traditional Medicine. 2000. Geneva. WHO/EDM/TRM/2000.1.
- (2) Drašar, P.; Moravcova, J. Recent advances in analysis of Chinese medical plants and traditional medicines. *Journal of Chromatography B* 2004, *812*, 3-21.
- (3) WHO. WHO guidelines on good manufacturing practices (GMP) for herbal medicines. 2007. Geneva. ISBN: 978 92 4 154716 1.
- (4) WHO. National policy on traditional medicine and regulation of herbal medicines: Report of a WHO global survey. 2005. France. ISBN: 92 4 159323 7.
- (5) Liang, Y.; Xie, P.; Chan, K. Quality control of herbal medicines. *Journal of Chromatography B* 2004, *812*, 53-70.
- (6) Secretaria de Salud; Comisión Permanente de la Farmacopea de los Estados Unidos Mexicanos. *Farmacopea Herbolaria de los Estados Unidos Mexicanos*. 2001. México. ISBN: 9789688119822.
- (7) Li, B.; Hu, Y.; Liang, Y.; Xie, P.; Du, Y. Quality evaluation of fingerprints of herbal medicine with chromatographic data. *Anal. Chim. Acta* 2004, *514*, 69-77.
- (8) Mahady, G. B. Global Harmonization of Herbal Health Claims. *The Journal of Nutrition* 2001, *131*, 1120S-1123S.
- (9) Perez-Meseguer, J.; Garza-Juarez, A.; Salazar-Aranda, R.; Salazar-Cavazos, M. L.; de la Torre Rodriguez, Y. C.; Rivas-Galindo, V.; Waksman de Torres, N. Development and validation of an HPLC-DAD analytical procedure for quality control of damiana (*Turnera diffusa*), using an antioxidant marker isolated from the plant. *J. AOAC Int.* 2010, *93*, 1161-1168.
- (10) European Pharmacopoeia. European Pharmacopoeia 5.0, Volumes 1 and 2. The European Pharmacopoeia Commission. 2005. Strasbourg: EDQM.
- (11) FDA. Dietary Supplements. <http://www.fda.gov/food/dietarysupplements/default.htm> (con acceso el 21/06, 2012).
- (12) FDA. *Guidance for Industry, Botanical Drug Products*. Rockville, MD. CDER/FDA 2004.
- (13) Yap, K. Y.; Chan, S. Y.; Lim, C. S. Infrared-based protocol for the identification and categorization of ginseng and its products. *Food Res. Int.* 2007, *40*, 643-652.
- (14) Xie, P.; Chen, S.; Liang, Y.; Wang, X.; Tian, R.; Upton, R. Chromatographic fingerprint analysis—a rational approach for quality assessment of traditional Chinese herbal medicine. *Journal of Chromatography A* 2006, *1112*, 171-180.
- (15) Lau, A.; Seo, B.; Woo, S.; Koh, H. High-performance liquid chromatographic method with quantitative comparisons of whole chromatograms of raw and steamed *Panax notoginseng*. *Journal of Chromatography A* 2004, *1057*, 141-149.
- (16) Mok, D. K. W.; Chau, F. Chemical information of Chinese medicines: A challenge to chemist. *Chemometrics Intellig. Lab. Syst.* 2006, *82*, 210-217.
- (17) Gallardo-Velázquez, T.; Osorio-Revilla, G.; Loa, M. Z.; Rivera-Espinoza, Y. Application of FTIR-HATR spectroscopy and multivariate analysis to the quantification of adulterants in Mexican honeys. *Food Res. Int.* 2009, *42*, 313-318.
- (18) Laasonen, M.; Harmia-Pulkkinen, T.; Simard, C. L.; Michiels, E.; Rasanen, M.; Vuorela, H. Fast identification of *Echinacea purpurea* dried roots using near-infrared spectroscopy. *Anal. Chem.* 2002, *74*, 2493-2499.
- (19) Dumarey, M.; van Nederkassel, A. M.; Deconinck, E.; Vander Heyden, Y. Exploration of linear multivariate calibration techniques to predict the total antioxidant capacity of green tea from chromatographic fingerprints. *Journal of Chromatography A* 2008, *1192*, 81-88.
- (20) Yang, L.; Wu, D.; Tang, X.; Peng, W.; Wang, X.; Ma, Y.; Su, W. Fingerprint quality control of Tianjihuang by high-performance liquid chromatography–photodiode array detection. *Journal of Chromatography A* 2005, *1070*, 35-42.

- (21) Ji, Y.; Xu, Q.; Hu, Y.; Heyden, Y. V. Development, optimization and validation of a fingerprint of Ginkgo biloba extracts by high-performance liquid chromatography. *Journal of Chromatography A* 2005, *1066*, 97-104.
- (22) Sun, Y.; Guo, T.; Sui, Y.; Li, F. Fingerprint analysis of Flos Carthami by capillary electrophoresis. *Journal of Chromatography B* 2003, *792*, 147-152.
- (23) Wang, L.,C.; Cao, Y.,H.; Xing, X.,P.; Ye, J.,N. Fingerprint Studies of *Radix Scutellariae* by Capillary Electrophoresis and High Performance Liquid Chromatography. *Chromatographia* 2005, 283-288.
- (24) Zhang, M. H.; Xu, Q. S.; Massart, D. L. Robust principal components regression based on principal sensitivity vectors. *Chemometrics Intellig. Lab. Syst.* 2003, *67*, 175-185.
- (25) EMEA Guideline on quality of herbal medicinal products/traditional herbal medicinal products. *CPMP/QWP/2819/00*. 2006, Rev.1.
- (26) EMEA Guideline on specifications: Test procedures and acceptance criteria for herbal substances, herbal preparations and herbal medicinal products/traditional herbal medicinal products. *CPMP/QWP/2820/00*. 2006, Rev.1.
- (27) Sumner, L. W.; Mendes, P.; Dixon, R. A. Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry* 2003, *62*, 817-836.
- (28) Tyler, V. E. Phytomedicines: Back to the Future. *J. Nat. Prod.* 1999, *62*, 1589-1592.
- (29) Chen, P.; Luthriau, D.; Harringtonohio, P. d. B.; Harnly, J. M. Discrimination Among *Panax* Species Using Spectral Fingerprinting. *Journal of AOAC International* 2011, *94*, 1411-1421.
- (30) Garza-Juarez, A.; de la Luz Salazar-Cavazos, M.; Salazar-Aranda, R.; Perez-Meseguer, J.; Waksman de Torres, N. Correlation between Chromatographic Fingerprint and Antioxidant Activity of *Turnera diffusa* (Damiana). *Planta Med.* 2011, *77*, 958-963.
- (31) Peng, L.; Wang, Y.; Zhu, H.; Chen, Q. Fingerprint profile of active components for *Artemisia selengensis* Turcz by HPLC–PAD combined with chemometrics. *Food Chem.* 2011, *125*, 1064-1071.
- (32) Cserháti, T. *Multivariate Methods in Chromatography*; John Wiley & Sons, Ltd. 2008. ISBN: 9780470754146.
- (33) Garza-Juarez, A.; Waksman-De-Torres, N.; Ramirez-Duron, R.; Salazar Cavazos, M. Development and validation of fingerprints of *Turnera diffusa* extracts obtained by use of high-performance liquid chromatography with diode array detection and chemometric methods. *Acta Chromatographica* 2009, *21*, 217-235.
- (34) Dejaegher, B.; Alaerts, G.; Matthijs, N. Methodology to develop liquid chromatographic fingerprints for the quality control of herbal medicines. *Acta Chromatographica* 2010, *22*, 237-258.
- (35) Chen, Y.; Xie, M.; Yan, Y.; Zhu, S.; Nie, S.; Li, C.; Wang, Y.; Gong, X. Discrimination of *Ganoderma lucidum* according to geographical origin with near infrared diffuse reflectance spectroscopy and pattern recognition techniques. *Anal. Chim. Acta* 2008, *618*, 121-130.
- (36) Jin, Z.; Xu, S.; Sun, S.; Zhou, Q. Analysis of *Acanthopanax senticosus* Harms for different parts using Fourier transform infrared spectroscopy. *Guangpuxue Yu Guangpu Fenxi* 2008, *28*, 2859-2863.
- (37) Gong, F.; Liang, Y.; Xie, P.; Chau, F. Information theory applied to chromatographic fingerprint of herbal medicine for quality control. *Journal of Chromatography A* 2003, *1002*, 25-40.
- (38) Gong, F.; Liang, Y.; Fung, Y.; Chau, F. Correction of retention time shifts for chromatographic fingerprints of herbal medicines. *Journal of Chromatography A* 2004, *1029*, 173-183.
- (39) Circosta, C.; De Pasquale, R.; Samperi, S.; Pino, A.; Occhiuto, F. Biological and analytical characterization of two extracts from *Valeriana officinalis*. *J. Ethnopharmacol.* 2007, *112*, 361-367.
- (40) Ong, E. S. Extraction methods and chemical standardization of botanicals and herbal preparations. *Journal of Chromatography B* 2004, *812*, 23-33.

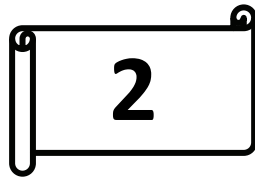


- (41) Ye, M.; Guo, D. Analysis of bufadienolides in the Chinese drug ChanSu by high-performance liquid chromatography with atmospheric pressure chemical ionization tandem mass spectrometry. *Rapid Communications in Mass Spectrometry* 2005, *19*, 1881-1892.
- (42) Belliardo, F.; Bicchi, C.; Cordero, C.; Liberto, E.; Rubiolo, P.; Sgorbini, B. Headspace-Solid-Phase Microextraction in the Analysis of the Volatile Fraction of Aromatic and Medicinal Plants. *Journal of Chromatographic Science* 2006, *44*, 416-429.
- (43) Deng, C.; Ji, J.; Li, N.; Yu, Y.; Duan, G.; Zhang, X. Fast determination of curcumol, curdione and germacrone in three species of Curcuma rhizomes by microwave-assisted extraction followed by headspace solid-phase microextraction and gas chromatography–mass spectrometry. *Journal of Chromatography A* 2006, *1117*, 115-120.
- (44) Lucero, M. E.; Fredrickson, E. L.; Estell, R. E.; Morrison, A. A.; Richman, D. B. Volatile Composition of *Gutierrezia sarothrae* (Broom Snakeweed) as Determined by Steam Distillation and Solid Phase Microextraction. *J. Essent. Oil Res.* 2006, *18*, 121-125.
- (45) Cao, J.; Qi, M.; Fang, L.; Zhou, S.; Fu, R.; Zhang, P. Solid-phase microextraction–gas chromatographic–mass spectrometric analysis of volatile compounds from Curcuma wenyujin Y.H. Chen et C. Ling. *J. Pharm. Biomed. Anal.* 2006, *40*, 552-558.
- (46) Krist, S.; Stuebiger, G.; Unterweger, H.; Bandion, F.; Buchbauer, G. Analysis of Volatile Compounds and Triglycerides of Seed Oils Extracted from Different Poppy Varieties (*Papaver somniferum* L.). *J. Agric. Food Chem.* 2005, *53*, 8310-8316.
- (47) Guo, F.; Huang, L.; Zhou, S.; Zhang, T.; Liang, Y. Comparison of the volatile compounds of *Atractylodes* medicinal plants by headspace solid-phase microextraction-gas chromatography–mass spectrometry. *Anal. Chim. Acta* 2006, *570*, 73-78.
- (48) Flamini, G.; Cioni, P. L.; Morelli, I. Composition of the essential oils and in vivo emission of volatiles of four *Lamium* species from Italy: *L. purpureum*, *L. hybridum*, *L. bifidum* and *L. amplexicaule*. *Food Chem.* 2005, *91*, 63-68.
- (49) Jiang, Y.; David, B.; Tu, P.; Barbin, Y. Recent analytical approaches in quality control of traditional Chinese medicines—A review. *Anal. Chim. Acta* 2010, *657*, 9-18.
- (50) Johnson, K. J.; Wright, B. W.; Jarman, K. H.; Synovec, R. E. High-speed peak matching algorithm for retention time alignment of gas chromatographic data for chemometric analysis. *Journal of Chromatography A* 2003, *996*, 141-155.
- (51) Guo, F.; Liang, Y.; Xu, C.; Huang, L.; Li, X. Comparison of the volatile constituents of *Artemisia capillaris* from different locations by gas chromatography–mass spectrometry and projection method. *Journal of Chromatography A* 2004, *1054*, 73-79.
- (52) van Nederkassel, A. M.; Xu, C. J.; Lancelin, P.; Sarraf, M.; MacKenzie, D. A.; Walton, N. J.; Bensaid, F.; Lees, M.; Martin, G. J.; Desmurs, J. R.; Massart, D. L.; Smeyers-Verbeke, J.; Vander Heyden, Y. Chemometric treatment of vanillin fingerprint chromatograms: Effect of different signal alignments on principal component analysis plots. *Journal of Chromatography A* 2006, *1120*, 291-298.
- (53) Suntornsuk, L. Capillary electrophoresis of phytochemical substances. *J. Pharm. Biomed. Anal.* 2002, *27*, 679-698.
- (54) Chen, J.; Zhu, H.; Chu, V. M.; Jang, Y. S.; Son, J. Y.; Kim, Y. H.; Son, C. G.; Seol, I. C.; Kang, J. S. Quality control of a herbal medicinal preparation using high-performance liquid chromatographic and capillary electrophoretic methods. *J. Pharm. Biomed. Anal.* 2011, *55*, 206-210.
- (55) Gotti, R. Capillary electrophoresis of phytochemical substances in herbal drugs and medicinal plants. *J. Pharm. Biomed. Anal.* 2011, *55*, 775-801.
- (56) Pauli, G. F.; Jaki, B. U.; Lankin, D. C. Quantitative <sup>1</sup>H NMR: Development and Potential of a Method for Natural Products Analysis. *J. Nat. Prod.* 2005, *68*, 133-149.
- (57) Eisenreich, W.; Bacher, A. Advances of high-resolution NMR techniques in the structural and metabolic analysis of plant biochemistry. *Phytochemistry* 2007, *68*, 2799-2815.

- (58) van Beek, T. A.; Montoro, P. Chemical analysis and quality control of Ginkgo biloba leaves, extracts, and phytopharmaceuticals. *Journal of Chromatography A* 2009, *1216*, 2002-2032.
- (59) Bilia, A. R.; Bergonzi, M. C.; Mazzi, G.; Vincieri, F. F. NMR spectroscopy: a useful tool for characterisation of plant extracts, the case of supercritical CO<sub>2</sub> arnica extract. *J. Pharm. Biomed. Anal.* 2002, *30*, 321-330.
- (60) Jaki, B.; Sticher, O.; Veit, M.; Frahlich, R.; Pauli, G. F. Evaluation of Glucoiberin Reference Material from Iberis amara by Spectroscopic Fingerprinting. *J. Nat. Prod.* 2002, *65*, 517-522.
- (61) Ward, J. L.; Harris, C.; Lewis, J.; Beale, M. H. Assessment of <sup>1</sup>H NMR spectroscopy and multivariate analysis as a technique for metabolite fingerprinting of Arabidopsis thaliana. *Phytochemistry* 2003, *62*, 949-957.
- (62) Dufour, É. In *Chapter 1 - Principles of Infrared Spectroscopy*; Infrared Spectroscopy for Food Quality Analysis and Control; Academic Press: San Diego, 2009; pp 1-27.
- (63) Cozzolino, D.; Fassio, A.; Restaino, E.; Fernandez, E.; La Manna, A. Verification of Silage Type Using Near-Infrared Spectroscopy Combined with Multivariate Analysis. *J. Agric. Food Chem.* 2008, *56*, 79-83.
- (64) Xie, L.; Ying, Y.; Ying, T. Classification of tomatoes with different genotypes by visible and short-wave near-infrared spectroscopy with least-squares support vector machines and other chemometrics. *J. Food Eng.* 2009, *94*, 34-39.
- (65) Ren, G.; Chen, F. Simultaneous Quantification of Ginsenosides in American Ginseng (Panax quinquefolium) Root Powder by Visible/Near-Infrared Reflectance Spectroscopy. *J. Agric. Food Chem.* 1999, *47*, 2771-2775.
- (66) Woo, Y.; Kim, H.; Cho, J. Identification of Herbal Medicines Using Pattern Recognition Techniques with Near-Infrared Reflectance Spectra. *Microchemical Journal* 1999, *63*, 61-70.
- (67) Lau, C.; Chan, C.; Chau, F.; Mok, D. K. Rapid analysis of Radix puerariae by near-infrared spectroscopy. *Journal of Chromatography A* 2009, *1216*, 2130-2135.
- (68) Blanco, M.; Villarroya, I. NIR spectroscopy: a rapid-response analytical tool. *TrAC Trends in Analytical Chemistry* 2002, *21*, 240-250.
- (69) Fan, X.; Cheng, Y.; Ye, Z.; Lin, R.; Qian, Z. Multiple chromatographic fingerprinting and its application to the quality control of herbal medicines. *Anal. Chim. Acta* 2006, *555*, 217-224.
- (70) Yan, S.; Xin, W.; Luo, G.; Wang, Y.; Cheng, Y. An approach to develop two-dimensional fingerprint for the quality control of Qingkailing injection by high-performance liquid chromatography with diode array detection. *Journal of Chromatography A* 2005, *1090*, 90-97.
- (71) Gómez-Ariza, J. L.; García-Barrera, T.; Lorenzo, F. Anthocyanins profile as fingerprint of wines using atmospheric pressure photoionisation coupled to quadrupole time-of-flight mass spectrometry. *Anal. Chim. Acta* 2006, *570*, 101-108.
- (72) Zhang, J.; Ke, Y.; Liu, H.; Yang, X. Fingerprint Analysis of Flos carthami and Safflower Yellow Pigments by High-Performance Liquid Chromatography. *Anal. Lett.* 2005, *38*, 981-995.
- (73) Steinmann, D.; Ganzera, M. Recent advances on HPLC/MS in medicinal plant analysis. *J. Pharm. Biomed. Anal.* 2011, *55*, 744-757.
- (74) Ni, Y.; Lai, Y.; Brandes, S.; Kokot, S. Multi-wavelength HPLC fingerprints from complex substances: An exploratory chemometrics study of the Cassia seed example. *Anal. Chim. Acta* 2009, *647*, 149-158.
- (75) Zheng, L.; Watson, D. G.; Johnston, B. F.; Clark, R. L.; Edrada-Ebel, R.; Elseheri, W. A chemometric study of chromatograms of tea extracts by correlation optimization warping in conjunction with PCA, support vector machines and random forest data modeling. *Anal. Chim. Acta* 2009, *642*, 257-265.

- (76) Navarrete, A.; Avula, B.; Choi, Y. W.; Khan, I. A. Chemical fingerprinting of valeriana species: simultaneous determination of valerenic acids, flavonoids, and phenylpropanoids using liquid chromatography with ultraviolet detection. *J. AOAC Int.* 2006, *89*, 8-15.
- (77) Peng, L.; Wang, Y.; Zhu, H.; Chen, Q. Fingerprint profile of active components for *Artemisia selengensis* Turcz by HPLC–PAD combined with chemometrics. *Food Chem.* 2011, *125*, 1064-1071.
- (78) Hendriks, M. M. W. B.; Cruz-Juarez, L.; Bont, D. D.; Hall, R. D. Preprocessing and exploratory analysis of chromatographic profiles of plant extracts. *Anal. Chim. Acta* 2005, *545*, 53-64.
- (79) Daszykowski, M.; Walczak, B. Use and abuse of chemometrics in chromatography. *TrAC Trends in Analytical Chemistry* 2006, *25*, 1081-1096.
- (80) P. Gemperline. Practical Guide to Chemometrics, Second Edition, CRC Press, Boca Raton FL, 2006.
- (81) Brereton, R. G. Applied Chemometrics for Scientist, John Wiley & Sons, Ltd, Great Britain, 2007.
- (82) Wise, B.; Gallagher, N.; Bro, R.; Shaver, J.; Windig, W.; Koch, R. Chemometrics Tutorial for PLS\_Toolbox and Solo. 2006. ISBN: 0-9761184-1-6.
- (83) Barnes, R. J.; Dhanoa, M. S.; Lister, S. J. Standard normal variate transformation and detrending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy* 1989, *43*, 772-777.
- (84) Norris, K. H.; Williams, P. C. Optimization of Mathematical Treatments of Raw Near-Infrared Signal in the Measurement of Protein in Hard Red Spring Wheat. I. Influence of Particle Size.. *Cereal Chemistry* 1984, *61*, 158-165.
- (85) Savitzky, A.; Golay, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* 1964, *36*, 1627-1639.
- (86) Rinnan, Å.; Nørgaard, L.; Berg, F. v. d.; Thygesen, J.; Bro, R.; Engelsen, S. B. In *Chapter 2 - Data Pre-processing; Infrared Spectroscopy for Food Quality Analysis and Control*; Academic Press: San Diego, 2009; pp 29-50.
- (87) Nielsen, N. V.; Carstensen, J. M.; Smedsgaard, J. Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping. *Journal of Chromatography A* 1998, *805*, 17-35.
- (88) Skov, T.; van den Berg, F.; Tomasi, G.; Bro, R. Automated alignment of chromatographic data. *Journal of Chemometrics*, 2006, *20*, 484-497.
- (89) Tomasi, G.; van den Berg, F.; Andersson, C. Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. *Journal of Chemometrics* 2004, *18*, 231-241.
- (90) Harald Martens; Tormod Naes. *Multivariate Calibration*; Wiley & Sons, Great Britain, 1989.
- (91) Brereton, R. G. In *Calibration; Applied Chemometrics for Scientists*; John Wiley & Sons, Ltd: 2007; pp 193-220.
- (92) de Juan, A.; Tauler, R. Chemometrics applied to unravel multicomponent processes and mixtures: Revisiting latest trends in multivariate resolution. *Anal. Chim. Acta* 2003, *500*, 195-210.
- (93) Jaumot, J.; Gargallo, R.; de Juan, A.; Tauler, R. A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB. *Chemometrics Intellig. Lab. Syst.* 2005, *76*, 101-110.
- (94) Kessler, W.; Oelkrug, D.; Kessler, R. Using scattering and absorption spectra as MCR-hard model constraints for diffuse reflectance measurements of tablets. *Anal. Chim. Acta* 2009, *642*, 127-134.
- (95) Brereton, R. G. In *Pattern Recognition; Applied Chemometrics for Scientists*; John Wiley & Sons, Ltd: 2007; pp 145-191.
- (96) Berrueta, L. A.; Alonso-Salces, R. M.; Héberger, K. Supervised pattern recognition in food analysis. *Journal of Chromatography A* 2007, *1158*, 196-214.

(97) Liu, F.; He, Y. Classification of brands of instant noodles using Vis/NIR spectroscopy and chemometrics. *Food Res. Int.* 2008, *41*, 562-567.



## **Objetivos.**

---



La eficacia de gran cantidad de productos fitoterapéuticos ha sido demostrada desde el punto de vista científico; sin embargo, para asegurar la efectividad es necesario que cada medicamento de plantas medicinales cumpla con una serie de requisitos que garanticen su calidad y seguridad. Mientras que en países de la Unión Europea, el simple procedimiento de registro de productos fitoterapéuticos implica un estricto control de calidad, en países en desarrollo la situación es completamente contraria. No obstante, tanto en países desarrollados como en desarrollo, siguen apareciendo reacciones adversas por el consumo de suplementos dietéticos elaborados a base de plantas medicinales y los informes en la literatura señalan la mala calidad de una gran parte de los productos de origen herbolario que están disponibles en el mercado. Realizar el control de calidad en productos de origen herbolario no es tarea fácil, por lo complejo de la matriz y por la gran cantidad de componentes que presentan este tipo de muestras. Por ello la OMS recomienda utilizar los perfiles cromatográficos y espectroscópicos y hoy en día la construcción de huellas dactilares es una de las aproximaciones más poderosas para el control de calidad de plantas medicinales, aunque para poder manejar los resultados obtenidos es imprescindible aplicar herramientas quimiométricas. Además, el correlacionar la información obtenida de las huellas dactilares de las plantas y la actividad biológica, lo que se puede traducir en eficacia de las mismas, es uno de los requerimientos de investigación urgentes para el control de calidad de productos fitoterapéuticos.

Todo ello conlleva a que se haya realizado esta Tesis Doctoral con los siguientes objetivos:

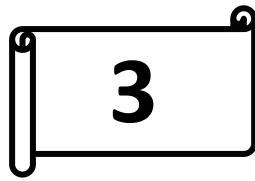
### **2.1. Objetivo General.**

Establecer métodos de control de calidad para plantas con uso medicinal a partir del tratamiento de sus perfiles cromatográficos o espectroscópicos (huellas dactilares) con técnicas quimiométricas.

## 2.2. Objetivos Específicos.

- 1) Obtener el perfil cromatográfico o espectroscópico de las plantas medicinales seleccionadas: *Panax ginseng*, *Eleutherococcus senticosus*, *Valeriana officinalis* y *Turnera diffusa*.
- 2) Pre-tratamiento de los datos obtenidos en el objetivo 1 y construcción de los modelos quimiométricos cualitativos, semi-cuantitativos o cuantitativos.
- 3) Comparar los métodos quimiométricos desarrollados y seleccionar el más adecuado para el control de calidad de cada una de las plantas medicinales elegidas.
- 4) Correlacionar, en su caso, la información obtenida de la huella dactilar de la planta con la actividad biológica.





***Eleutherococcus senticosus.***

---



### **3.1. Introducción.**

#### **3.1.1. Descripción de la planta medicinal.**

*Eleutherococcus senticosus* (Rupr. y Maxim.) Maxim. también conocido como *Acanthopanax senticosus*, ciwujia o ginseng siberiano, es un arbusto espinoso que pertenece a la familia *Araliaceae*, que crece en Siberia y partes del noreste de China y Corea [1-3]. En las últimas tres décadas ha habido un creciente interés en esta planta debido al amplio espectro de propiedades terapéuticas de sus raíces y rizomas; que incluyen efectos anticancerígeno, antioxidante, antibacterial, antiviral, hipocolesterolémico, actividades antiinflamatoria e inmunoestimulante, hipoglucemiante y propiedades adaptogénicas [1,2,4]. A pesar de que la acción farmacológica de *E. senticosus* podría parecerse a la de *Panax ginseng* y de que ambos son de la familia *Araliaceae*, su composición química es muy diferente y por lo tanto no pueden ser utilizados de manera intercambiable. Los compuestos aislados en *Eleutherococcus* son de grupos químicos muy variados e incluyen fenilpropanoides, lignanos, saponinas, cumarinas, el ácido betulínico triterpeno, provitaminas, vitaminas y otros compuestos que, aún hoy, están sin identificar o sin nombre [1]. Componentes característicos de esta planta son los eleuterósidos, compuestos heterogéneos que son responsables, en cierta medida, de sus actividades biológicas. El eleuterósido E (siringinoresinol *di-O-β-D-glucósido*, un lignano) y eleuterósido B (siringina, un derivado de fenilpropano) se han propuesto como los compuestos marcadores en la identificación y el análisis de *E. senticosus* [5].

#### **3.1.2. Problemática con el control de calidad del *Eleutherococcus senticosus*.**

Davydov et al. han descrito problemas de nomenclatura e identidad de *E. senticosus* debido a la ubiquidad de los compuestos marcadores seleccionados en su control de calidad [1]. Por otra parte, algunos productores sin escrúpulos continuamente están tratando de desarrollar maneras para obtener un perfil químico de sus productos similar al perfil auténtico presentado por la planta medicinal en cuestión. En estas circunstancias específicas, el uso de compuestos marcadores no sólo

es incapaz de confirmar la identidad de una planta concreta, sino que también podría ignorar la influencia de otros componentes químicos de origen intrínseco y, por lo tanto, en algunos casos, su uso puede ser inadecuado para los propósitos de control de calidad [6,7]. Además, sería deseable que el tiempo del análisis fuera lo más corto posible, aunque como la mayoría de estos procedimientos se basan en aplicaciones cromatográficas son bastante lentos, por ejemplo, la determinación de eleuterósidos puede tardar de 35 - 40 min por análisis sin incluir la preparación de las muestras [2,5,8].

### **3.1.3. Descripción del estudio.**

En este estudio se ha propuesto un procedimiento para una rápida identificación de las raíces de *E. senticosus* empleando huellas dactilares obtenidas mediante la espectroscopia de infrarrojo cercano (NIR). El procedimiento de preparación de la muestra es simple; sólo es necesario moler y tamizar la muestra para registrar los espectros. De esta manera, en lugar de un análisis cromatográfico largo, puede lograrse en unos pocos minutos un examen rápido por espectroscopia NIR, incluyendo la preparación de la muestra sin el uso de disolventes [9]. La complejidad de los espectros NIR requiere el uso de quimiometría para extraer y visualizar la información analítica útil; así, se han utilizado y comparado el PCA (principal component analysis), el método SIMCA (soft independent modeling of class analogy), el DA (discriminant analysis) y el PLS-DA (partial least squares - discriminant analysis). Por lo que sabemos, sólo hay un informe de uso de la espectroscopia molecular (FT-IR) para el establecimiento de huellas dactilares de *E. senticosus* [10], por lo que decidió investigar este tema.

## **3.2. Metodología específica.**

### **3.2.1. Descripción de las muestras.**

Con el fin de incluir la mayor variabilidad posible en las muestras, independientemente de su origen geográfico, las muestras fueron recolectadas de

varios proveedores durante los años 2007-2009. Las raíces de *E. senticosus* (ES), *Panax ginseng* (PG), *Panax quinquefolium* (PQ), *Panax notoginseng* (PN), *Lepidium meyenii* (LM), *Withania somnifera* (WS), *Angelica sinensis* (AS), *Pfaffia paniculata* (PP) y jengibre (una especia de cocina con propiedades estimulantes [G]) fueron obtenidas de Amorós Nature (Girona, Catalunya, España), Galke (Gittelde, Alemania), Richters (Ontario, Canada), Desert Botanicals (Gilbert, AZ, EE.UU.), Frontier Natural Products Co-op (Norway, IA, EE.UU.), MaxNature (Los Angeles, CA, EE.UU.), Modern Jungle (EE.UU.) y herboristerías de la zona metropolitana de Barcelona (España) y la ciudad de Monterrey (México). Varios proveedores autenticaron sus productos y proporcionaron un certificado de identidad. Por lo tanto, el conjunto de muestras consistió en miembros de la familia *Araliaceae* (ES, PG y PQ), y otros tipos de plantas que no están relacionados con *Araliaceae*, que son llamados "ginseng" de forma genérica. En la Tabla 3.1 se presenta un resumen de las muestras en estudio.

**Tabla 3.1.** Resumen de las muestras y conjuntos de espectros.

Muestras		Lotes			Número de espectros en		
ID <sup>a</sup>	Origen <sup>b</sup>	QS <sup>c</sup>	NSSQC <sup>d</sup>	Número Total <sup>e</sup>	PCA <sup>f</sup>	SIMCA, PLS-DA, DA	
						CS <sup>g</sup>	PS <sup>g</sup>
ES	CA, DE, ES, MX, US	5	15	20	45	42	18
PG	DE, ES, US	3	5	10	21	21	9
LM	DE, ES, US	3	3	5	12	0	15
PQ	CA, US	3	3	4	9	12	0
AS	ES, US	2	2	4	9	0	12
WS	DE, US	3	3	4	9	0	12
PP	DE, US	3	3	3	6	0	9
PN	US	1	1	1	3	0	3
Jengibre	US	0	0	1	3	0	3

<sup>a</sup> Identidad: ES, *Eleutherococcus senticosus*; PG, *Panax ginseng*; PQ, *Panax quinquefolium*; PN, *Panax notoginseng*; AS, *Angelica sinensis*; LM, *Lepidium meyenii*; PP, *Pfaffia paniculata* and WS, *Withania somnifera*.

<sup>b</sup> CA, Canadá; DE, Alemania; ES, España; MX, México and US, Estados Unidos.

<sup>c</sup> Proveedores de calidad; número de proveedores con una certificación de calidad.

<sup>d</sup> Número de muestras procedentes de proveedores con alguna certificación de calidad.

<sup>e</sup> Diferentes lotes de plantas adquiridos de varios proveedores y herboristerías en tres años (es decir, todas las muestras proceden de diferentes lotes, proveedores o años).

<sup>f</sup> Número de espectros empleados para construir un modelo de PCA global (N.B. éste no es el modelo empleado para la reducción de variables en el análisis discriminante).

<sup>g</sup> Número de espectros empleados en el conjunto de calibración (CS) o en el conjunto de predicción (PS) para la construcción y evaluación de los modelos de clasificación.

Las raíces se utilizaron tal como fueron adquiridas; es decir, no fue aplicado ningún secado adicional en el laboratorio para evitar la modificación de sus características originales y posibles cambios en la composición química. Las muestras recibidas como raíces enteras o troceadas se molieron en un molinillo Turmix Mill (Barcelona, España) y luego fueron sometidas a un tamizado por vibración empleando una tamizadora RP-15 (CISA, Barcelona, España). Las muestras recibidas en forma de polvo se tamizaron directamente. En todos los casos, fue obtenido un polvo con un tamaño de partícula igual o inferior a 100  $\mu\text{m}$ . Además, como parte del conjunto de predicción, se prepararon en el laboratorio diez mezclas sintéticas, mezclando diferentes cantidades de las plantas en estudio con muestras de *E. senticosus*; estas mezclas contenían finalmente 5%, 10% o 20% del adulterante seleccionado. En la Tabla 3.2 se presenta la composición de cada muestra adulterada.

**Tabla 3.2.** Composición de las mezclas sintéticas.

Mezcla no. <sup>a</sup>	Identidad del adulterante	Adulterante (%) <sup>b</sup>	ES (%) <sup>b</sup>
1	<i>P. ginseng</i>	5	95
2	<i>P. quinquefolium</i>	5	95
3	<i>P. notoginseng</i>	5	95
4	<i>Jengibre</i>	5	95
5	<i>P. ginseng</i>	10	90
6	<i>L. meyenii</i>	10	90
7	<i>A. sinensis</i>	10	90
8	<i>P. ginseng</i>	20	80
9	<i>W. somnifera</i>	20	80
10	<i>P. paniculata</i>	20	80

<sup>a</sup> Cada mezcla fue preparada por triplicado y se registró el espectro de cada replicado.

<sup>b</sup> Porcentajes por pesada. Peso total de 3 g.

### 3.2.2. Análisis espectroscópico en el infrarrojo cercano.

Se registraron los espectros NIR utilizando un espectrómetro con monocromador de barrido NIRSystems modelo 5000 (Foss NIRSystems, Silver Spring, MD), equipado con un módulo RCA (Rapid Content Analyzer), una lámpara halógena de cuarzo y un detector de sulfuro de plomo. Los espectros fueron obtenidos en modo de reflectancia sobre la región espectral de 1100 - 2498 nm con una resolución de 2

nm, empleando como referencia un patrón interno de cerámica, que viene provisto con el instrumento. La lectura de las muestras se realizó en una cubeta de muestra circular de vidrio (40 mm de diámetro y 10 mm de espesor) llena hasta el borde. Se aplicó un peso constante de 100 g durante aproximadamente 30 segundos en la muestra con el fin de obtener una compactación del polvo relativamente constante. Cada muestra se midió por triplicado, empleando tres porciones de muestra procesadas de forma independiente, donde cada espectro fue el promedio de 64 barridos sucesivos (obteniendo 700 datos por barrido). El control del espectrómetro y la manipulación de los archivos espectrales se realizaron utilizando el software Vision, versión 2.51, de NIRSystems Foss. Los datos de reflectancia fueron almacenados como el recíproco del logaritmo de la reflectancia,  $\log(1/R)$ . Se revisó todos los días de trabajo el buen funcionamiento del instrumento utilizando las opciones de diagnóstico proporcionados por el fabricante del mismo.

### **3.2.3. Análisis multivariante.**

Los espectros sin tratar fueron exportados desde el programa Vision como archivos NSAS para el análisis de datos. La construcción de los modelos de PCA, SIMCA y PLS-DA fue realizada utilizando el programa Unscrambler (v 9.8; CAMO, Oslo, Noruega), mientras que el DA se llevó a cabo con el programa SPSS (v 18.0, IBM, IL, EE.UU.). Con el fin de minimizar el efecto de dispersión de la radiación, los espectros necesitan ser pre-tratados; con este fin se ensayaron los pre-tratamientos de SNV (standard normal variate), primera y segunda derivada [11]. Las principales bandas de absorción debidas al agua fueron eliminadas para reducir la influencia de la humedad en los espectros NIR [12].

El PCA es a menudo el primer paso para el análisis de los espectros obtenidos con el fin de detectar patrones en las muestras registradas empleando NIR. En este caso se utilizó con el fin de examinar los agrupamientos naturales de las plantas en estudio, la detección de muestras anómalas, así como para desarrollar el método SIMCA y realizar el DA en las componentes principales (PCs) que explican más varianza del sistema.

Los resultados del método SIMCA fueron visualizados utilizando el gráfico de Cooman, que muestra la discriminación entre dos clases; al representar las muestras del conjunto de validación en este gráfico, es más fácil evaluar qué tan certera es la clasificación.

En este estudio se utilizó el DA para discriminar *E. senticosus* (grupo 1), de los otros miembros de la familia Araliaceae, *Panax ginseng* (grupo 2) y *Panax quinquefolium* (grupo 3) y de otras plantas que pudieran ser confundidas con ginseng. El DA requiere un mayor número de muestras que de variables; por esta razón, se realizó previamente un PCA y el DA fue realizado utilizando los scores de las primeras componentes principales, las que contienen la mayor parte de la información útil del sistema estudiado. Las reglas de clasificación se elaboraron utilizando únicamente las plantas de los grupos 1 – 3.

Para la construcción del modelo de PLS-DA, se asignaron a las muestras del conjunto de calibración dos variables indicando la categoría de las mismas: 1 para *E. senticosus* y 0 para otras plantas. Se considera que una muestra del conjunto de predicción está correctamente clasificada si su valor calculado está dentro de un intervalo de valores calculado a partir de los valores experimentales. El número óptimo de factores se determinó con el criterio del valor mínimo de PRESS (prediction sum of squares) [13]. Como figuras de mérito de los modelos construidos, fueron evaluados el RMSEC, el RMSEP y el coeficiente de determinación  $r^2$ . En la construcción de los modelos de calibración de PCA y PLS-DA, fue utilizada la validación cruzada (CV), empleando tres muestras por segmento y estableciendo la selección de muestra como aleatoria. La comparación de los resultados obtenidos con SIMCA, PLS-DA y DA se hizo en términos de sensibilidad, especificidad, la tasa general de éxito en la clasificación (GSR) y el porcentaje de mezclas clasificadas correctamente (MCC).



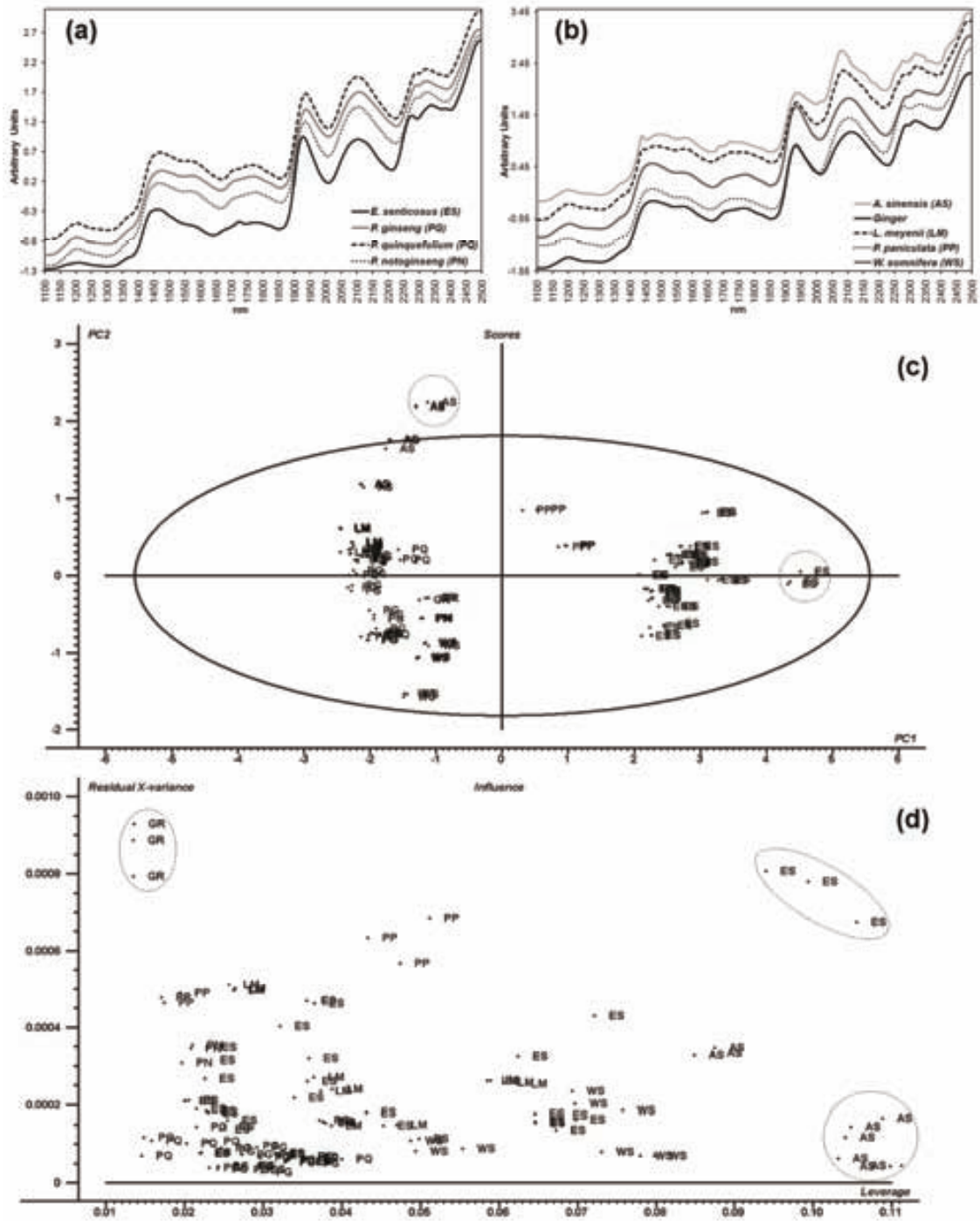
### **3.3. Resultados y discusión.**

#### **3.3.1. Características de los espectros de infrarrojo cercano.**

Las moléculas poliatómicas presentes en la matriz de las plantas medicinales exhiben múltiples vibraciones armónicas y de combinación, haciendo que sus bandas espectrales se superpongan. Como resultado, se pueden observar bandas espectrales muy amplias y sin rasgos distintivos; las Figuras 3.1a y 3.1b presentan espectros registrados.

No obstante, los espectros NIR contienen información molecular de las muestras, y esta información puede ser extraída por medio de métodos quimiométricos. Las bandas de humedad originadas por el enlace O-H, se encuentran alrededor de 1350 nm y 1950 nm, que son respectivamente bandas de armónicos y de combinación. Con el fin de excluir a estas bandas del análisis, las longitudes de onda comprendidas de 1306 a 1396 nm y de 1846 a 2016 nm, fueron excluidas en el pre-tratamiento y la construcción del modelo matemático. Es evidente que no se puede realizar la identificación de la especie de la planta a través de la inspección visual los espectros. Aparecen algunas diferencias en los espectros a longitudes de onda por encima de 2050 nm; sin embargo, también sufren un efecto relativamente importante de dispersión. Aunque no son visualmente perceptibles, existen diferencias en todo el espectro, por lo que se pueden construir los modelos de clasificación utilizando todas las longitudes de onda que quedan.

De todos los pre-tratamientos ensayados, el SNV efectuó la mejor corrección de los artefactos espectrales y produjo los modelos más sencillos, por lo que este pre-tratamiento fue seleccionado para ser aplicado en el resto del estudio.



**Figura 3.1.** (a) Espectros de infrarrojo cercano de los miembros de la familia *Araliaceae* y (b) plantas no relacionadas. (c) Representación de los scores de la PC1 vs. PC2 de un PCA global, presentando la elipse de confianza al 95%. (d) Gráfico de influencia; las muestras anómalas presentan valores altos de leverage y de varianza residual. En (c) y (d), se señalan las muestras sospechosas de ser anómalas.

### 3.3.2. Estudio del agrupamiento natural de las muestras.

Se realizó un PCA global empleando algunas muestras de cada tipo de planta (véase la Tabla 3.1). El conjunto de calibración incluyó 117 espectros, tomados al azar de los 156 espectros disponibles para ser evaluados; excluyendo las mezclas preparadas en el laboratorio. Este conjunto de calibración, que incluyó todos los tipos de plantas, nos permitió examinar la presencia de pautas. Las otras muestras se utilizaron para construir el conjunto de predicción y, posteriormente, fueron utilizadas como muestras externas (es decir, objetos completamente nuevos no utilizados en el desarrollo del modelo). El modelo de PCA fue construido empleando 4 PCs, presentando una varianza explicada del 97.90%. La capacidad del modelo en la predicción de nuevos objetos se evaluó por medio de la varianza explicada obtenida de la validación cruzada, la cual presentó un valor de 97.76%, lo que indica una buena capacidad predictiva. Los elevados valores de varianza explicada alcanzados en la calibración y la validación del modelo, así como la pequeña diferencia que presentan entre ellos, sugiere que el sistema no se ve afectado de forma importante por muestras anómalas. En el gráfico de scores se pueden observar las pautas de agrupación de acuerdo con la especie de planta (Figura 3.1c). La primera componente principal separó el *E. senticosus* de otras especies, al posicionar este grupo en valores positivos de scores mientras que la mayoría de las otras muestras se ubican en valores negativos. Las muestras de *P. paniculata* fueron las más cercanas al grupo del *E. senticosus*. Las fronteras del grupo *E. senticosus* parecen estar bien definidas, sólo una muestra apareció a una distancia moderada del grupo.

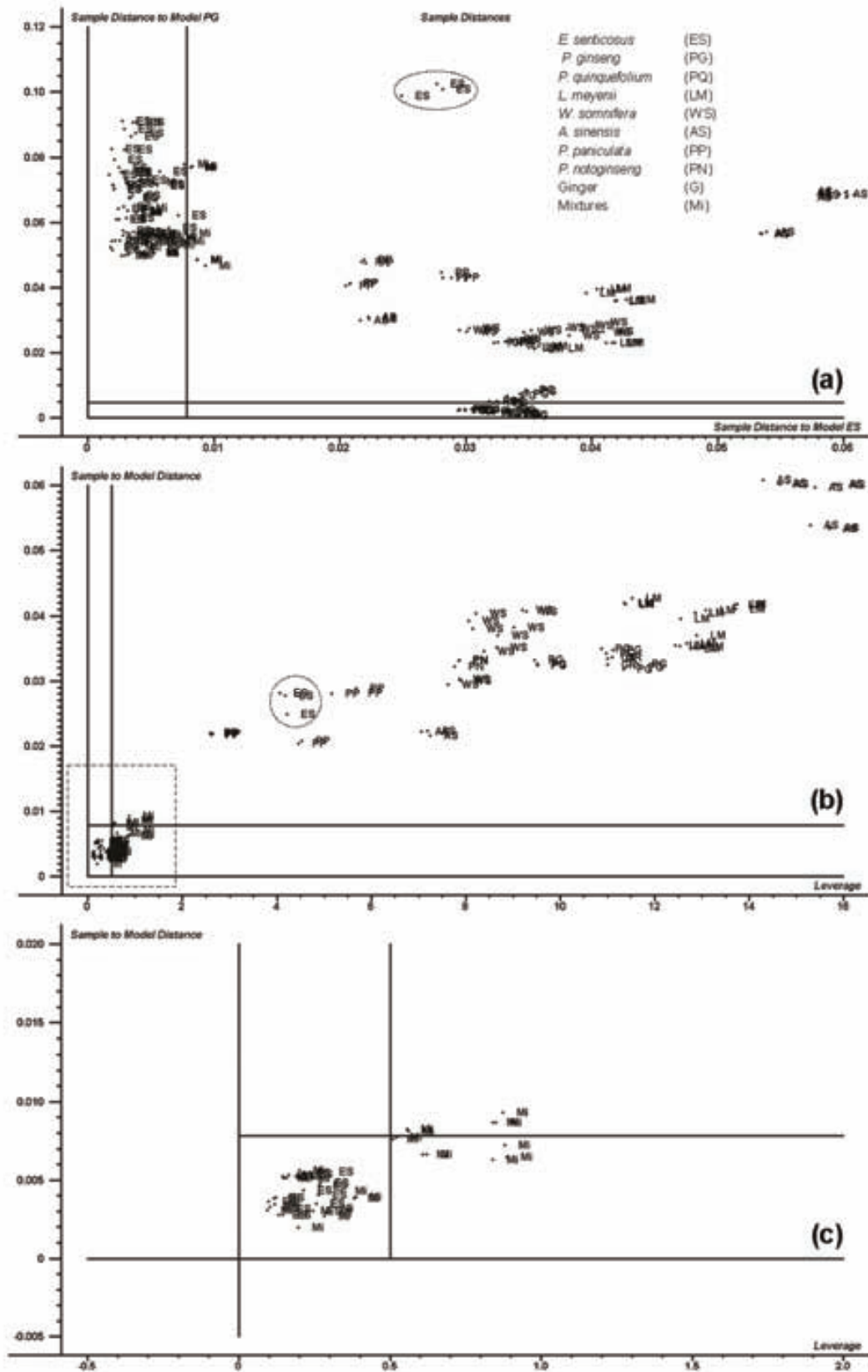
Así, parece posible la identificación de *E. senticosus* frente a otras especies de plantas, independientemente si están o no relacionadas con la familia *Araliaceae*. La muestra sospechosa ubicada entre los valores de score 4 y 5 de PC1, la muestra ES23, y aquellas fuera de la elipse de confianza del 95% (AS18, AS41) fueron examinadas usando el gráfico de influencia, con el fin de decidir si eran o no muestras anómalas (Figura 3.1d). A partir de la figura se determinó que la muestra ES23 era una muestra anómala evidente. Tomar una decisión para las otras muestras sospechosas (G, AS18 y AS41) fue más complicado. Aunque se sospechaba que estas muestras podrían ser

anómalas, fueron conservadas ya que son descritas correctamente por el modelo, tal como se puede observar del análisis de los residuales y de las distancias que presentaron al modelo construido. El comportamiento presentado por la muestra ES23 se debió a longitudes de onda que estaban por encima de 2200 nm.

Al igual que con todos los métodos no supervisados de clasificación, el PCA proporciona un panorama general para explicar las pautas de las muestras estudiadas, pero la asignación de nuevas muestras a una clase puede ser una tarea difícil con este método, ya que no calcula una regla que defina fronteras o regiones que diferencien entre los grupos de muestras en estudio. Por lo tanto, fueron empleados métodos supervisados con el fin de lograr una clasificación más objetiva.

### **3.3.3. Modelado blando independiente de analogías de clases (SIMCA).**

Como las muestras de la familia *Araliaceae* tienen espectros similares, consideramos apropiado estudiar la posibilidad de que las muestras de *E. senticosus* (ES) fueran mal clasificadas en las clases *P. ginseng* (PG) y *P. quinquefolium* (PQ). En consecuencia, para la construcción del modelo SIMCA se consideraron tres clases: ES, PG y PQ. El número de espectros empleados en cada clase, se indica en la Tabla 3.1. El número de componentes principales necesarias para definir adecuadamente cada clase fueron de 6, 4 y 3, respectivamente, presentando en todos los casos una varianza explicada > 97%, obtenida mediante la validación cruzada. Todas las variables incluidas en la construcción del modelo SIMCA exhibieron una buena capacidad de modelado y de discriminación. Adicionalmente, las distancias presentadas entre los modelos de PCA de cada clase fueron grandes, indicando que hubo una buena resolución entre ellas. En la Figura 3.2a se presenta un gráfico de Cooman, que muestra la distancia de las muestras de validación con respecto a las clases del *E. senticosus* y *P. ginseng*. A partir de esta gráfica llegamos a la conclusión de que las muestras de validación de *E. senticosus* tienen una buena probabilidad de ser asignadas a la clase correcta, dentro de un nivel de significación del 5%.



**Figura 3.2.** (a) Gráfico de Cooman que presenta la distancias ortogonales de las muestras de validación a las clases (modelos) del *E. senticosus* y el *P. ginseng*. (b) Leverage (distancia al centro del modelo de *E. senticosus*) vs. distancia de la muestra a la clase del *E. senticosus*. En (a) y (b) fue marcada con un círculo la muestra anómala ES23. (c) Ampliación del recuadro.

Como era de esperar, la muestra ES23 estaba más lejos de la clase correspondiente que otras muestras de *E. senticosus*, lo cual es un resultado coherente con las observaciones realizadas en los scores y loadings obtenidos del PCA global realizado previamente. En ningún caso las muestras de *E. senticosus* estuvieron cerca de la clase PG o de la clase PQ; además, ninguna muestra de *P. ginseng* empleada en la validación presentó proximidad a la clase del *E. senticosus*. Esta comprobación cruzada, permitió confirmar la correcta asignación de clase entre los miembros de la familia *Araliaceae*.

La clasificación final de las muestras del conjunto de predicción fue realizada con el gráfico de la distancia de la muestra a la clase del *E. senticosus* y el leverage (Figura 3.2b). Como era de esperar, todas las muestras de *E. senticosus* estuvieron dentro de los límites de clasificación, y la muestra ES23 fue excluida correctamente de la clase. El resto de las otras especies de plantas en estudio excedieron los valores límite de clasificación, claramente indicando que no pertenecen a la clase del *E. senticosus*.

La semejanza a *E. senticosus* presentada por las plantas estudiadas, expuesta en orden decreciente, es como sigue: PP, PN, WS, PG, LM and AS. Esta información también está de acuerdo con los resultados obtenidos en el PCA. Los resultados para las mezclas preparadas en el laboratorio se muestran en la Figura 3.2c, las muestras clasificadas como *E. senticosus* están dentro del cuadrado formado por los límites y los ejes. Las últimas cuatro mezclas de la Tabla 3.2 fueron excluidos por el modelo de la clase ES, por lo que es posible detectar muestras adulteradas, con certeza, desde un 20% de adulterante. A pesar de que la mezcla siete se preparó con un 10% de otra especie de planta y se identificó como adulterada, SIMCA no pudo detectar las mezclas cinco y seis, que presentan la misma proporción de adulterante. Los resultados de la mezcla siete podría atribuirse a la especie de planta utilizada, ya que AS es el producto más diferente a *E. senticosus* y es fácilmente discriminado por el modelo. En resumen, sólo es posible detectar mezclas que tengan alrededor de un 20% de contaminante.

**Tabla 3.3.** Resultados obtenidos en la clasificación con los modelos construidos.

Modelo	Tipo de Muestra	n <sup>a</sup>	Clasificado en <sup>b</sup>		Sensibilidad (%)	Especificidad (%)	MCC (%)	GSR (%)
			Clase ES	Clase NOES				
SIMCA	ES	15	15	0	100	100	40	84
	OHPs	66	0	66				
	Mezclas	30	18	12				
PLS-DA	ES	15	11	4	73	100	83	92
	OHPs	66	0	66				
	Mezclas	30	5	25				
DA	ES	15	15	0	100	96	0	70
	OHPs	66	3	63				
	Mezclas	30	30	0				

<sup>a</sup> Número de espectros en el conjunto de prueba (PS).

<sup>b</sup> Clase ES, muestras que pertenecen a la clase *E. senticosus*;  
Clase NOES, muestras que no pertenecen a la clase *E. senticosus*.

La Tabla 3.3 recoge los resultados generales obtenidos con SIMCA. La tasa general de éxito en la clasificación (GSR) representa el porcentaje total de las muestras correctamente clasificadas. Utilizando SIMCA se obtuvo un valor de GSR bastante bueno, sólo disminuido a causa de mezclas con cantidades de adulterante que eran inferiores al 20%. El porcentaje de mezclas correctamente clasificadas (MCC) es el porcentaje de mezclas clasificadas como ajenos a la clase *E. senticosus*. Dos parámetros relacionados afectan a este valor, el producto herbal utilizado para preparar las mezclas y la cantidad mínima de adulterante discriminada por el modelo. La sensibilidad es la proporción de muestras que pertenecen a la clase *E. senticosus* que fueron correctamente identificadas por el modelo matemático. La especificidad es la proporción de muestras que no pertenecen a la clase *E. senticosus* que fueron identificadas como extrañas por el modelo (aquí denotadas de manera general como OHPs, otros productos herbales). Estos dos criterios son similares a los propuestos por Lau, et al. [14]. El valor de sensibilidad obtenido indica que hay una probabilidad muy pequeña de obtener un falso negativo al realizar la prueba de identidad usando el

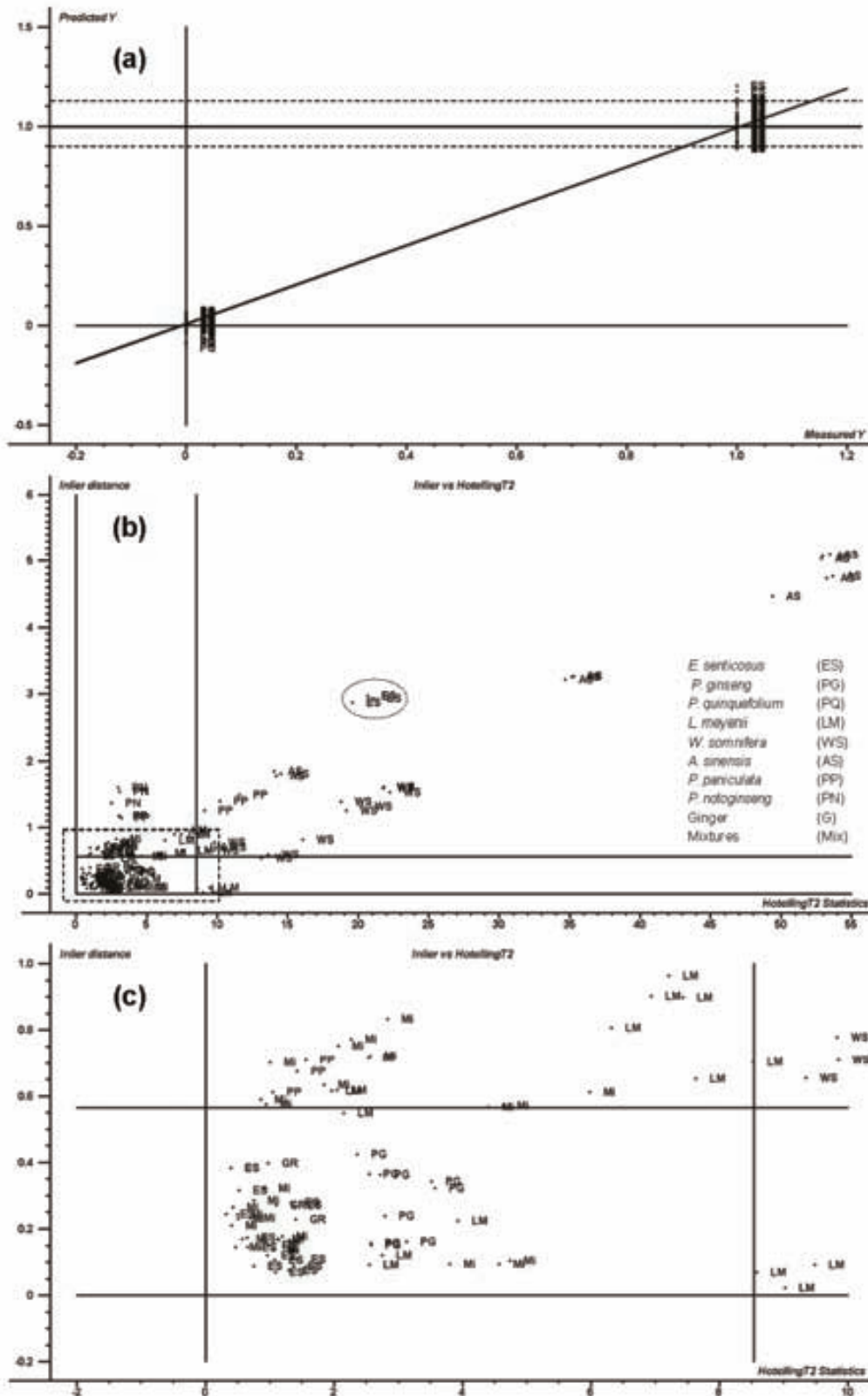
modelo construido. Además, de acuerdo con la especificidad, los OHPs en todos los casos serían excluidos de la clase del *E. senticosus*.

#### **3.3.4. Análisis discriminante con mínimos cuadrados parciales (PLS-DA).**

Las variaciones en las variables espectrales (X) y las variables categóricas (Y) fueron descritas por tres variables latentes (LVs), presentando una varianza explicada de 94.75% y 96.59%, respectivamente, obtenida por la validación cruzada. Este número relativamente pequeño de LVs sugiere una baja correlación en los espectros de diferentes clases, pero similitudes en los espectros dentro de las clases. Además, el bajo número de LVs proporciona una indicación de la buena diferenciación entre las clases construidas. Estas suposiciones fueron confirmadas mediante la inspección del gráfico de scores, donde sólo fue necesaria la LV1 para lograr una separación de la clase ES de las otras dos. No hubo presencia de muestras anómalas en el conjunto de calibración, de acuerdo a lo observado en un gráfico de leverage vs. la varianza residual de las variables X. En la Figura 3.3a, se presenta un gráfico de los valores Y de referencia vs. los valores Y calculados por el modelo (predicción de la validación cruzada), para todas las muestras de la clase ES. La línea de regresión obtenida en la predicción de las muestras utilizadas en la calibración, mostró una pendiente de 0.982, una ordenada de 0.011 y  $r^2 = 0.983$ . Así mismo, el RMSECV no superó el valor de 0.07 (unidades arbitrarias, au). Considerando estos resultados, se concluyó que el modelo PLS-DA desarrollado es útil para clasificar las nuevas muestras del conjunto externo de predicción.

Para ser clasificada correctamente, una muestra debe cumplir dos criterios: debe tener un espectro que no sea significativamente diferente de los espectros de las muestras que forman el conjunto de calibración y el valor de la variable categórica Y calculada, ha de caer dentro de un cierto intervalo de aceptación, calculado a partir de los resultados experimentales.





**Figura 3.3.** (a) Gráfico de los valores Y de referencia vs. los valores Y calculados; se presentan las líneas de regresión para la calibración y la validación. Las líneas discontinuas muestran el intervalo de aceptación y la línea continua el valor categórico asignado. (b) Representación de los valores de Hotelling T<sup>2</sup> vs. los valores del estadístico “inlier”; se consideró que todas las muestras con valores fuera de los límites no pertenecen a la clase *E. senticosus*. El círculo señala la muestra ES23. (c) Ampliación del recuadro.

Si una nueva muestra desconocida tiene un espectro significativamente diferente de los espectros de las muestras utilizadas en el conjunto de calibración, al realizar su predicción se observará en el gráfico de scores que se posiciona lejos de la muestra de calibración más próxima, en tal caso se marcará como un "inlier". La detección de inliers es la primera etapa para asignar las nuevas muestras a alguna de las clases. Con este fin, se utilizaron de manera conjunta la distancia de Mahalanobis entre cada nueva muestra calculada y su vecino de calibración más cercano, y el estadístico de Hotelling. Las muestras auténticas de *E. senticosus* del conjunto de predicción deberían estar bien descritas por el modelo construido y no ser marcadas como inliers. Sin embargo, cabe esperar la presencia de inliers, ya que en la construcción del modelo PLS-DA se han dejado fuera algunas especies de plantas (por ejemplo, *P. notoginseng*, *A. sinensis* y *W. somnifera*), las cuales serán posteriormente clasificadas. Esta es una situación común en los análisis de rutina. Los resultados obtenidos con el conjunto de predicción fueron representados como se muestra en la Figura 3.3b. El límite establecido por la distancia inlier fue de 0.57, mientras que el límite de Hotelling fue de 8.54. Ninguno de los inliers estaba relacionado con la clase del *E. senticosus*, excepto la muestra atípica ES23. No fue inesperado que algunos de los otros productos herbales (OHPs) y algunas mezclas se ajustaran al modelo debido a que sus espectros son semejantes a los del *E. senticosus* (véase la Figura 3.3c).

Cuando un método de regresión se utiliza para construir un modelo de clasificación, uno de los parámetros más importantes a definir es el intervalo de aceptación en el cual un nuevo objeto se considera clasificado correctamente en la clase, lo que se denomina "cut-off value". Se han utilizado varias alternativas para elegir un intervalo apropiado; por ejemplo véanse los trabajos desarrollados por Lau, et al. [14], Chen, et al. [15] y Woodcock, et al. [16]. En este estudio, se determinó el intervalo aceptable para una clasificación correcta considerando la línea de validación de la clase del *E. senticosus* y los errores en las predicciones de la validación cruzada. Para cada muestra de calibración, se halló la diferencia entre los valores Y calculados por el modelo y los de referencia, encontrando un valor mínimo (m) y máximo (M); se añadió el error asociado a la predicción, el RMSECV, y el promedio de los tres valores

absolutos fue el valor del intervalo aceptado para clasificar nuevas muestras en una clase u otra:

$$\text{Cutoff value} = \pm (|m| + |M| + |\text{RMSECV}|) / 3 \quad (3.1)$$

El siguiente paso para clasificar las muestras restantes, las que están dentro de los límites de los estadísticos de Hotelling y de la distancia inlier, fue comprobar el ajuste de sus valores calculados de  $Y$  en el intervalo de aceptación definido previamente. El intervalo hallado fue de  $\pm 0.10$ . Esto significa que todas las muestras con un valor calculado de  $Y$  de 0.90 a 1.10, y que no se han considerado inliers, serán clasificadas como pertenecientes a la clase del *E. senticosus*.

Los resultados generales después de los dos pasos se presentan en la Tabla 3.3. El valor de la sensibilidad disminuyó debido a las muestras con falso negativo, las cuales eran todas réplicas de la muestra ES33 y una réplica de la muestra ES35. Por otro lado, el método construido presentó una buena especificidad; ninguna de las muestras de otras plantas medicinales fue clasificada erróneamente. Además, la mayoría de las mezclas probadas fueron clasificadas correctamente fuera de la clase del *E. senticosus*; sólo la mezcla 3 y un replicado de las mezclas 4 y 5 fueron asignados erróneamente. Por lo tanto, mediante el uso de PLS-DA, es posible detectar muestras adulteradas a partir de 5% de adulterante, pero teniendo en cuenta que la especie de la planta empleada como adulterante no debe de ser de los miembros de la familia *Araliaceae*, especialmente el *P. notoginseng*. El PLS-DA presentó un mejor valor de GSR que SIMCA, principalmente debido a una mayor capacidad para detectar mezclas y una buena especificidad, aunque es posible rechazar algunas muestras auténticas de *E. senticosus*.

### **3.3.5. Análisis discriminante (DA).**

El método DA presenta la importante limitación de requerir más muestras que variables para que la matriz de covarianza pueda ser invertible. Se llevó a cabo un PCA con el fin de reducir el número de variables espectrales a unas pocas componentes

principales y se utilizaron como variables de entrada las primeras 10 PCs, que explicaban una varianza > 98%. Las muestras de las clases de *Araliaceae* previamente definidas (Tabla 3.1) fueron utilizadas para la construcción del modelo. El método de inclusión de variables por etapas (stepwise) fue aplicado a fin de incluir en el modelo discriminante las PCs estadísticamente significativas. Para tal fin, fue utilizado el criterio de maximización de la distancia de Mahalanobis, ya que éste considera la dirección y dispersión de los datos multivariantes [13]. Al final fue utilizado un modelo con las primeras 8 PCs como variables independientes.

La prueba de Box rechazó la igualdad de la covarianza para las clases establecidas, de modo que, en la clasificación fueron utilizadas las matrices de covarianza por separado. Al haber sido empleadas tres clases fueron obtenidas dos funciones discriminantes, aquí nombradas como CF1 y CF2. La matriz de estructura sugirió la mayor correlación absoluta entre la variable PC1 y CF1, mientras que la variable PC4 hizo lo mismo con CF2. Como era de esperar, CF1 fue la principal función para la clasificación, su correlación canónica fue de 0.997 con una varianza explicada del 94.10%. Para el conjunto de calibración, el porcentaje de muestras correctamente clasificadas fue del 100%.

Las muestras del conjunto de predicción fueron proyectadas sobre el modelo de PCA construido previamente y, a continuación, se utilizaron los scores obtenidos en el análisis discriminante. Los resultados alcanzados se pueden ver en la Tabla 3.3. Todas las muestras auténticas de *E. senticosus* fueron asignadas correctamente a su clase; sin embargo, la especificidad se vio afectada por la clasificación errónea de la muestra ES23. Los OHPs fueron excluidos de la clase ES; desgraciadamente, todas las mezclas fueron clasificadas en la clase ES, independientemente de la cantidad o el tipo de adulterante presente en la mezcla. Por lo tanto, el método DA no es adecuado para la detección de muestras adulteradas, tal como ha indicado el valor de MCC, pero aun así es útil para discriminar *E. senticosus* no procesado de otros productos fitoterapéuticos procedentes de diferentes orígenes. De todos los modelos ensayados, el método DA obtuvo el valor más bajo de GSR como resultado de la limitación en la detección de mezclas y la posibilidad de falsos positivos.

### 3.4. Conclusiones del capítulo.

Mediante el uso de la espectroscopia NIR y la quimiometría se consiguió la diferenciación rápida de materiales sin procesar de *E. senticosus* de otras ocho especies de plantas, relacionadas y no relacionadas con la familia *Araliaceae*.

Los tres métodos de clasificación evaluados, SIMCA, DA, y PLS-DA, permitieron hacer una buena discriminación entre *E. senticosus* y otras plantas; sin embargo, las pruebas con mezclas preparadas en laboratorio indicaron que sólo el PLS-DA y SIMCA eran realmente útiles para la detección de falsificaciones y adulteraciones. El método SIMCA y el PLS-DA obtuvieron una tasa general de éxito en la clasificación de 84% y 92%, respectivamente. La sensibilidad y la especificidad de todos los modelos (supervisados) de clasificación construidos fueron superiores al 73%.

El DA requiere la compresión previa de las variables originales por medio de un PCA y, en consecuencia, la reducción de la especificidad fue evidente, por lo que se vio afectada la capacidad para detectar muestras adulteradas.

El método SIMCA, intuitivo y sencillo, sería útil si el objetivo principal del estudio fuese detectar confusiones de especie, ya que tiene una excelente sensibilidad y especificidad.

El PLS-DA proporcionó los mejores resultados y es capaz de descubrir mezclas de plantas con una cantidad de adulterante menor a las que puede detectar SIMCA; la clasificación de las mezclas de laboratorio mostró que es posible detectar adulteraciones con una cantidad de material herbal cercana al 5%, dependiendo de la proximidad que tenga esta planta con la familia *Araliaceae*. Los resultados sugieren que la espectroscopia NIR puede ser utilizada para autenticar *E. senticosus*.

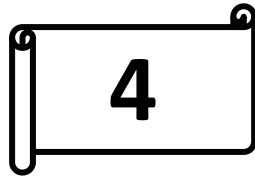


## Referencias.

- (1) Davydov, M.; Krikorian, A. D. *Eleutherococcus senticosus* (Rupr. & Maxim.) Maxim. (Araliaceae) as an adaptogen: a closer look. *J. Ethnopharmacol.* 2000, *72*, 345-393.
- (2) Zgórká, G.; Kawka, S. Application of conventional UV, photodiode array (PDA) and fluorescence (FL) detection to analysis of phenolic acids in plant material and pharmaceutical preparations. *J. Pharm. Biomed. Anal.* 2001, *24*, 1065-1072.
- (3) Niu, H.; Hsu, F.; Liu, I. Role of sympathetic tone in the loss of syringin-induced plasma glucose lowering action in conscious Wistar rats. *Neurosci. Lett.* 2008, *445*, 113-116.
- (4) Liu, K. Y.; Wu, Y.; Liu, I.; Yu, W. C.; Cheng, J. Release of acetylcholine by syringin, an active principle of *Eleutherococcus senticosus*, to raise insulin secretion in Wistar rats. *Neurosci. Lett.* 2008, *434*, 195-199.
- (5) Apers, S.; Naessens, T.; Van Miert, S.; Pieters, L.; Vlietinck, A. Quality control of roots of *Eleutherococcus senticosus* by HPLC. *Phytochem. Anal.* 2005, *16*, 55-60.
- (6) Mok, D. K. W.; Chau, F. Chemical information of Chinese medicines: A challenge to chemist. *Chemometrics Intellig. Lab. Syst.* 2006, *82*, 210-217.
- (7) Xie, P.; Chen, S.; Liang, Y.; Wang, X.; Tian, R.; Upton, R. Chromatographic fingerprint analysis—a rational approach for quality assessment of traditional Chinese herbal medicine. *Journal of Chromatography A* 2006, *1112*, 171-180.
- (8) European Pharmacopoeia. European Pharmacopoeia 5.0, Volumes 1 and 2. The European Pharmacopoeia Commission. 2005. Strasbourg: EDQM.
- (9) Pereira, A. F. C.; Pontes, M. J. C.; Neto, F. F. G.; Santos, S. R. B.; Galvão, R. K. H.; Araújo, M. C. U. NIR spectrometric determination of quality parameters in vegetable oils using iPLS and variable selection. *Food Res. Int.* 2008, *41*, 341-348.
- (10) Jin, Z.; Xu, S.; Sun, S.; Zhou, Q. Analysis of *Acanthopanax senticosus* Harms for different parts using Fourier transform infrared spectroscopy. *Guangpuxue Yu Guangpu Fenxi* 2008, *28*, 2859-2863.
- (11) Rinnan, Å.; Nørgaard, L.; Berg, F. v. d.; Thygesen, J.; Bro, R.; Engelsen, S. B. In *Chapter 2 - Data Pre-processing; Infrared Spectroscopy for Food Quality Analysis and Control*; Academic Press: San Diego, 2009; pp 29-50.
- (12) Ren, G.; Chen, F. Determination of moisture content of ginseng by near infra-red reflectance spectroscopy. *Food Chem.* 1997, *60*, 433-436.
- (13) Brereton, R. G. *Applied Chemometrics for Scientist*, John Wiley & Sons, Ltd, Great Britain, 2007.
- (14) Lau, C.; Chan, C.; Chau, F.; Mok, D. K. Rapid analysis of *Radix puerariae* by near-infrared spectroscopy. *Journal of Chromatography A* 2009, *1216*, 2130-2135.
- (15) Chen, Y.; Xie, M.; Yan, Y.; Zhu, S.; Nie, S.; Li, C.; Wang, Y.; Gong, X. Discrimination of *Ganoderma lucidum* according to geographical origin with near infrared diffuse reflectance spectroscopy and pattern recognition techniques. *Anal. Chim. Acta* 2008, *618*, 121-130.
- (16) Woodcock, T.; Downey, G.; O'Donnell, C. P. Near infrared spectral fingerprinting for confirmation of claimed PDO provenance of honey. *Food Chem.* 2009, *114*, 742-746.







**Panax ginseng.**

---



#### **4.1. Introducción.**

##### **4.1.1. Descripción de la planta medicinal.**

*Panax ginseng* es una famosa planta medicinal que se usa ampliamente en la medicina tradicional oriental, pertenece a la familia *Araliaceae* y crece principalmente en Corea y China [1-3]. Los componentes activos principales de *P. ginseng*, y también de otras especies pertenecientes al género *Panax*, son saponósidos del dammarano comúnmente referidos como ginsenósidos [1,4]. Ha habido un creciente interés en esta planta debido al amplio espectro de propiedades terapéuticas de sus raíces y rizomas, que incluyen efectos anticancerígeno, antioxidante, antibacterial, antiviral, hipocolesterolémico, actividades antiinflamatoria e inmunoestimulante, hipoglucemiante y propiedades adaptogénicas [5-7]. A pesar de que la acción farmacológica de otros miembros de la familia *Araliaceae* (p. ej. *E. senticosus*) podría parecerse a la de *Panax ginseng*, ya se ha mencionado en el capítulo 3 que la composición química de *Araliaceae* es muy diferente y por lo tanto no pueden ser utilizados de manera intercambiable [8].

##### **4.1.2. Problemática con el control de calidad del *Panax ginseng*.**

El *P. ginseng* tiene que ser diferenciado de otras especies de *Panax*, de otros integrantes de la familia *Araliaceae* y de otros tipos de plantas que no están relacionados con *Araliaceae*, que se consideran como ginseng. Dado que ha habido informes de reacciones adversas, es obvio que es necesaria una mejora en el control de calidad para evitar indeseables efectos secundarios [5,9].

En la actualidad, el método más común de análisis es CLAR de los extractos de ginseng con detección ya sea por espectrometría UV o de masas; sin embargo, los métodos cromatográficos destruyen las muestras de ginseng, consumen tiempo y disolventes y se centran principalmente en los ginsenósidos [4,6]. Se han desarrollado métodos que utilizan la espectroscopia de infrarrojo medio o cercano para el control de calidad de las especies de *Panax*; sin embargo, éstos se centran en la determinación

de la humedad y de ginsenósidos específicos [7,10], en la distinción de áreas de cultivo y calidades (grados) del ginseng [6,11] o en discriminar el ginseng de un pequeño número de plantas de aspecto similar [3-5,12]. A partir de aquí, es evidente que sigue siendo necesaria una estrategia complementaria para el establecimiento de la huellas dactilar del *P. ginseng*; sobre todo para discriminarlo de un número mayor de posibles falsificaciones, incluyendo las mezclas de plantas.

#### **4.1.3. Descripción del estudio.**

En este estudio se ha desarrollado un procedimiento para una identificación rápida de las raíces de *P. ginseng* empleando huellas dactilares obtenidas mediante la espectroscopia de infrarrojo cercano (NIR). El procedimiento de preparación de la muestra es simple, sólo es necesario moler y tamizar para registrar los espectros. De esta manera, en lugar de un largo análisis cromatográfico, en unos pocos minutos puede lograrse un examen rápido por espectroscopia NIR, incluyendo la preparación de la muestra sin el uso de disolventes.

La complejidad de los espectros NIR requiere el uso de quimiometría para extraer y visualizar la información analítica útil; en este trabajo, se han utilizado y comparado el PCA (principal component analysis), el método SIMCA (soft independent modeling of class analogy), el DA (discriminant analysis) y el PLS-DA (partial least squares - discriminant analysis).

Además, se han utilizado mezclas preparadas en el laboratorio en la evaluación de la capacidad de los modelos para detectar muestras adulteradas. Si la intención del análisis de las mezclas de plantas es la exploración de su composición, una opción útil podría ser el uso de los espectros NIR y el MCR-ALS (multivariate curve resolution - alternating least squares), ya que el análisis habitual por CLAR de mezclas de plantas podría llevar mucho tiempo y consume cantidades considerables de disolventes. Así, se ha llevado a cabo el análisis semicuantitativo de mezclas binarias de *P. ginseng* y un adulterante utilizando MCR-ALS.

## **4.2. Metodología específica.**

### **4.2.1. Descripción de las muestras.**

Las raíces de *Panax ginseng* (PG), *E. senticosus* (ES), *Panax quinquefolium* (PQ), *Panax notoginseng* (PN), *Lepidium meyenii* (LM), *Withania somnifera* (WS), *Angelica sinensis* (AS), *Pfaffia paniculata* (PP) y jengibre (ginger, una especia de cocina con propiedades estimulantes [G]) fueron obtenidas de Amorós Nature (Girona, Catalunya, España), Galke (Gittelde, Alemania), Richters (Ontario, Canada), Desert Botanicals (Gilbert, AZ, EE.UU.), Frontier Natural Products Co-op (Norway, IA, EE.UU.), MaxNature (Los Angeles, CA, EE.UU.), Modern Jungle (EE.UU.) y herboristerías de la zona metropolitana de Barcelona (España) y la ciudad de Monterrey (México).

Así, el conjunto de muestras consistió en miembros de la familia *Araliaceae* (PG, ES, PQ y PN), y otros tipos de plantas que no están relacionados con *Araliaceae*, y que son considerados como ginseng. Varios proveedores poseen una certificación de calidad (p. ej. ISO, GMP o de otro tipo); otros son capaces de de autenticar sus productos a través de botánicos expertos que identifican las plantas de sus propios cultivos; de este modo, pueden proporcionar un comprobante de identidad.

Al igual que en el estudio de *E. senticosus*, las muestras fueron recolectadas durante los años 2007 a 2009 con el fin de incluir la mayor variabilidad posible. En la Tabla 4.1 se presenta un resumen de las muestras en estudio. El tratamiento también fue similar; las raíces se utilizaron tal como fueron adquiridas, es decir no fue aplicado ningún secado adicional en el laboratorio para evitar la modificación de sus características originales y posibles cambios en la composición química. Las muestras recibidas como raíces enteras o troceadas se molieron en un molinillo Turmix Mill (Barcelona, España) y fueron posteriormente tamizadas por vibración con una criba vibratoria RP-15 (CISA, Barcelona, España). Las muestras recibidas en forma de polvo se tamizaron directamente. En todos los casos, fue obtenido un polvo con un tamaño de partícula igual o inferior a 100  $\mu\text{m}$ .

**Tabla 4.1.** Características de las muestras y conjuntos de espectros.

Muestras		Lotes			Número de espectros en		
ID <sup>a</sup>	Origen <sup>b</sup>	QS <sup>c</sup>	NSSQC <sup>d</sup>	Número Total <sup>e</sup>	PCA <sup>f</sup>	SIMCA, PLS-DA, DA	
						CS <sup>g</sup>	PS <sup>g</sup>
PG	DE, ES, US	3	12	18	54	36	18
ES	CA, DE, ES, MX, US	5	7	14	42	27	15
LM	DE, ES, US	3	3	5	15	0	15
PQ	CA, US	3	3	4	12	12	0
AS	ES, US	2	2	4	12	0	12
WS	DE, US	3	3	4	12	0	12
PP	DE, US	3	3	3	9	0	9
PN	US	1	1	1	3	0	3
Jengibre	US	0	0	1	3	0	3
Mezclas <sup>h</sup>	---	---	---	---	30	0	30

<sup>a</sup> Identidad: PG, Panax ginseng; ES, Eleutherococcus senticosus; PQ, Panax quinquefolium; PN, Panax notoginseng; AS, Angelica sinensis; LM, Lepidium meyenii; PP, Pfaffia paniculata and WS, Withania somnifera.

<sup>b</sup> CA, Canadá; DE, Alemania; ES, España; MX, México and US, Estados Unidos.

<sup>c</sup> Proveedores de calidad; número de proveedores con una certificación de calidad.

<sup>d</sup> Número de muestras procedentes de proveedores con alguna certificación de calidad.

<sup>e</sup> Diferentes lotes de plantas adquiridos de varios proveedores y herboristerías en tres años (es decir, todas las muestras proceden de diferentes lotes, proveedores o años).

<sup>f</sup> Número de espectros empleados para construir un modelo de PCA global (N.B. éste no es el modelo empleado para la reducción de variables en el análisis discriminante).

<sup>g</sup> Número de espectros empleados en el conjunto de calibración (CS) o en el conjunto de predicción (PS) para la construcción y evaluación de los modelos de clasificación.

<sup>h</sup> ---, No aplica.

Además, como parte del conjunto de predicción, fueron preparadas en el laboratorio diez mezclas sintéticas mediante el mezclado de diferentes cantidades de las plantas en estudio con muestras de *P. ginseng*. En la Tabla 4.2 se presenta la composición de cada muestra adulterada, incluyendo aquéllas utilizadas en el análisis semicuantitativo por MCR. Todas las muestras fueron almacenadas a temperatura de laboratorio (25 °C aprox.), protegidas de la luz y la humedad, y analizadas dentro de la fecha de caducidad dada por el fabricante.

**Tabla 4.2.** Composición de las mezclas.

Modo	Número/Tipo de mezcla <sup>a</sup>	Identidad del Adulterante	PG (%) <sup>b</sup>	Adulterante (%) <sup>b</sup>
Clasificación (SIMCA, PLS-DA, DA)	1	<i>P. quinquefolium</i>	95	5
	2	<i>E. senticosus</i>	95	5
	3	<i>P. notoginseng</i>	95	5
	4	<i>Ginger</i>	95	5
	5	<i>P. quinquefolium</i>	90	10
	6	<i>L. meyenii</i>	90	10
	7	<i>A. sinensis</i>	90	10
	8	<i>P. quinquefolium</i>	80	20
	9	<i>W. somnifera</i>	80	20
	10	<i>P. paniculata</i>	80	20
Semi- cuantificación (MCR)	Simulación numérica	<i>E. senticosus</i>	90	10
			85	15
			80	20
			75	25
			70	30
	Experimental	<i>E. senticosus</i>	90	10
			80	20
			70	30
			60	40
			60	40

<sup>a</sup> Cada mezcla se preparó por triplicado y se registró el espectro de cada réplica (exceptuando las mezclas simuladas).

<sup>b</sup> Porcentajes por pesada. Peso total de 3 g.

#### 4.2.2. Análisis espectroscópico en el infrarrojo cercano.

Los espectros NIR fueron obtenidos usando un espectrómetro con monocromador de barrido NIRSystems modelo 5000 (Foss NIRSystems, Silver Spring, MD), equipado con un módulo RCA (Rapid Content Analyzer), una lámpara halógena de cuarzo y un detector de sulfuro de plomo. Los espectros fueron obtenidos en modo de reflectancia, empleando como referencia un patrón de cerámica, provista de fábrica con el instrumento, sobre la región espectral de 1100 - 2498 nm con una resolución de 2 nm. Las muestras fueron analizadas en una cubeta de muestra circular de vidrio (40 mm de diámetro y 10 mm de espesor) llena hasta el borde. Se aplicó un peso constante de 100 g, durante aproximadamente 30 segundos, en la muestra con el fin de obtener en la cubeta una compactación del polvo relativamente constante.

Cada muestra se registró por triplicado, empleando tres porciones de muestra procesadas de forma independiente, donde cada espectro fue el promedio de 64 barridos sucesivos (obteniendo 700 datos por barrido). El control del espectrómetro y la manipulación de los archivos espectrales se realizaron utilizando el programa Vision, versión 2.51, de NIRSystems Foss. Los datos de reflectancia fueron almacenados como el recíproco del logaritmo de la reflectancia,  $\log(1/R)$ . Se revisó el buen funcionamiento del instrumento todos los días de trabajo, utilizando las opciones de diagnóstico proporcionados por el fabricante.

#### **4.2.3. Análisis multivariante.**

Los espectros sin tratar fueron exportados desde el programa Vision como archivos NSAS para el análisis de datos. El PCA, SIMCA y PLS-DA fueron realizados utilizando el programa PLS-Toolbox (v. 6.5.1; Eigenvector Research, Inc.) para uso en el entorno computacional MATLAB (v 7.9, The MathWorks, Inc.), mientras que el DA fue llevado a cabo con el programa SPSS (v 19.0; IBM). En la Tabla 4.1 se presentan los espectros utilizados para construir y probar los modelos matemáticos. Con el fin de minimizar el efecto de dispersión de la radiación, los espectros necesitan ser pretratados; con este fin se utilizaron los pretratamientos de SNV (standard normal variate), primera y segunda derivada [13]. Se eliminaron las principales bandas de absorción del agua para reducir la influencia de la humedad en los espectros NIR [10]. El cálculo de MCR-ALS fue realizado utilizando el software MATLAB y la interfaz gráfica de usuario desarrollada por Jaumot et al. [14], que está disponible libremente en [http://www.ub.edu/mcr/web\\_mcr/welcome.htm](http://www.ub.edu/mcr/web_mcr/welcome.htm).

Como es habitual, el estudio previo de las muestras se realizó mediante un PCA. También se empleó esta técnica para el modelado de cada una de las tres clases consideradas, *P. ginseng*, *P. quinquefolium* y *E. senticosus*, y que se utilizaron en el método de clasificación SIMCA. El resto de las muestras y las mezclas de laboratorio se utilizaron en el conjunto de prueba. Al aplicar el método SIMCA, los límites para el espacio ( $T^2$ ) definido por el modelo de *P. ginseng* y su espacio residual (Q) fueron



establecidos a un nivel de confianza del 95% y las muestras con distancias dentro de los límites fueron clasificadas como miembros de la clase.

Se empleó el DA para discriminar *P. ginseng* (grupo 1) de los otros miembros de la familia Araliaceae, de *E. senticosus* (grupo 2), de *P. quinquefolium* (grupo 3) y de plantas consideradas como ginseng; los grupos 1 a 3 se utilizaron para elaborar las reglas de clasificación. Puesto que el DA requiere un mayor número de muestras que de variables, se realizó previamente una reducción de variables con un PCA, seleccionando las componentes que explican la mayor variabilidad de los datos.

En el caso del PLS-DA, las variables categóricas fueron asignadas a las muestras del conjunto calibración de la siguiente manera: 1 a *P. ginseng* y 0 para las otras plantas (ES y PQ). Los criterios para considerar que una muestra está correctamente clasificada se han mencionado en la sección 1.6.3.2. y su aplicación ha sido descrita en las secciones 3.2.3 y 3.3.4.

Los modelos de calibración de PCA y PLS fueron validados empleando la validación cruzada con el método de bloques contiguos y tres muestras por segmento, los replicados de muestra son eliminados o introducidos juntos en el conjunto de calibración. El número óptimo de factores para construir el modelo se determinó con el criterio del valor mínimo de PRESS (Prediction Sum of Squares) [15]. Los estadísticos evaluados en la construcción de los modelos de calibración incluyeron RMSEC (root mean square error of calibration), RMSECV (root mean square error of cross validation) y el coeficiente de determinación  $r^2$ .

La comparación de los resultados obtenidos con SIMCA, PLS-DA y DA se hizo en términos de sensibilidad, especificidad, la tasa general de éxito en la clasificación (GSR) y el porcentaje de mezclas clasificadas correctamente (MCC). Podemos definir estos términos de manera análoga a la descrita en la sección 3.3.3. La sensibilidad es la proporción de muestras que pertenecen a la clase *P. ginseng* que son correctamente identificadas por el modelo matemático. La especificidad es la proporción de muestras que no pertenecen a la clase *P. ginseng* (es decir, ES, PQ, PN, LM, WS, AS y PP) que son

correctamente identificados por el modelo como extrañas. El valor de GSR se obtiene calculando el porcentaje global de muestras predichas que fueron correctamente clasificadas con respecto a la clase *P. ginseng*. El porcentaje de mezclas clasificadas correctamente es útil para comparar la capacidad de los modelos para discriminar muestras adulteradas.

Para evaluar la viabilidad de aplicar el algoritmo MCR-ALS se utilizaron dos conjuntos de datos, una primera matriz D se construyó con espectros de mezclas simuladas matemáticamente, las cuales fueron formadas a partir de los espectros puros de *P. ginseng* y un adulterante seleccionado. La otra matriz D se construyó registrando espectros de mezclas de plantas preparadas en el laboratorio y cuya composición se muestra en la Tabla 4.2. A fin de facilitar el proceso de resolución, ambas matrices se aumentaron con varios espectros de *P. ginseng* puro. El número propuesto de factores se estableció en 2 y como estimaciones iniciales se utilizaron los espectros “puros” de las plantas. Asimismo, fue evaluada la utilidad de las limitaciones de no negatividad, closure y las restricciones de igualdad. El valor de convergencia seleccionado fue del 0.1%.

### **4.3. Resultados y discusión.**

#### **4.3.1. Características de los espectros de infrarrojo cercano.**

Las moléculas complejas, presentes en la matriz de las raíces de ginseng y las plantas analizadas, exhiben múltiples vibraciones armónicas y de combinación, haciendo que sus bandas se superpongan; las Figuras 4.1a y 4.1b presentan los espectros registrados. Como resultado, se obtienen espectros con bandas muy amplias y sin rasgos distintivos, características comunes en el análisis por NIR de las plantas medicinales [16]. Al igual que para *E. senticosus*, las bandas de humedad fueron excluidas en el pre-tratamiento y la construcción de los modelos matemáticos. De todos los pre-tratamientos ensayados, el SNV efectuó la mejor corrección de la dispersión de la radiación y permitió obtener los modelos más sencillos, por lo que se aplicó en el resto del estudio.

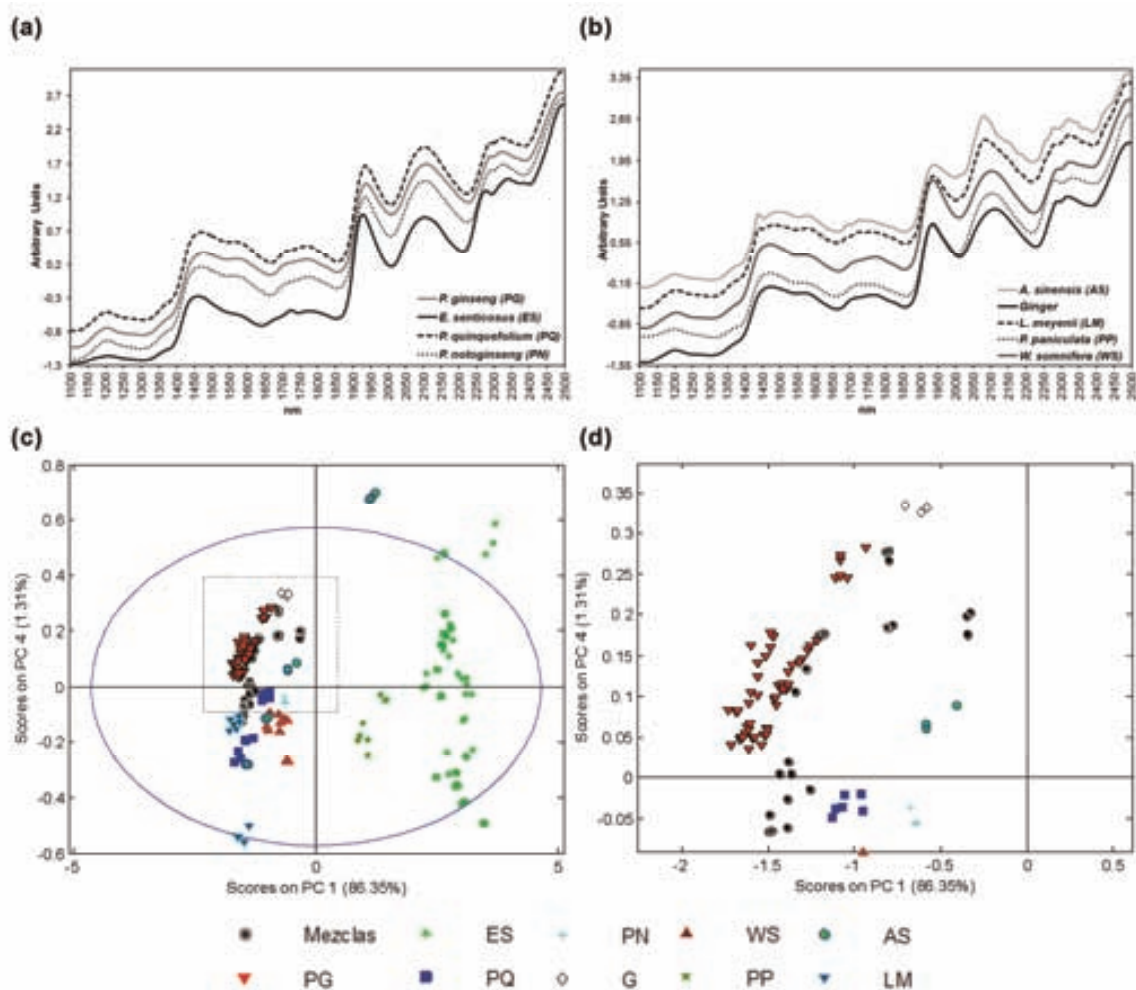
#### 4.3.2. Estudio de la similitud de las muestras.

Se realizó un PCA global empleando las muestras indicadas en la Tabla 4.1 que con 5 componentes principales (PCs) explicaba una varianza de 98.86% con un RMSECV de 0.302. Este valor de RMSECV y el alto porcentaje de varianza explicada empleando pocos PCs sugiere que el modelo no es afectado de forma importante por ninguna muestra anómala. En los gráficos de scores se observó un agrupamiento complejo de las muestras según la especie de la planta; el gráfico de PC1 vs. PC4 fue el mejor para examinar la forma de agrupamiento del *P. ginseng* y para representar la distribución natural y el agrupamiento de las otras especies de plantas alrededor del grupo del *P. ginseng*, véanse las Figuras 4.1c y 4.1d. La proximidad de las otras especies de plantas al grupo del *P. ginseng*, con la excepción de las muestras de *E. senticosus* y *P. paniculata*, indica que puede ser difícil conseguir una buena clasificación; nótese que algunas de las mezclas están situadas junto a muestras *P. ginseng* y también la proximidad de *P. quinquefolium*. Afortunadamente, las fronteras del grupo del *P. ginseng* parecen ser claras y el grupo es compacto, sólo unas pocas muestras aparecen a una cierta distancia del grupo; estas características harán posible discriminar las otras especies de plantas cuando se construya el correspondiente modelo de la clase. Así, es posible la identificación del *P. ginseng* frente a otras especies de plantas, independientemente si están o no relacionadas con la familia *Araliaceae*.

Las muestras sospechosas fuera de la elipse de confianza del 95% fueron examinadas mediante la revisión de sus valores de residual (Q) y de leverage ( $T^2$ ) y, finalmente, se decidió conservarlas ya que se llegó a la conclusión de que obedecían el modelo de PCA construido. La inspección de los residuales de los loadings mostró la posibilidad de que el SNV no haya corregido completamente los efectos de la dispersión de la radiación en longitudes de onda por encima de 2270 nm.

Al igual que todos los métodos no supervisados de clasificación, el PCA proporciona un panorama general para explicar las pautas de las muestras estudiadas, pero la asignación de nuevas muestras a una clase puede ser una tarea difícil con este método, ya que no calcula una regla que defina fronteras o regiones para diferenciar

entre los grupos obtenidos a partir de las muestras en estudio. Por lo tanto, fueron empleados métodos supervisados con el fin de lograr una clasificación más objetiva.



**Figura 4.1.** (a) Espectros de infrarrojo cercano de los miembros de la familia *Araliaceae* y (b) plantas no relacionadas utilizadas como sustitutos del ginseng. (c) Gráfico de scores PC1 vs. PC4 de un PCA global presentando la elipse de confianza al 95%. (d) Ampliación del recuadro.

#### 4.3.3. SIMCA (Soft Independent Modeling of Class Analogy).

Se consideraron tres clases para construir el modelo SIMCA, como se describe en la Tabla 4.1. El principal interés es estudiar los posibles errores de clasificación de las muestras de *P. ginseng* en las clases PQ y ES, puesto que las muestras pertenecientes a la familia *Araliaceae* tienen características similares. Además, la clase PG debe ser capaz de discriminar las muestras adulteradas y otras especies de plantas diferentes a las que constituyen la clase. Al construir el modelo SIMCA se necesitaron 5 PCs para definir correctamente la clase PG, 5 para la clase ES y 3 para la clase PQ,

obteniendo una varianza explicada > 98% y un valor de RMSECV < 0.08. Todas las variables incluidas en la construcción del modelo SIMCA exhibieron una buena capacidad de modelado. Ninguna muestra de *P. ginseng* del conjunto de predicción mostró cercanía suficiente a la clase PQ o la clase ES para ser clasificados erróneamente; esto permite confirmar la asignación correcta de clase entre los miembros de la familia *Araliaceae*.

**Tabla 4.3.** Comparación de los resultados de la calcificación.

Modelo	ID	n <sup>a</sup>	Clasificado <sup>b</sup>		Sensibilidad (%)	Especificidad (%)	MCC (%) <sup>c</sup>	GSR (%) <sup>d</sup>
			PG	AHG				
SIMCA	PG	18	18	0	100	100	80	95
	OHs <sup>e</sup>	69	0	69				
	Mezclas	30	6	24				
PLS-DA	PG	18	18	0	100	91	60	85
	OHs <sup>e</sup>	69	6	63				
	Mezclas	30	12	18				
DA	PG	18	18	0	100	87	53	80
	OHs <sup>e</sup>	69	9	60				
	Mezclas	30	14	16				

<sup>a</sup> Número de espectros en el conjunto de prueba (PS).

<sup>b</sup> PG, *P. ginseng*; AHG, otro grupo de plantas.

<sup>c</sup> Porcentaje de mezclas correctamente clasificadas como ajenas a la clase PG.

<sup>d</sup> Tasa general de éxito en la clasificación.

<sup>e</sup> Otras plantas, aquí se encuentran *E. senticosus*, *P. notoginseng*, *L. meyenii*, *W. somnifera*, *A. sinensis*, *P. paniculata* y el jengibre (ginger).

La Tabla 4.3 resume el resultado al utilizar SIMCA. El valor de sensibilidad obtenido indica que hay una probabilidad muy pequeña de obtener un falso negativo al realizar la prueba de identidad usando el modelo construido. Además, de acuerdo con la especificidad, las otras especies de plantas serían excluidas de la clase PG en todos los casos. Utilizando SIMCA se obtuvo un valor de GSR bastante bueno, sólo disminuido a causa de los espectros de las mezclas uno y cinco, ambas preparadas con ginseng Americano (PQ), que fueron clasificadas erróneamente como *P. ginseng*. Este

error de clasificación no es inesperado debido a que estas mezclas fueron las situadas junto al grupo del *P. ginseng* en la Figura 4.1d. Aunque algunas de las especies de plantas utilizadas en la preparación de mezclas eran lo suficientemente diferentes como para ser discriminadas por el modelo SIMCA, incluso en una cantidad del 5%, otras especies con espectros similares al del *P. ginseng*, por ejemplo el *P. quinquefolium*, podrían ser erróneamente identificadas en mezclas a un 10%; por lo tanto, sólo es posible detectar con fiabilidad las muestras que contengan alrededor de un 20% de adulterante.

#### **4.3.4. PLS-DA (partial least squares - discriminant analysis).**

Utilizando la matriz X de espectros, las clases codificadas en el vector Y fueron descritas con tres variables latentes (LVs), presentando una varianza explicada de 94% y 98%, respectivamente. El utilizar un número de LVs igual al número de clases sugiere una baja correlación química en los espectros de diferentes clases, pero similitudes de los espectros dentro de las clases; es decir, el número bajo de LVs proporciona una idea de la buena capacidad del modelo para describir las clases propuestas. Los estadísticos de la regresión mostraron un RMSEC de 0.128, un RMSECV de 0.182 y  $r^2$  obtenida de la validación cruzada de 0.869. Este valor de  $r^2$  puede parecer bajo; sin embargo, durante la construcción del modelo no se detectaron muestras anómalas. Un valor de corte de  $\pm 0.20$  se determinó empleando la ecuación 3.1 (sección 3.3.4), considerando la línea de validación de la clase PG y el error en las predicciones. Esto significa que todas las muestras calculadas que se han estimado dentro de los límites establecidos por el valor de Hotelling  $T^2$  y el valor de residuales Q a un nivel de confianza del 95% y con un valor Y de 0.80 a 1.20 serían clasificadas como pertenecientes a la clase del PG.

Los resultados obtenidos se muestran en la Tabla 4.3. El valor de especificidad fue de un 91% debido a muestras que dieron falso positivo, todas ellas replicados de muestras de *L. meyenii*. Por otro lado, el modelo construido presenta una buena sensibilidad ya que ninguna de las muestras de *P. ginseng* fue clasificada erróneamente. Desafortunadamente, más de un tercio de los espectros de las mezclas

analizadas (1, 3, 5 y 6) fueron clasificadas erróneamente ya que las especies adulterantes de esas muestras no fueron discriminadas por el modelo PLS-DA; ver la Tabla 4.2. De acuerdo con estos resultados, sólo es posible detectar las mezclas con una cantidad de adulterante que esté necesariamente en torno a un 20%, ya que fueron confundidas varias especies de plantas aparte de *P. quinquefolium*.

Aunque el PLS-DA presentó un valor menor de GSR que SIMCA, principalmente debido a una inferior capacidad para detectar mezclas y su menor especificidad, es posible rechazar la mayoría de las muestras de otras especies y hay una probabilidad muy pequeña de clasificar una muestra auténtica de *P. ginseng* como un falso negativo.

#### **4.3.5. DA (discriminant analysis).**

Debido a las características del DA, no es posible establecer que una muestra dada no pertenezca a una de las clases definidas previamente en el modelo; es decir, todas las muestras a clasificar deben ser asignadas a una clase y las especies no incluidas en el modelo serán clasificadas en alguna de ellas aunque no sea la que le corresponde. En consecuencia, se debe determinar qué muestras podrían ser confundidas con respecto a la clase del PG. Como ya se ha comentado en la sección 1.6.2.3, el método DA presenta la limitación de requerir más muestras que variables para que la matriz de covarianza pueda ser invertible; consecuentemente, se llevó a cabo un PCA con las muestras de las clases de *Araliaceae* previamente definidas (Tabla 4.1), con el fin de reducir el número de variables espectrales a unas pocas PCs. De esta manera, se utilizaron como variables de entrada las primeras 10 PCs del modelo de PCA que explican el 99.96% de la varianza. Se aplicó el método de inclusión de variables por etapas (stepwise) a fin de incluir en el modelo discriminante las componentes estadísticamente significativas; para tal fin, se utilizó el criterio de maximización de la distancia de Mahalanobis, ya que con él se considera la dirección y dispersión de los datos multivariantes [15]. Al final, fue obtenido un modelo con 9 PCs como variables independientes; la PC3 no fue significativo y no fue incluido en el modelo definitivo. La prueba de Box rechazó la igualdad de la covarianza para las

clases establecidas, de modo que, en la clasificación fueron utilizadas las matrices de covarianza por separado; es decir, las fronteras no fueron líneas rectas.

Se obtuvieron dos funciones discriminantes canónicas (CF) y la matriz de estructura sugirió la mayor correlación absoluta entre la variable PC1 y CF1; mientras que el resto de las variables hicieron lo mismo con CF2, siendo la variable PC4 la más importante. Como era de esperar, CF1 fue la principal función para la clasificación, su correlación canónica fue de 0.998 con una varianza explicada del 93.7%. El porcentaje de muestras correctamente clasificadas en el conjunto de calibración fue del 100%. Las muestras del conjunto de predicción fueron proyectadas sobre el modelo de PCA y posteriormente el modelo DA fue aplicado en los valores de score obtenidos. Los resultados obtenidos se muestran en la Tabla 4.3. Todas las muestras auténticas de *P. ginseng* fueron correctamente asignadas a su clase y la mayoría de las otras especies fueron debidamente excluidas de la clase PG; sin embargo, varias muestras de *L. meyenii* fueron asignadas erróneamente, lo que disminuyó el valor de la especificidad. Las mezclas con 5% y 10% de adulterante fueron mal clasificadas, excluyendo la mezcla 2, la cual probablemente pudo ser fácilmente discriminada por el modelo debido a que los espectros de *E. senticosus* son los más distintos a los de *P. ginseng*, como se observa en los scores del PCA global. El método DA obtuvo el valor más bajo de GSR de todos los modelos matemáticos ensayados, como resultado de la limitación en la detección de las mezclas, fueron confundidos casi la mitad de los espectros de las mezclas, y también por el porcentaje de especificidad obtenido.

#### **4.3.6. Análisis semicuantitativo de mezclas.**

El éxito en la aplicación del MCR-ALS requiere que los espectros de los componentes puros tengan baja correlación. Se seleccionó el *E. senticosus* para preparar las mezclas ya que fue la especie que presentó una mayor diferencia en su espectro con el *P. ginseng*, tal como fue mencionado en las secciones 4.3.2 y 4.3.5, para poner a prueba el procedimiento de semicuantificación propuesto.

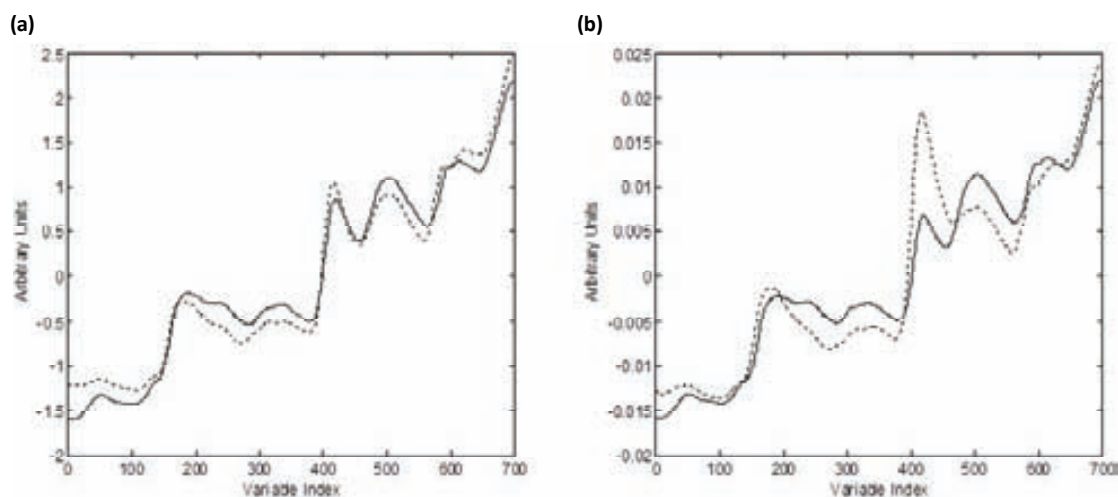


Como únicamente hay dos componentes en las mezclas, se utilizó un espectro promedio de *P. ginseng* puro y otro espectro promedio de *E. senticosus* puro como estimaciones iniciales para la resolución. En la primera etapa, se trabajó con un conjunto de espectros de mezclas simuladas, obtenidos por ordenador a partir de los espectros de los componentes, variando la composición de 5 a 5%, y se llegó a la conclusión de que las limitaciones (constraints) útiles y con sentido químico fueron "no negatividad", aplicada a los perfiles de las contribuciones (concentraciones), lo que implica que los componentes fueron siempre positivos, y "closure", se supone que el perfil de contribución de cada mezcla puede cumplir con un balance de masa constante igual a la unidad, representando el 100% de los componentes de la misma.

Utilizando estas restricciones, se obtuvo la contribución de *P. ginseng* en cada mezcla simulada y la diferencia entre los valores reales y los calculados por el modelo. El promedio de los valores absolutos de estas diferencias fue de 3.64%, sugiriendo que al añadir *E. senticosus* a una muestra de *P. ginseng*, el algoritmo no es capaz de distinguir cambios en la concentración nominal de *P. ginseng* en la muestra de  $\pm$  ese valor. Por consiguiente, las mezclas preparadas en el laboratorio fueron elaboradas presentando variaciones en la composición de *P. ginseng* de 10 en 10%.

Posteriormente, se concatenaron la matriz de las mezclas simuladas y la de las mezclas preparadas en el laboratorio, junto con varios espectros de *P. ginseng* puro, y se aplicó el algoritmo MCR-ALS con las restricciones enunciadas anteriormente. La convergencia fue alcanzada con 5 iteraciones y el modelo explicó el 99.9% de la varianza, presentando una diferencia entre los datos iniciales y los reproducidos por MCR-ALS (fitting error) de 0.932.

Las Figuras 4.2a y 4.2b muestran los espectros originales y los obtenidos después de la resolución de la mezcla, respectivamente. Se puede observar que los perfiles espectrales resultantes del MCR-ALS son buenas aproximaciones de los espectros originales; sin embargo, está presente una ambigüedad en la intensidad. La ambigüedad probablemente se debe a que los espectros originales no pudieron someterse a restricciones.



**Figura 4.2.** (a) Estimaciones iniciales empleadas para iniciar el algoritmo MCR-ALS y (b) perfiles obtenidos después de la resolución de las mezclas. Las líneas continuas representan al *P. ginseng* y las líneas discontinuas al *E. senticosus*.

Se calculó el porcentaje de error únicamente a partir de los perfiles de contribución (concentración) obtenidos de los replicados de las mezclas preparadas en el laboratorio. Los valores obtenidos se encontraban en un intervalo de 0.36 a 8.60% y el valor medio fue de 5.53%; lo que indica que los resultados de la contribución deberían ser expresados utilizando un intervalo de 10% (por ejemplo, si 84.92% es la contribución de *P. ginseng* obtenida en una mezcla por el algoritmo, sólo se puede decir que el resultado se encuentra entre 80 y 90%).

De acuerdo con los resultados observados para el *P. ginseng* puro, se puede considerar que una muestra no está mezclada con otra especie de planta si la resolución proporciona un valor de contribución igual o superior al 95%. Así, todos los resultados sugieren que el MCR-ALS es capaz de proporcionar una estimación rápida de los componentes en las mezclas de *P. ginseng* y *E. senticosus*.

#### 4.4. Conclusiones del capítulo.

La construcción de huellas dactilares NIR desarrollado en esta Tesis, representa una buena y rápida alternativa a los análisis tradicionales por CLAR para la diferenciación de *P. ginseng* de las otras ocho especies de plantas estudiadas.

Los tres métodos de clasificación aplicados, SIMCA, DA y PLS-DA, permitieron obtener una buena discriminación de *P. ginseng*, al conseguir igual valor de sensibilidad (100%). Sin embargo, las pruebas en mezclas preparadas en el laboratorio mostraron que SIMCA tiene una mejor capacidad para detectar las muestras adulteradas (MCC = 80%). Además, su especificidad (100%) fue mejor que la de los otros métodos de clasificación; todas estas características hicieron que SIMCA consiguiera el mejor rendimiento, como indicó el valor de GSR (95%).

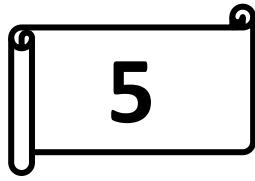
La espectroscopia NIR con MCR-ALS podría ser utilizada para realizar una determinación rápida y semicuantitativa de la composición de una mezcla, pero teniendo en cuenta que se han de aplicar las restricciones apropiadas y que las plantas “puras” deberán presentar espectros suficientemente diferentes para que puedan ser resueltos por el algoritmo. El porcentaje de error promedio fue de 5.53% y mostró que la composición de mezcla debe cambiar en cantidades de aproximadamente un 10% con el fin de obtener buenos resultados. Las ventajas de este método incluyen el no requerimiento de un conjunto de calibración o un método de referencia para determinar los componentes de la mezcla. Una estrategia de trabajo podría ser utilizar SIMCA para clasificar una muestra y, si es etiquetada como una mezcla, estimar su composición utilizando MCR-ALS. En el presente estudio, esta estrategia funcionó para mezclas de *P. ginseng* and *E. senticosus*.



## Referencias.

- (1) Bonfill, M.; Casals, I.; Palazón, J.; Mallol, A.; Morales, C. Improved high performance liquid chromatographic determination of ginsenosides in *Panax ginseng*-based pharmaceuticals using a diol column. *Biomedical Chromatography* 2002, 16, 68-72.
- (2) Kim, S. N.; Ha, Y. W.; Shin, H.; Son, S. H.; Wu, S. J.; Kim, Y. S. Simultaneous quantification of 14 ginsenosides in *Panax ginseng* C.A. Meyer (Korean red ginseng) by HPLC-ELSD and its application to quality control. *J. Pharm. Biomed. Anal.* 2007, 45, 164-170.
- (3) Woo, Y.; Kim, H.; Cho, J. Identification of Herbal Medicines Using Pattern Recognition Techniques with Near-Infrared Reflectance Spectra. *Microchemical Journal* 1999, 63, 61-70.
- (4) Chen, P.; Luthriau, D.; Harringtonohio, P. d. B.; Harnly, J. M. Discrimination Among *Panax* Species Using Spectral Fingerprinting. *Journal of AOAC International* 2011, 94, 1411-1421.
- (5) Yap, K. Y.; Chan, S. Y.; Lim, C. S. The reliability of traditional authentication – A case of ginseng misfit. *Food Chem.* 2008, 107, 570-575.
- (6) Zhang, Y.; Chen, J.; Lei, Y.; Zhou, Q.; Sun, S.; Noda, I. Evaluation of different grades of ginseng using Fourier-transform infrared and two-dimensional infrared correlation spectroscopy. *J. Mol. Struct.* 2010, 974, 94-102.
- (7) Ren, G.; Chen, F. Simultaneous Quantification of Ginsenosides in American Ginseng (*Panax quinquefolium*) Root Powder by Visible/Near-Infrared Reflectance Spectroscopy. *J. Agric. Food Chem.* 1999, 47, 2771-2775.
- (8) Lucio-Gutiérrez, J. R.; Coello, J.; Maspoch, S. Application of near infrared spectral fingerprinting and pattern recognition techniques for fast identification of *Eleutherococcus senticosus*. *Food Res. Int.* 2011, 44, 557-565.
- (9) Blackwell, R. Adverse events involving certain Chinese herbal medicines and the response of the profession. *Journal of Chinese medicine* 1996, 12.
- (10) Ren, G.; Chen, F. Determination of moisture content of ginseng by near infra-red reflectance spectroscopy. *Food Chem.* 1997, 60, 433-436.
- (11) Woo, Y.; Cho, C.; Kim, H.; Yang, J.; Seong, K. Classification of cultivation area of ginseng by near infrared spectroscopy and ICP-AES. *Microchemical Journal* 2002, 73, 299-306.
- (12) Yap, K. Y.; Chan, S. Y.; Lim, C. S. Infrared-based protocol for the identification and categorization of ginseng and its products. *Food Res. Int.* 2007, 40, 643-652.
- (13) Rinnan, Å.; Nørgaard, L.; Berg, F. v. d.; Thygesen, J.; Bro, R.; Engelsen, S. B. In *Chapter 2 - Data Pre-processing; Infrared Spectroscopy for Food Quality Analysis and Control*; Academic Press: San Diego, 2009; pp 29-50.
- (14) Jaumot, J.; Gargallo, R.; de Juan, A.; Tauler, R. A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB. *Chemometrics Intellig. Lab. Syst.* 2005, 76, 101-110.
- (15) Brereton, R. G. *Applied Chemometrics for Scientist*, John Wiley & Sons, Ltd, Great Britain, 2007.
- (16) Laasonen, M.; Harmia-Pulkkinen, T.; Simard, C. L.; Michiels, E.; Rasanen, M.; Vuorela, H. Fast identification of *Echinacea purpurea* dried roots using near-infrared spectroscopy. *Anal. Chem.* 2002, 74, 2493-2499.





***Valeriana officinalis.***

---





## 5.1. Introducción.

### 5.1.1. Descripción de la planta medicinal.

*Valeriana officinalis* pertenece al género *Valeriana* (familia *Valerianaceae*), compuesto por cerca de 250 especies con muchas más subespecies, varias de ellas referidas genéricamente como valeriana. Las raíces y los rizomas de las especies pertenecientes a este género se utilizan en la medicina tradicional de muchas culturas como un sedante suave [1]. Aunque han sido identificados más de 150 compuestos en los extractos de valeriana, no está del todo claro qué constituyentes son los responsables de sus efectos biológicos, lo que apunta a la posibilidad de una interacción sinérgica. A este respecto, sustancias como los esteres de terpenos (denominados valepotriatos), los baldriales y el ácido valerénico, así como los derivados de éste, son considerados como los más importantes. Sin embargo, estudios recientes plantean que los lignanos, flavonoides y el ácido clorogénico también desempeñan un papel relevante [2]. Además de *V. officinalis*, otras importantes especies comerciales son *Valeriana wallichii* (India) y *Valeriana edulis* (México), las cuales también se utilizan en este trabajo; su composición química difiere considerablemente provocando que las fitomedicinas preparadas a partir de ellas tengan una composición muy diversa [3,4].

### 5.1.2. Problemática con el control de calidad de *Valeriana officinalis*.

Considerando la disponibilidad comercial y el uso vago del nombre "valeriana", no es de extrañar que algunos estudios sugieran que sea común la adulteración de *V. officinalis* con otras especies de *Valeriana* [2,4]. Este tema adquiere relevancia porque sólo ha sido confirmada por la Agencia Europea del Medicamento (EMA) [5] la eficacia terapéutica de *V. officinalis* y también, porque algunos hechos experimentales plantean dudas respecto a la participación real de los componentes químicos de las diferentes valerianas con actividades similares [2]. Los métodos actuales para el control de calidad de *V. officinalis* se centran principalmente en la detección de la presencia sólo de valepotriatos y del ácido valerénico [2,3]; en tales condiciones,

podrían pasar inadvertidas las adulteraciones con varias especies de valeriana de importancia comercial que contienen otros componentes químicos además del ácido valerénico. Así, la práctica de identificar *V. officinalis* midiendo algunos pocos marcadores o componentes activos está lejos de ser satisfactoria. Además, generalmente no se considera el uso de la quimiometría para mejorar los resultados del control de calidad de la planta. Por lo tanto, todavía es un reto para los analistas de productos naturales la obtención de información útil para distinguir entre las especies de *Valeriana*.

### **5.1.3. Antecedentes.**

Tal como se ha mencionado en la sección 1.5.4, el método analítico más popular para la obtención de huellas dactilares de las plantas medicinales sigue siendo la cromatografía líquida de alto rendimiento acoplada a un detector de diodos en línea (CLAR-DAD) [6-8]. Sin embargo, la forma habitual de trabajar es utilizar un solo cromatograma a una longitud de onda fija (unidimensional); de este modo, no se consideran picos cromatográficos que pueden aparecer a otras longitudes de onda y que pueden ser útiles para la identificación de la especie estudiada. Han sido propuestas diferentes estrategias para utilizar más información del cromatograma; por ejemplo, las huellas dactilares bidimensionales desarrolladas por Yan, S., et al. [9] y el método de fusión utilizado por Fan, X., et al. [10] y por Ni, Y., et al [11]. No obstante, deben tenerse en cuenta algunas consideraciones cuando se utilizan estos métodos, debido a que pueden existir cambios de un cromatograma a otro [8,12]. El método bidimensional de obtención de huellas dactilares, presentado por Yan, S., et al., sólo considera el uso de la primera componente principal (PC1), posiblemente excluyendo datos relevantes, ya que la información útil puede ser dividida en varios componentes principales. El método de fusión pondera los vectores considerados, es decir, las respuestas en las longitudes de onda a lo largo del tiempo de análisis, mediante un coeficiente de combinación; los detalles del método pueden consultarse en [11]. Su aplicación a más de dos vectores puede ser compleja, ya que es necesario calcular varios coeficientes de combinación.

En la obtención de la huella dactilar de una planta por cromatografía, la forma y posición de los picos desempeñan un papel fundamental; por lo tanto, la estabilidad del sistema cromatográfico es importante para su correcto uso, ya que cualquier perturbación debido a cambios inesperados en las muestras analizadas o en el dispositivo de medida, podrían inducir a huellas dactilares alteradas. Así mismo, los datos deben ser corregidos de variaciones no deseadas, tales como cambios de la línea base o desplazamientos del tiempo de retención, ya que éstos influyen en gran medida en el resultado del análisis [13]. Sin embargo, ninguno de los enfoques descritos para aumentar la información tiene en cuenta la influencia de esas variaciones en la construcción de la huella dactilar o en la interpretación de los resultados. A partir de aquí, es evidente que aún son necesarias estrategias complementarias para la obtención de huellas dactilares de las plantas medicinales; especialmente para las muestras con perfiles cromatográficos complejos y para las mezclas de plantas.

#### **5.1.4. Descripción del estudio.**

En este trabajo se describe un tratamiento de la información que implica el uso de cromatogramas a varias longitudes de onda para la construcción de huellas dactilares aumentadas de plantas medicinales. Tomando los cromatogramas en un intervalo de longitudes de onda, se obtiene la matriz de datos de una muestra dada, con dimensiones (tiempo de retención x longitudes de onda); posteriormente, se calcula la transpuesta de la matriz y se seleccionan, tanto las longitudes de onda de trabajo como el intervalo o intervalos de tiempo de retención de interés para conseguir el mejor control de calidad posible. Entonces, los perfiles cromatográficos recuperados son sometidos a pretratamientos con el fin de eliminar las variaciones cromatográficas no deseadas y reordenados en un único vector de datos, que es la huella dactilar final (o ampliada) de la muestra. El conjunto de todas las huellas dactilares ampliadas es el que se utiliza para construir los modelos de exploración y de clasificación por medio de PCA (principal component analysis), SIMCA (soft independent modeling of class analogy) y PLS-DA (partial least squares discriminant analysis).

Dado que esta estrategia permite la inclusión de tantas longitudes de onda como seleccione el investigador, la mayoría de la información cromatográfica relevante puede ser considerada en la clasificación, lo que permite obtener mejores resultados que en el caso de utilizar una sola longitud de onda.

A fin de mostrar la utilidad del uso de cromatogramas con múltiples longitudes de onda, se ha aplicado este procedimiento mejorado de obtención de huellas dactilares a cromatogramas de *Valeriana officinalis*. Esta planta fue seleccionada debido a que hay todavía muchas dudas respecto a la identidad de las especies de *Valeriana*, que aumentan cuando se analizan mezclas de especies.

## **5.2. Metodología específica.**

### **5.2.1. Descripción de las muestras.**

Las raíces secas de *Valeriana officinalis* (VOF), *Valeriana wallichii* (VIN) y *Valeriana edulis* (VMX), y las hojas de *Passiflora incarnata* (PI), fueron obtenidas de Amóros Nature (Girona, Cataluña, España), Galke (Gittelde, Alemania), Richter (Ontario, Canadá), Desert Botanicals (Gilbert, AZ, EE.UU.), Frontier Natural Products Co-op (Noruega, IA, EE.UU.), MaxNature (Los Angeles, CA, EE.UU.), Modern Jungle (EE.UU.) y herboristerías de las áreas metropolitanas de Barcelona (España), Oporto (Portugal) y Monterrey (México). Varios proveedores poseen una certificación de calidad (p. ej. ISO, GMP o de otro tipo) e identifican las plantas de sus propios cultivos mediante estudios botánicos; de este modo, son capaces de autenticar sus productos y proporcionar un comprobante de identidad. En la Tabla 5.1 se presenta un resumen de las muestras en estudio. Con el fin de incluir la mayor variabilidad posible en las muestras, aparte de su origen geográfico, todas ellas fueron recolectadas durante los años 2007 a 2009; por consiguiente, el número de lotes corresponde al número de muestras de cada especie de valeriana, porque todas las muestras provienen de diferentes partidas, proveedores o años. Las muestras de valeriana fueron almacenadas protegidas de la luz y la humedad a temperatura de laboratorio (25 °C aprox.) y fueron analizadas dentro de la fecha de caducidad, dada por el fabricante.

**Tabla 5.1.** Características de las muestras y conjuntos de objetos.

Muestras			Lotes				Número of objetos en		
ID <sup>a</sup>	Origen <sup>b</sup>	PT <sup>c</sup>	QS <sup>d</sup>	Número total <sup>e</sup>	NC <sup>f</sup>	NP <sup>f</sup>	PCA <sup>g</sup>	SIMCA, PLS-DA	
								CS <sup>h</sup>	PS <sup>h</sup>
VOF	BG, CA, DE, ES, HR, MX, PT, US	RHR	5	29	15	5	87	45	42
VIN	DE, IN	RHR	2	2	0	2	6	0	6
VMX	DE, MX	RHR	1	13	7	3	39	21	18
PI	ES	RHL	0	1	0	0	1	0	1
Mezclas <sup>i</sup>	---	---	---	---	---	---	9	0	9

<sup>a</sup> Identidad: VOF, *V. officinalis*; VIN, *V. wallichii*; VMX, *V. edulis* y PI, *P. incarnata*.

<sup>b</sup> BG, Bulgaria; CA, Canadá; HR, Croacia; DE, Alemania; IN, India; MX, México; PT, Portugal; ES, España y US, Estados Unidos.

<sup>c</sup> Tipo de producto, RHR, raíces sin procesar; RHL, hojas sin procesar.

<sup>d</sup> Proveedores de calidad; número de proveedores con una certificación de calidad.

<sup>e</sup> Diferentes lotes de plantas adquiridos de varios proveedores y herboristerías en tres años .

<sup>f</sup> Número de muestras en el conjunto de calibración (NC) o en el conjunto de prueba (NP) procedentes de proveedores con una certificación de calidad.

<sup>g</sup> Número de huellas dactilares aumentadas.

<sup>h</sup> Número de huellas dactilares (a una única longitud de onda o con múltiples longitudes de onda) empleadas en el conjunto de calibración (CS) o en el conjunto de predicción (PS) para la construcción y evaluación de los modelos matemáticos.

<sup>i</sup> ---, No aplica.

## 5.2.2. Reactivos.

Acetonitrilo y metanol grado HPLC (ambos Chromasolv) fueron adquiridos a Sigma-Aldrich (Steinheim, Alemania). El ácido fosfórico (85%, Biochemika) y bifenilo calidad GC, utilizado como marcador externo para la corrección de los desplazamientos del tiempo de retención, se obtuvieron de Fluka (Buchs, Suiza). El agua para la fase móvil cromatográfica y para la preparación de los extractos de valeriana se purificó en un sistema Milli-Q (Millipore, Madrid, España). Etanol grado HPLC (96% v/v) fue adquirido a Scharlau (Sentmenat, España).

### 5.2.3. Preparación de las muestras y de las mezclas de laboratorio.

Las muestras de valeriana analizadas recibidas como raíces enteras o en forma de raíces troceadas, así como las hojas de *P. incarnata*, fueron molidas con un molinillo Turmix Mill (Barcelona, España) durante 1.5 minutos aproximadamente; las muestras recibidas en forma de polvo se utilizaron directamente. De catorce lotes de *V. officinalis* disponibles en el conjunto de predicción, nueve fueron seleccionados al azar para preparar nueve mezclas en el laboratorio. Con este propósito, varias cantidades de *V. wallichii*, *V. edulis* y *P. incarnata*, procedentes de plantas distintas, fueron mezcladas con *V. officinalis*, originando mezclas conteniendo 5%, 10% o 20% del adulterante seleccionado.

Para cada muestra se pesó con precisión 1.00 g del polvo y se añadieron 100  $\mu$ L de bifenilo disuelto en metanol (2 mg ml<sup>-1</sup>). Después de un mezclado manual y un período de reposo de 10 minutos, se adicionaron 3 mL de una solución de metanol - agua (9:1) a cada muestra. Se colocaron en un baño de ultrasonidos 15 minutos para posteriormente centrifugarlas durante 10 min a 3500 rpm. El sobrenadante fue transferido a un matraz aforado de 10 mL. El procedimiento se repitió dos veces más, y los respectivos sobrenadantes fueron combinados. El volumen final se enrasó a 10 mL con la solución de metanol – agua (9:1). Antes de ser inyectadas en el cromatógrafo, todas las soluciones obtenidas fueron filtradas utilizando un filtro de membrana de nylon con tamaño de poro de 0.45  $\mu$ m. Los extractos fueron sometidos a análisis por CLAR en las 12 horas siguientes de su preparación para mantener la estabilidad y la integridad de su contenido. Se preparó un replicado de cada muestra durante tres días no consecutivos (aunque las mezclas preparadas en el laboratorio fueron únicas), utilizando cada día una porción de muestra procesada de forma independiente.

### 5.2.4. Instrumentación y procedimiento cromatográfico.

Los análisis por CLAR se realizaron con un sistema HPLC-DAD de Agilent serie 1100, equipado con un desgasificador de vacío G1379A, una bomba binaria G1312A, un muestreador automático con un inyector ajustable de 100  $\mu$ L de loop G1313A, un

compartimiento de columna con temperatura controlada G1316A, un detector de diodos en línea (diode array) G1315B y un ordenador equipado con el programa de control ChemStation para LC 3D (Rev. B.01.03). La separación fue realizada utilizando una columna Agilent Zorbax Eclipse XDB-C18; 4.6 mm x 250 mm, 5 µm de tamaño de partícula operada a 35 °C.

La fase móvil consistió en (A) agua y (B) acetonitrilo-metanol (1 + 1), ambas fases con 0.05% de H<sub>3</sub>PO<sub>4</sub>. La secuencia del gradiente de elución fue establecida como sigue: inicial a 5% de B; entonces, el porcentaje de B se incrementó de 5 hasta 17% de 0 a los 10 min, de 17 a 50% de los 10 a 25 min y, finalmente, de 50 hasta 100% de los 25 a 45 min. Cada análisis fue seguido por un lavado de 5 minutos con 100% de B y a continuación el sistema fue restablecido a las condiciones iniciales (95% A, 5% B) durante 10 min. El flujo de la fase móvil se mantuvo a 1.2 mL min<sup>-1</sup> y el volumen de inyección fue de 20 µL. Los espectros UV-Vis fueron registrados en el intervalo de 200 a 500 nm cada 2 nm, estableciendo el tiempo de respuesta en 2 s (tiempo de muestreo del detector de 0.0067 min).

#### **5.2.5. Análisis multivariante.**

Los datos CLAR-DAD obtenidos de los triplicados de 44 muestras de valeriana, las 9 mezclas preparadas y un replicado de una muestra de PI, un total de 142 cromatogramas, se exportaron desde el programa de control del cromatógrafo en formato CVS y fueron importados a MATLAB 7.9 (The MathWorks, Inc); que fue utilizado para el manejo de las matrices de datos cromatográficos. Utilizando el programa PLS-Toolbox 5.5 (Eigenvector Research), se realizó la co-adición (combinación de datos adyacentes), la corrección de la línea base, la normalización, el desdoblamiento de la matriz de datos y la construcción de los modelos matemáticos. Los códigos para Matlab desarrollados por Skov [14] y por Tomasi [15], que están disponibles libremente en [www.models.kvl.dk](http://www.models.kvl.dk), fueron empleados para la alineación de los cromatogramas mediante COW (correlation optimized warping).

Se llevó a cabo un PCA para decidir cuál era el método de alineación que proporcionaba mejores resultados (muestras más agrupadas). Se utilizó el gráfico de scores para observar las pautas de las muestras, las agrupaciones, las similitudes, diferencias y muestras anómalas.

El método SIMCA se utilizó para clasificar las muestras en dos clases, *V. officinalis* y *V. edulis*. El resto de las especies utilizadas y las mezclas de laboratorio fueron incluidas siempre en el conjunto de predicción. Los límites para el espacio definido por el modelo de *V. officinalis* ( $T^2$ ) y su espacio residual (Q) fueron establecidos a un nivel de confianza del 95%; las muestras calculadas dentro de los límites fueron clasificadas como miembros de la clase.

En la construcción del modelo PLS-DA, la matriz X se compone de las huellas dactilares aumentadas y el vector Y es un conjunto de variables categóricas que indica a qué clase pertenecen los objetos. Las variables categóricas se asignaron como 1 a *V. officinalis* y 0 a *V. edulis*.

Para construir los modelos de calibración de PCA y PLS-DA se utilizó la validación cruzada (CV), empleando el método de bloques contiguos con 22 divisiones (los replicados de muestra son eliminados o introducidos juntos en el proceso de construcción del modelo). El número óptimo de factores se determinó con el criterio del valor mínimo de PRESS (prediction sum of squares) en la validación cruzada [16]. Los estadísticos evaluados en la construcción de los modelos de calibración incluyeron RMSEC (root mean square error of calibration), RMSECV (root mean square error of cross validation) y el coeficiente de determinación  $r^2$ . Se compararon los resultados obtenidos con SIMCA y PLS-DA en términos de sensibilidad, especificidad, de tasa general de éxito en la clasificación (GSR) y de porcentaje de mezclas clasificadas correctamente (MCC). La sensibilidad es la proporción de muestras que pertenecen a la clase *V. officinalis* que son correctamente identificadas por el modelo matemático, mientras que la especificidad es la proporción de muestras que no pertenecen a la clase *V. officinalis* (es decir, VIN, VMX, PI y mezclas), que son correctamente identificados por el modelo como extrañas. Estos dos criterios son similares a los



propuestos por C. Lau, et al [17]. El valor de GSR se obtiene calculando el porcentaje global de muestras predichas que fueron correctamente clasificadas con respecto a la clase *V. officinalis*. El valor de MCC es útil para comparar la capacidad de los modelos para discriminar muestras adulteradas; se ve afectado por la planta utilizada para preparar las mezclas y la cantidad mínima de adulterante discriminado por el modelo.

De forma paralela, se construyeron modelos matemáticos con los datos pretratados de huellas dactilares a una sola longitud de onda (226 nm) y se compararon los resultados obtenidos.

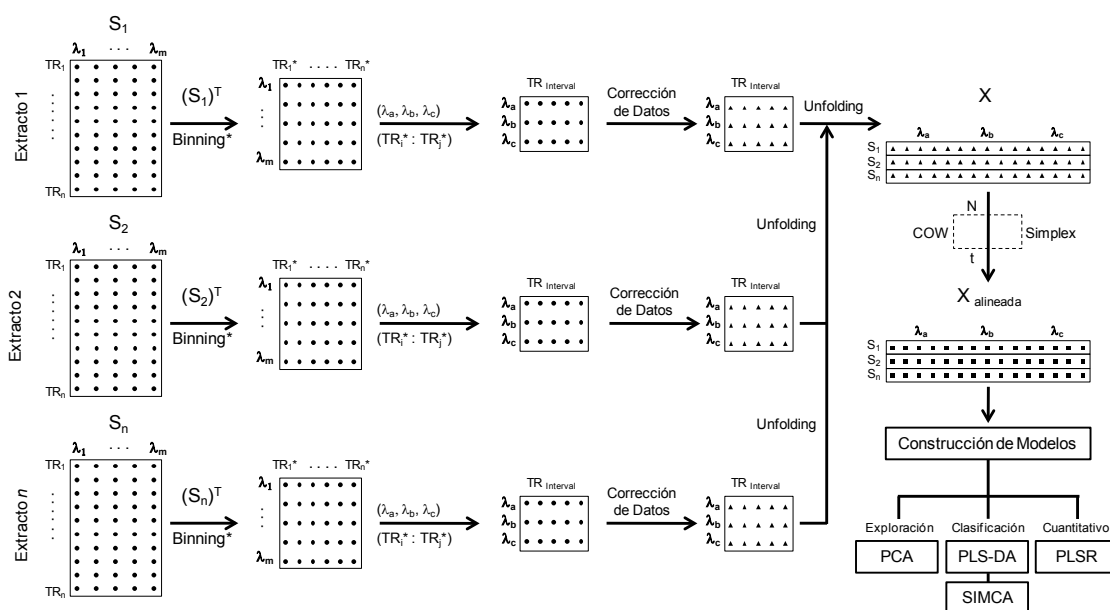
#### **5.2.6. Construcción de las huellas dactilares con múltiples longitudes de onda.**

En la Figura 5.1 se representa el procedimiento general para la construcción de las huellas dactilares aumentadas para los cromatogramas de plantas.

Para cada muestra se obtuvo una matriz de datos cromatográficos ( $S_n$ ) de dimensiones 7643 x 151 (tiempo de retención x longitudes de onda); a continuación, se calculó la transpuesta de la matriz ( $S_n^T$ ) y la dimensión del tiempo de retención fue co-adicionada, es decir, las variables adyacentes se combinaron, utilizando el valor promedio de grupos de 5 variables, obteniendo una matriz de (151 x 1529).

##### **5.2.6.1. Reducción de variables.**

Fueron seleccionadas las longitudes de onda y el intervalo de tiempo de retención que mostraron la mayoría de los picos. Durante la optimización del método cromatográfico, se observó que las respuestas más intensas estuvieron en cuatro longitudes de onda: 226, 254, 280 y 326 nm; por lo que los cromatogramas de estas longitudes de onda se utilizaron para construir la huella dactilar aumentada de cada muestra. El intervalo de tiempo de retención seleccionado fue de los 10 a los 44 minutos, ya que todos los picos más intensos se encontraban en este rango y se evitó la inclusión en los modelos matemáticos de ruido o señales no informativas. De aquí que las dimensiones de la matriz se redujeran a 4 x 1020.



**Figura 5.1.** Esquema de la construcción de las huellas dactilares aumentadas.  $S_1, S_2, \dots, S_n$ ; representan los datos cromatográficos originales en dos dimensiones (tiempo de retención [TR] x longitudes de onda) de los diferentes extractos de plantas. X es la nueva matriz de datos construida con huellas dactilares aumentadas. COW, N y t son el acrónimo y los símbolos para correlation optimized warping, longitud de segmento y el parámetro slack, respectivamente.  $X_{alineada}$  es la matriz de datos final que se utiliza en la construcción de los modelos matemáticos.

### 5.2.6.2. Ajuste de la línea de base y normalización de los datos.

Las variaciones en la línea base fueron minimizadas por medio del método de mínimos cuadrados asimétricos ponderados (WLS) utilizando un polinomio de tercer grado; se pueden encontrar más detalles sobre el método en la sección 1.6.1.5 y en [18]. La variabilidad en la escala entre las muestras fue minimizada normalizando cada vector de datos; es decir, en la matriz con la línea base corregida fue normalizado cada cromatograma, empleando la suma del cuadrado de los valores de todas las variables para un dado vector de datos, obteniendo una matriz con vectores de longitud unidad.

### 5.2.6.3. Desdoblamiento de la matriz (unfolding).

La matriz corregida fue desdoblada de tal manera que los vectores de datos para cada longitud de onda fueron colocados uno tras otro para construir un único vector de datos con dimensiones de 1 x 4080; que es la huella dactilar aumentada. Finalmente, las huellas dactilares aumentadas de todas las muestras se ordenaron en

una matriz (**X**) para la subsiguiente construcción de los modelos de exploración o clasificación.

### **5.2.7. Optimización de los parámetros para COW (correlation optimized warping).**

Antes de la construcción de la matriz final de huellas dactilares aumentadas (**X**), los desplazamientos del tiempo de retención deben de ser corregidos. El algoritmo COW fue utilizado para este propósito, mediante la ejecución de los códigos de Matlab mencionados en la sección 5.2.5. Se seleccionó, a partir de las huellas dactilares aumentadas de muestras autenticadas de *V. officinalis*, un vector de datos que fue usado como referencia, y los otros cromatogramas se alinearon con respecto es éste. Se utilizó como criterio el máximo acumulado del producto de los coeficientes de correlación. La teoría de los algoritmos que se han utilizado en la selección del cromatograma de referencia y la optimización del COW puede ser consultada en [14].

Los valores de la longitud del segmento (**N**) y el grado de flexibilidad (**t**) fueron determinados por medio de un diseño simplex, realizado dentro de un espacio de optimización definido: longitud del segmento de 25 a 250; flexibilidad de 2 a 20. Estos valores fueron elegidos de acuerdo con las anchuras de los picos y los desplazamientos observados en los cromatogramas. El número de puntos de inicio en la red de búsqueda se ajustó a 3, el número máximo de pasos de optimización a 25 y la fracción de la máxima desviación del centro en la alineación COW al 15%.

## **5.3. Resultados y discusión.**

### **5.3.1. Optimización y validación de la metodología.**

Los parámetros de CLAR-DAD han sido optimizados en función de las condiciones descritas en la literatura [2,3,19,20]. Fueron evaluadas varias composiciones de fase móvil constituidas por agua - metanol o agua y metanol - acetonitrilo, separando una primera serie de extractos de una muestra autenticada de *V. officinalis*. Se utilizaron caudales de 0.8, 1.2 y 1.5 mL min<sup>-1</sup> y temperaturas de

columna de 25, 30 y 35 °C. Se ensayaron dos volúmenes de inyección de 10 y 20 µL; fue seleccionado el segundo ya que se mejoró la respuesta sin aumentar excesivamente la anchura de pico. Una buena separación y adecuadas condiciones de trabajo fueron logradas empleando las condiciones presentadas en la sección 5.2.4.

Las respuestas de los componentes químicos detectables de las valerianas en estudio aparecen, principalmente, en la región espectral de 200 a 400 nm. La mayoría de los eluatos presentaron una absorbancia máxima en alguna de las siguientes longitudes de onda: 226, 254, 280 y 326 nm. Esta observación se corresponde con el conocido hecho de que las principales familias de compuestos presentes en *Valerianaceae* (por ejemplo, valepotriatos, lignanos, flavonoides y el ácido clorogénico) tienen la mayor respuesta en estas longitudes de onda. Debido a ello, fueron las longitudes de onda seleccionadas para realizar el estudio. El resto de la optimización se realizó a 226 nm, ya que todos los grupos de picos cromatográficos exhibieron respuesta en esta longitud de onda.

En la extracción se seleccionó metanol (MeOH) frente a (EtOH) al conseguir picos cromatográficos con una mejor simetría. El rendimiento de la extracción mejoró mediante la adición de agua, ensayando soluciones de MeOH-H<sub>2</sub>O (9:1) y MeOH-H<sub>2</sub>O (8:2). La primera composición proporcionó picos con una forma más simétrica, por lo que fue el medio elegido.

Con el fin de evaluar la estabilidad de la metodología utilizada, se determinaron la repetibilidad y la precisión intermedia. Se obtuvo la repetibilidad del método mediante el análisis, en un mismo día, de cuatro porciones independientes de una muestra de *V. officinalis*. Para cada cromatograma se registraron las áreas de 14 picos, que abarcaban toda la región cromatográfica de interés, y fue calculada la desviación estándar relativa (RSD) para cada conjunto de picos. Los valores de RSD para las áreas fueron menores del 4.61%. Para los tiempos de retención, el valor mayor de RSD obtenido fue de 0.12%, sugiriendo que no se tendrían cambios de importancia en el tiempo de retención. Desafortunadamente, en las etapas posteriores del trabajo

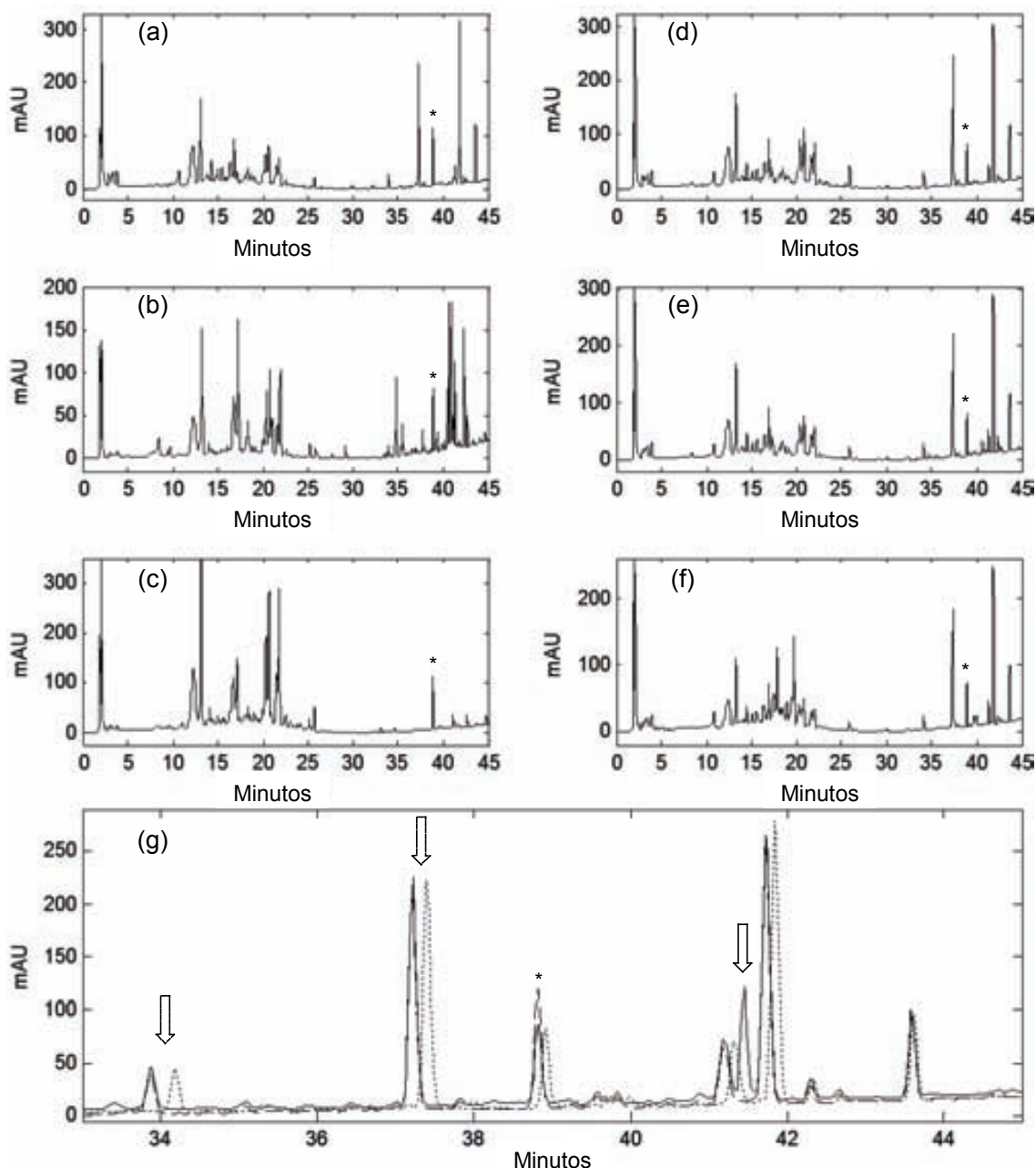
aparecieron ligeros cambios en el tiempo de retención, probablemente debido a cambios menores en la composición de la fase móvil.

Puesto que la composición química de las plantas en estudio puede variar considerablemente, no es posible determinar la precisión intermedia del método a partir de los compuestos endógenos que se producen de forma natural en las muestras, de modo que para determinar este parámetro sólo se utilizó el marcador externo. A partir de los replicados de 8 muestras, procesadas y analizadas en tres días diferentes se obtuvieron las áreas de los picos del bifenilo, presentando un RSD de 7.37%. Así, todos los resultados indican una buena estabilidad del procedimiento cromatográfico utilizado.

### **5.3.2. Características de los cromatogramas y de las huellas dactilares aumentadas.**

Los cromatogramas obtenidos fueron complejos, mostrando más de 40 picos y es evidente que algunos de ellos no están completamente resueltos, Figura 5.2. A pesar del control riguroso de las condiciones de separación, aparecieron variaciones cromatográficas habituales en el análisis por CLAR de extractos de plantas: cambios en la línea base, cambios en el tiempo de retención y diferente número de picos entre los replicados de las mismas muestras. Estas variaciones están relacionadas con la incertidumbre en la composición del extracto, los cambios inesperados en las muestras analizadas o las perturbaciones introducidas por el mismo dispositivo de medición [8,12]. Esta complejidad representa un reto para la correcta identificación de las especies de plantas y en la construcción de los modelos de clasificación.

Los conjuntos de picos se distribuyen en dos regiones del cromatograma; la primera se encuentra entre los tiempos de retención de 10 a 27 min y la otra de 33 a 44 min. De acuerdo con la literatura [2,3,19,20], se conoce que los principales constituyentes presentes en el primer conjunto cromatográfico son el ácido clorogénico, lignanos y flavonoides; mientras que el segundo grupo de picos consiste principalmente en ácido valerénico y sus derivados (sesquiterpenos) y valepotriatos (por ejemplo, valtrato, didrovaltrato, acevaltrato).



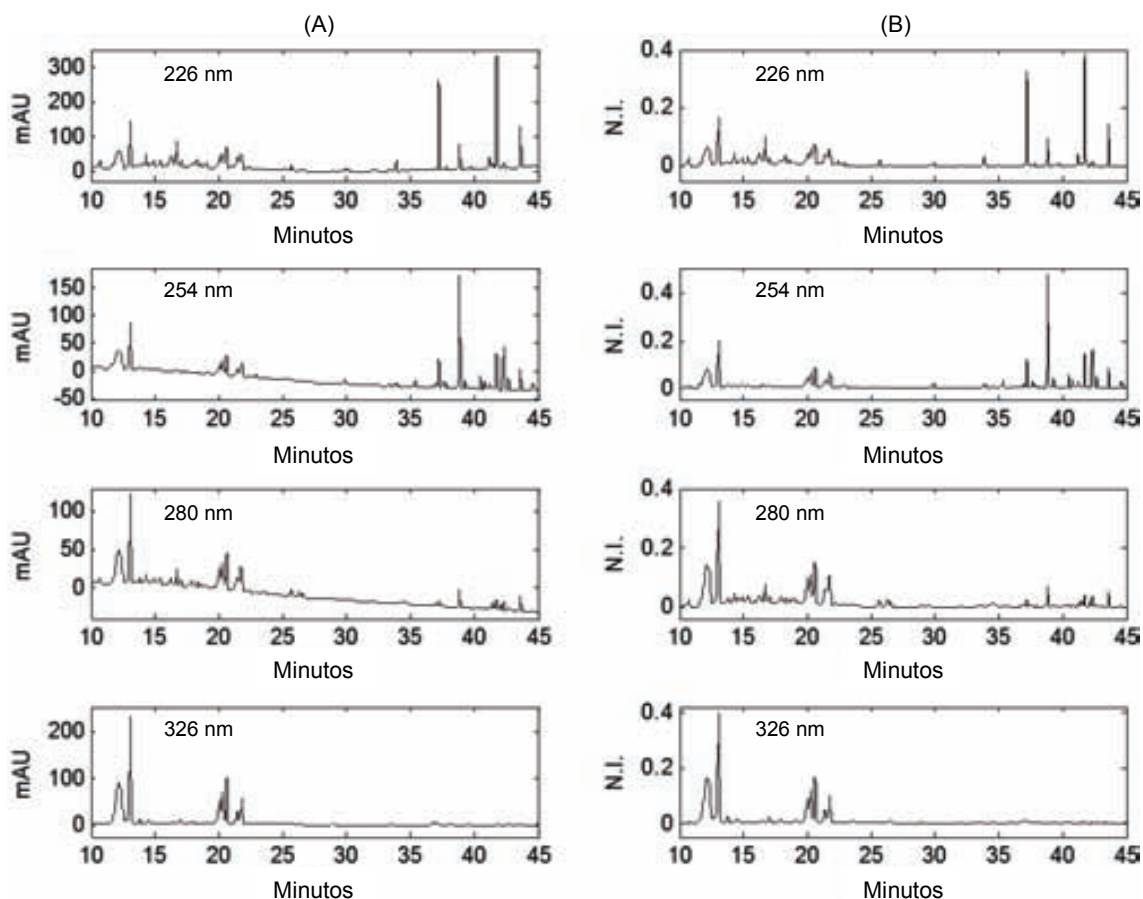
**Figura 5.2.** Comparación de los cromatogramas a 226 nm de las diferentes muestras estudiadas. (a) *V. officinalis*, VOF; (b) *V. edulis*, VMX; (c) *V. wallichii*, VIN; (d) mixture VOF - VIN; (e) mezcla VOF - VMX; (f) mezcla VOF - *P. incarnata*, (todas las mezclas antes mencionadas están preparadas en una proporción de 9:1) y (g) sección de los cromatogramas de tres muestras diferentes de VOF. El asterisco indica la posición del pico del bifenilo; la flecha de línea continua indica la presencia de un pico inesperado en una de las muestras y las flechas de líneas discontinuas muestran algunos ejemplos del desplazamiento en el tiempo de retención.

Además, se puede notar que el contenido de sesquiterpenos y valepotriatos en *V. edulis* (Fig. 5.2b) y *V. wallichii* (Fig. 5.2c) difieren en gran medida de *V. officinalis* (Fig. 5.2a). Sin embargo, la utilidad de esta diferencia para la identificación de las especies de valeriana es limitado, ya que el contenido químico en muestras de plantas puede

variar considerablemente y porque no es de ninguna manera fácil asignar la identidad de la valeriana por inspección visual de los cromatogramas, sobre todo empleando una única longitud de onda y tratándose de mezclas; tal como se puede observar en las Figuras 5.2d-f. Esta tarea es aún más complicada debido a la presencia de picos inesperados, lo cual ocurrió espontáneamente en algunos extractos, y también debido a los cambios del tiempo de retención, como se muestra en la Fig. 5.2g. Bajo estas circunstancias, pueden ocurrir identificaciones erróneas de las muestras adulteradas no sólo al utilizar otras especies de valeriana en cantidades del 20%, sino también cuando se utiliza una hierba completamente ajena (género y especie diferentes) como *P. incarnata*.

La Figura 5.3a muestra las variaciones presentadas por los cromatogramas de una muestra *V. officinalis* en las longitudes de onda seleccionadas que fueron utilizadas para construir las huellas dactilares aumentadas. La más evidente de estas variaciones fue el cambio en la pendiente de la línea base, presente en unos pocos cromatogramas, pero que se produjo en todas las especies de valeriana en estudio. Hubo una ligera tendencia hacia valores positivos a 226 nm y una tendencia importante a los valores negativos a 254 y 280 nm. En tales condiciones, no era viable desdoblarse la matriz de datos cromatográficos para construir el vector de la huella dactilar aumentada; de modo que, los datos tuvieron que ser corregidos.

La Fig. 5.3b muestra el resultado de la aplicación de los pretratamientos a los vectores de la matriz de datos cromatográficos. El orden en que se utilizaron fue el siguiente: combinación de datos adyacentes, corrección de la línea base y posterior normalización. La combinación de datos adyacentes mantuvo la información cromatográfica, es decir, no afectó las características del cromatograma, y redujo el número de datos en la dimensión del tiempo de retención de la matriz. La corrección de la línea base eliminó, con éxito, las tendencias exhibidas por la señal de fondo de los cromatogramas. Finalmente, la normalización de los vectores cromatográficos hizo posible reducir la influencia ejercida por la absortividad de los constituyentes químicos de las muestras; también evita obtener resultados de clasificación sesgados hacia la variabilidad generada por la concentración total de la muestra.



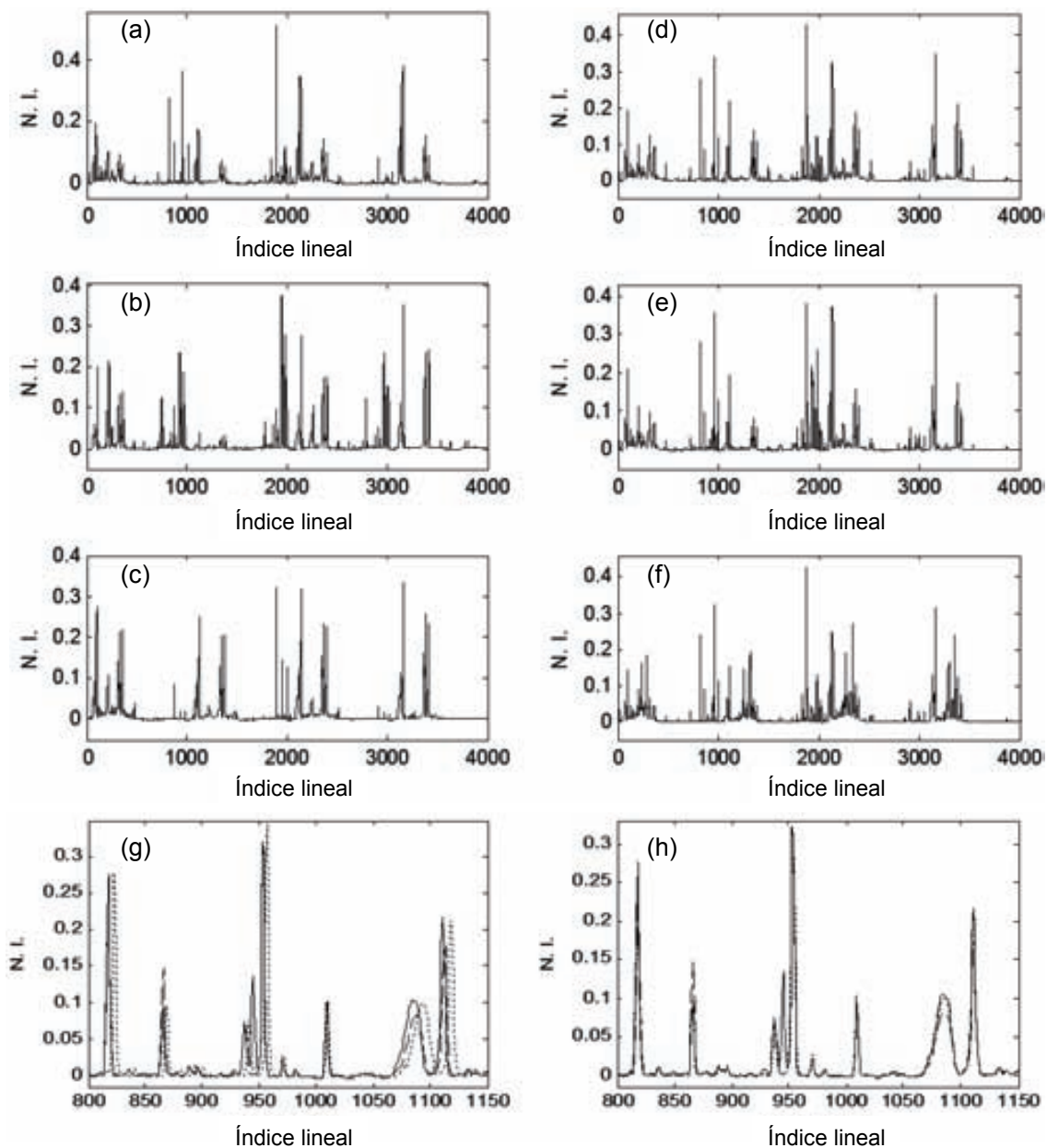
**Figura 5.3.** Cromatogramas de una muestra de *V. officinalis* a las longitudes de onda utilizadas en la construcción de las huellas dactilares aumentadas. (A) Antes de los pretratamientos y (B) después de la combinación de datos, corrección de línea base y normalización; N.I. se refiere a valores normalizados de intensidad.

Las matrices de datos corregidas fueron desdobladas a lo largo de la dimensión del tiempo de retención para generar las huellas dactilares con múltiples longitudes de onda; Figura 5.4. Como resultado del desdoblamiento de la matriz, se perdió la dimensión del tiempo de retención y esta tuvo que ser reemplazada con una nueva dimensión mediante la creación de un índice lineal; no obstante, las unidades del índice lineal pueden ser traducidas a las unidades del tiempo de retención.

Como era de esperar, las huellas dactilares aumentadas presentaron más conjuntos de picos que los exhibidos al usar una sola longitud de onda, permitiendo una mejor y más simple comparación entre las señales cromatográficas y haciendo más sencilla la identificación de las especies de valeriana. Así, las huellas dactilares aumentadas representan un incremento en la información proporcionada por la variabilidad química de la muestra, no sólo debido al incremento en el número de



picos, sino también debido a la minimización de la variabilidad no informativa producida por otros factores, por ejemplo el dispositivo de medición, como resultado del pretratamiento de los datos.



**Figura 5.4.** Comparación de las huellas dactilares aumentadas de (a) *V. officinalis*, VOF; (b) *V. edulis*, VMX; (c) *V. wallichii*, VIN; (d) mezcla VOF - VIN; (e) mezcla VOF - VMX; (f) mezcla VOF - *P. incarnata*, mezclas preparadas en proporción 9:1; (g) sección de las huellas dactilares aumentadas de tres muestras de VOF antes de aplicar correlation optimized warping y (h) las mismas muestras después de la alineación. N. I., valores normalizados de intensidad.

### 5.3.3. COW (correlation optimized warping) dirigida mediante PCA (principal component analysis).

La corrección de los desplazamientos del tiempo de retención fue el último pretratamiento de datos realizado. La capacidad del algoritmo para realizar correcciones simultáneas con cromatogramas obtenidos a varias longitudes de onda no había sido probada. Como parámetros de entrada se utilizó la huella dactilar aumentada utilizada como referencia, la longitud del segmento y el grado de flexibilidad (slack size). Utilizando el algoritmo en quince huellas dactilares aumentadas, obtenidas a partir de material autenticado de *V. officinalis*, se determinó como vector de referencia la muestra que presentó el mayor parecido con todas las huellas dactilares aumentadas en consideración. La optimización simplex proporcionó una longitud de segmento de 24 y un grado de flexibilidad de 4. La Fig. 5.4g y la Fig. 5.4h presentan los resultados obtenidos con estos parámetros de alineación.

Los cambios del tiempo de retención fueron corregidos adecuadamente, a pesar de la presencia de picos no comunes entre las réplicas, por ejemplo, el pico próximo al valor 940. Además, los picos y las características no fueron distorsionados por el uso del algoritmo COW. Sin embargo, la aplicación de estos parámetros para alinear otras huellas dactilares aumentadas diferentes a la de *V. officinalis* (una posibilidad fácil de aparecer en análisis de rutina con muestras desconocidas) podría producir alteraciones en las formas de los picos.

Se utilizó el PCA para evaluar la forma de aplicar la alineación a cada huella dactilar aumentada construida. Se evaluaron tres alternativas: ninguna alineación, la alineación sobre el triplicado de la muestra (es decir, una de las replicas se utiliza como referencia para alinear las otras dos) y el uso de una referencia seleccionada para alinear cada muestra analizada. Los resultados obtenidos se presentan en la Tabla 5.2.

**Tabla 5.2.** Opciones para la alienación evaluadas por PCA. <sup>a</sup>

Acción	PCs <sup>b</sup>	EV (%) <sup>c</sup>	RMSECV <sup>d</sup>
Ninguna alineación	14	90.25	0.72
Alineación de los triplicados de muestra	12	92.16	0.63
Alineación con única referencia	12	87.28	0.70

<sup>a</sup> Para la construcción del modelo PCA se realizó el centrado de los datos y la validación cruzada, empleando 30 divisiones aleatorias para los subconjuntos y 5 iteraciones; se incluyeron las 142 huellas dactilares aumentadas.

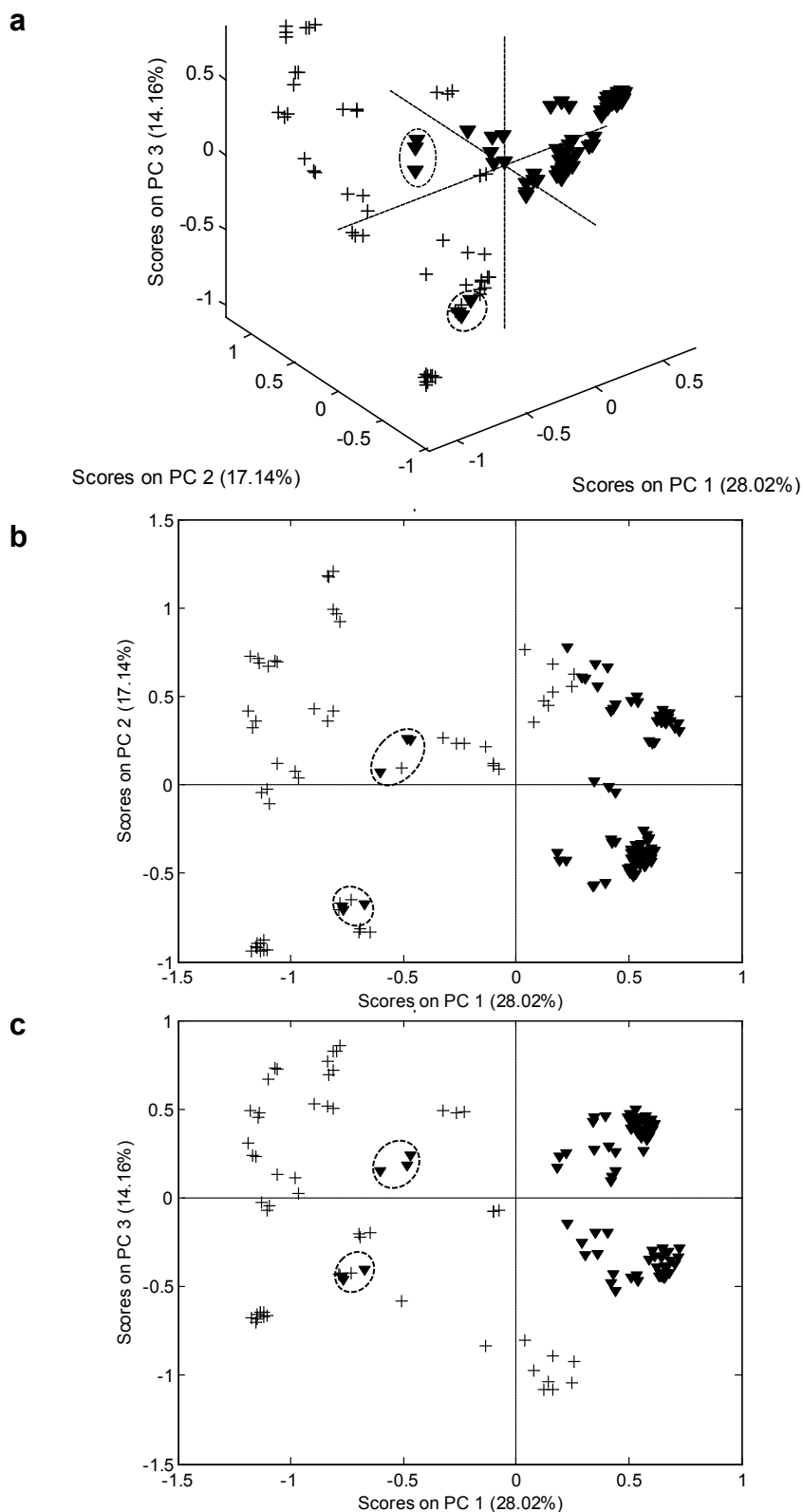
<sup>b</sup> Número de componentes principales usadas en el modelo PCA, seleccionadas con el criterio del mínimo PRESS.

<sup>c</sup> Varianza explicada acumulada.

<sup>d</sup> Raíz del error cuadrático medio de la validación cruzada.

La alineación sobre triplicados produjo el modelo de PCA más simple, con la mayor varianza explicada y el menor valor de RMSECV; indicando que se elimina la variabilidad no informativa de la matriz de huellas dactilares aumentadas. A pesar de que era evidente que aún queda una cierta dispersión de las muestras de *V. officinalis* en el gráfico de scores, esto no influirá en los resultados de la clasificación debido a que la variabilidad química entre las especies de plantas estudiadas es mayor que la variabilidad producida por los desplazamientos que quedan entre los triplicados corregidos por COW; es decir, esta dispersión se debe principalmente a diferencias en la información química en cada triplicado en lugar de cambios del tiempo de retención entre los triplicados, véase la Figura 5.5.

El gráfico de scores reveló dos triplicados sospechosos de *V. officinalis* situados cerca de los grupos de las "otras plantas". Ya que los triplicados estuvieron bien descritos por el modelo de PCA construido, es muy probable que estas muestras hayan sido etiquetadas erróneamente como *V. officinalis* en las herboristerías. Por lo tanto, las muestras sospechosas no fueron utilizadas en el conjunto de calibración y fueron marcadas como otras plantas para fines de la clasificación.



**Figura 5.5.** (▼) *V. officinalis*; (+) Otras plantas. (a) Gráfico 3D de los scores del PCA global de las huellas dactilares aumentadas después de la alineación. (b) Gráfico 2D de los scores en PC1 vs. PC2 y (c) PC1 vs. PC3. Aquí, "otras plantas" se refiere a *V. wallichii*, *V. edulis*, *P. incarnata* y las mezclas. Los triplicados sospechosos de muestras de *V. officinalis* han sido señalados con un círculo.

#### 5.3.4. Comparación de los resultados de los modelos de clasificación.

Los modelos SIMCA y PLS-DA se construyeron utilizando 66 objetos entre muestras de *V. officinalis* y de *V. edulis*, véase la Tabla 5.1. El conjunto de predicción constaba de 76 objetos. En ningún caso fueron detectadas muestras anómalas durante la construcción del modelo.

Al construir el modelo SIMCA con las huellas dactilares aumentadas se necesitaron 9 PCs para definir correctamente la clase VOF y 6 para la clase VMX; obteniendo una varianza explicada  $> 95\%$  y un valor de RMSECV  $< 0.27$ . El gráfico de Hotelling  $T^2$  vs. residuales Q mostró que todas las muestras de calibración estaban dentro de los límites con un nivel de confianza del 95%. Todas las variables incluidas en la construcción de SIMCA exhibieron una buena capacidad de modelado, tal como se concluyó de la interpretación del gráfico de loadings de cada sub-modelo de clase.

En PLS-DA, las variaciones en la matriz X y el vector de variables categóricas Y fueron descritas con 8 variables latentes (LVs), presentando una varianza explicada del 91% y el 99%, respectivamente. Los estadísticos de la regresión mostraron un RMSEC de 0.015, un RMSECV de 0.025 y un  $r^2$  de la CV de 0.997. El intervalo para la aceptación de una muestra en una determinada clase fue calculado empleando la ecuación 3.1 y se estableció en  $\pm 0.20$ . Esto significa que todas las muestras calculadas con un valor de Y de 0.80 a 1.20 y que caen dentro de los límites establecidos por el valor de Hotelling  $T^2$  y el valor de los residuales Q, para un nivel de confianza del 95%, serán clasificadas como pertenecientes a la clase VOF.

Con el fin de comprobar las ventajas de utilizar varias longitudes de onda, los resultados obtenidos por el procedimiento propuesto fueron comparados con los obtenidos por la aplicación del mismo tratamiento de datos pero a cromatogramas registrados a una sola longitud de onda (226 nm). En la Tabla 5.3 se muestra el comportamiento de cada método de clasificación en los dos casos.

**Tabla 5.3.** Comparación de los resultados de la clasificación.

Modo	Modelo	Tipo de muestra	n <sup>a</sup>	Clasificada en <sup>b</sup>		Sensibilidad (%)	Especificidad (%)	MCC (%) <sup>c</sup>	GSR (%) <sup>d</sup>	
				VOF	AHG					
Huellas dactilares aumentadas	SIMCA								97	
		VOF	36	34	2	94				
		OHs <sup>e</sup>	31	0	31		100			
		Mezclas	9	0	9			100		
	PLS-DA									100
		VOF	36	36	0	100				
		OHs <sup>e</sup>	31	0	31		100			
		Mezclas	9	0	9			100		
	Longitud de onda única (226 nm)	SIMCA								90
VOF			36	36	0	100				
OHs <sup>e</sup>			31	0	31		100			
		Mezclas	9	8	1			11		
PLS-DA										97
		VOF	36	36	0	100				
		OHs <sup>e</sup>	31	0	31		100			
		Mezclas	9	2	7			78		

<sup>a</sup> Número de huellas dactilares aumentadas o número de huellas dactilares convencionales empleadas en la predicción.

<sup>b</sup> VOF, *V. officinalis*; AHG, otro grupo de plantas.

<sup>c</sup> Porcentaje de mezclas correctamente clasificadas como ajenas a la clase VOF.

<sup>d</sup> Tasa general de éxito en la clasificación.

<sup>e</sup> Otras plantas, aquí se incluye *V. wallichii*, *V. edulis*, *P. incarnata* y las dos muestras sospechosas descritas en la sección 5.3.3.

Siempre se obtuvo alta sensibilidad y especificidad, independientemente del modo o el modelo utilizado para realizar la clasificación. Esta situación puede presentarse porque las diferencias químicas entre las plantas en estudio son grandes y porque los modelos de clasificación hacen estas diferencias más evidentes para el analista. La principal diferencia entre el uso de una o varias longitudes de onda para el desarrollo de las huellas dactilares, se produjo al clasificar de las muestras adulteradas. Los resultados obtenidos con huellas dactilares aumentadas fue mejor que el conseguido empleando huellas dactilares a una sola longitud de onda. Incluso si el adulterante se encuentra aproximadamente a un 5% y es de la familia *Valerianaceae*,

el resultado muestra una mejora en la detección de muestras adulteradas. Debido a los mejores resultados para los valores de MCC, el enfoque de las huellas dactilares aumentadas también alcanzó los valores más altos de GSR en cada modelo construido con esta metodología. Comparando los modelos SIMCA y los modelos PLS-DA en cada modo de establecimiento de la huella dactilar, es evidente que la última técnica proporciona mejores resultados.

#### **5.4. Conclusiones del capítulo.**

El método propuesto de utilización de huellas dactilares cromatográficas aumentadas representa una buena alternativa para mejorar los resultados de clasificación de especies de plantas con cromatogramas complejos. La clave para el correcto uso de esta metodología es la selección de las  $k$  longitudes de onda a utilizar para obtener, en los cromatogramas registrados a cada una de ellas, picos no coincidentes de los componentes químicos detectables de las plantas estudiadas y con alta señal (en este caso 226, 254, 280 y 326 nm); además, es necesaria la aplicación de unos pretratamientos adecuados para eliminar las variaciones no deseadas en los cromatogramas, entre las que se incluye la corrección de la línea base, la normalización y la corrección de los desplazamientos en el tiempo de retención.

La combinación de COW y el PCA ha permitido la alineación de cromatogramas complejos compilados utilizando varias longitudes de onda. Así se evitó la introducción de ruido y artefactos matemáticos.

Los dos métodos de clasificación utilizados, SIMCA y PLS-DA, permitieron hacer una buena discriminación entre *V. officinalis* y otras plantas, presentando una sensibilidad y especificidad superior al 94%. Las pruebas con mezclas preparadas en el laboratorio mostraron que los modelos de clasificación utilizados son sumamente útiles para la detección de falsificaciones y adulteraciones cuando se utilizan las huellas dactilares aumentadas. PLS-DA proporcionó los mejores resultados, ya que su valor GRS fue mayor que el obtenido con SIMCA.

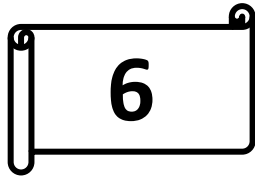




## Referencias.

- (1) Circosta, C.; De Pasquale, R.; Samperi, S.; Pino, A.; Occhiuto, F. Biological and analytical characterization of two extracts from *Valeriana officinalis*. *J. Ethnopharmacol.* 2007, *112*, 361-367.
- (2) Navarrete, A.; Avula, B.; Choi, Y. W.; Khan, I. A. Chemical fingerprinting of valeriana species: simultaneous determination of valerenic acids, flavonoids, and phenylpropanoids using liquid chromatography with ultraviolet detection. *J. AOAC Int.* 2006, *89*, 8-15.
- (3) Bos, R.; Woerdenbag, H. J.; Pras, N. Determination of valepotriates. *Journal of Chromatography A* 2002, *967*, 131-146.
- (4) Joshi, V. C.; Navarrete, A.; Khan, I. A. Authentication of *Valeriana procera* Kunth and comparative account of five *Valeriana* species. *J. AOAC Int.* 2005, *88*, 1621-1625.
- (5) Hattesoehl, M.; Feistel, B.; Sievers, H.; Lehnfeld, R.; Hegger, M.; Winterhoff, H. Extracts of *Valeriana officinalis* L. s.l. show anxiolytic and antidepressant effects but neither sedative nor myorelaxant properties. *Phytomedicine* 2008, *15*, 2-15.
- (6) Dejaegher, B.; Alaerts, G.; Matthijs, N. Methodology to develop liquid chromatographic fingerprints for the quality control of herbal medicines. *Acta Chromatographica* 2010, *22*, 237-258.
- (7) Jiang, Y.; David, B.; Tu, P.; Barbin, Y. Recent analytical approaches in quality control of traditional Chinese medicines—A review. *Anal. Chim. Acta* 2010, *657*, 9-18.
- (8) Liang, Y.; Xie, P.; Chan, K. Quality control of herbal medicines. *Journal of Chromatography B* 2004, *812*, 53-70.
- (9) Yan, S.; Xin, W.; Luo, G.; Wang, Y.; Cheng, Y. An approach to develop two-dimensional fingerprint for the quality control of Qingkailing injection by high-performance liquid chromatography with diode array detection. *Journal of Chromatography A* 2005, *1090*, 90-97.
- (10) Fan, X.; Cheng, Y.; Ye, Z.; Lin, R.; Qian, Z. Multiple chromatographic fingerprinting and its application to the quality control of herbal medicines. *Anal. Chim. Acta* 2006, *555*, 217-224.
- (11) Ni, Y.; Lai, Y.; Brandes, S.; Kokot, S. Multi-wavelength HPLC fingerprints from complex substances: An exploratory chemometrics study of the Cassia seed example. *Anal. Chim. Acta* 2009, *647*, 149-158.
- (12) Li, B.; Hu, Y.; Liang, Y.; Xie, P.; Du, Y. Quality evaluation of fingerprints of herbal medicine with chromatographic data. *Anal. Chim. Acta* 2004, *514*, 69-77.
- (13) Hendriks, M. M. W. B.; Cruz-Juarez, L.; Bont, D. D.; Hall, R. D. Preprocessing and exploratory analysis of chromatographic profiles of plant extracts. *Anal. Chim. Acta* 2005, *545*, 53-64.
- (14) Skov, T.; van den Berg, F.; Tomasi, G.; Bro, R. Automated alignment of chromatographic data. *Journal of Chemometrics* 2006, *20*, 484-497.
- (15) Tomasi, G.; van den Berg, F.; Andersson, C. Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. *Journal of Chemometrics* 2004, *18*, 231-241.
- (16) Brereton, R. G. *Applied Chemometrics for Scientist*, John Wiley & Sons, Ltd, Great Britain, 2007.

- (17) Lau, C.; Chan, C.; Chau, F.; Mok, D. K. Rapid analysis of Radix puerariae by near-infrared spectroscopy. *Journal of Chromatography A* 2009, *1216*, 2130-2135.
- (18) Daszykowski, M.; Walczak, B. Use and abuse of chemometrics in chromatography. *TrAC Trends in Analytical Chemistry* 2006, *25*, 1081-1096.
- (19) Bos, R.; Woerdenbag, H. J.; Hendriks, H.; Zwaving, J. H.; De Smet, P. A. G. M.; Tittel, G.; Wikström, H. V.; Scheffer, J. J. C. Analytical Aspects of Phytotherapeutic Valerian Preparations. *Phytochemical Analysis* 1996, *7*- 143-151.
- (20) Gao, X. Q.; Björk, L. Valerenic acid derivatives and valepotriates among individuals, varieties and species of Valeriana. *Fitoterapia* 2000, *71*, 19-24.



***Turnera diffusa.***

---



## 6.1. Introducción.

### 6.1.1. Descripción de la planta medicinal.

*Turnera diffusa* Wild. Ex Schult, también conocida como damiana, es un pequeño arbusto perteneciente a la familia *Turneraceae*, que crece de 1 a 2 metros de altura y tiene hojas serradas aromáticas de 10 a 25 cm de largo. En verano la damiana exhibe unas pequeñas flores amarillas, que son seguidas por pequeños frutos con un olor dulce y un sabor similar al higo. La parte medicinal de la planta son sus hojas, las cuales son cosechadas, de acuerdo con las prácticas de la etnofarmacología, durante la época de floración [1,2]. El arbusto de damiana se puede encontrar por todo México, América Central, las islas del Caribe y partes de América del Sur. Tiene varios usos tradicionales como afrodisíaco, para las dolencias hepáticas, depresión, ansiedad, neurosis, como expectorante, estimulante y tónico; también se utiliza para dar sabor a postres, bebidas, dulces, etc [3]. Además, se ha informado que la parte aérea (tallos y hojas) presentó una buena actividad antioxidante, similar a la exhibida por la quercetina, utilizada como control [4]. Los antioxidantes son conocidos para reducir el riesgo de padecer ciertos tipos de cáncer y muchas enfermedades crónico-degenerativas; por eso existe un creciente interés en la búsqueda de antioxidantes naturales, a fin de obtener diversos productos que son utilizados principalmente como suplementos dietéticos [5,6].

### 6.1.2. Problemática con el control de calidad de *Turnera diffusa*.

Esta planta fue seleccionada porque las organizaciones (EMA, FDA, etc.) todavía tienen que recomendar una metodología analítica definitiva para su control de calidad y porque se ha demostrado que muchos de los productos que dicen contener *T. diffusa* en el mercado mexicano contienen en realidad otras especies, de acuerdo a los trabajos realizados a partir de huellas dactilares obtenidas por TLC y CLAR-DAD [1,7,8].

### 6.1.3. Antecedentes.

Además de estudiar las identidades de las especies de las plantas medicinales, algunas publicaciones exploran la utilidad de las huellas dactilares obtenidas por CLAR-DAD, como una herramienta para predecir las propiedades terapéuticas de las mismas. Por ejemplo, Dumarey et al. [9] y Van Nederkassel [10] mostraron una correlación entre la huella dactilar cromatográfica (a 280 nm) de las muestras de té verde y su capacidad antioxidante, como una medida de los efectos protectores originados por el té. Sin embargo, en la correlación de las huellas dactilares con una posible actividad biológica, los químicos suelen utilizar un solo cromatograma a una longitud de onda fija, de modo que una gran cantidad de información contenida en otras longitudes de onda puede ser descartada, soslayando picos cromatográficos que pueden ser útiles para caracterizar las propiedades de una planta medicinal.

Un estudio previo sobre la correlación entre las huellas dactilares cromatográficas de *T. diffusa* (a 254 nm) y su actividad antioxidante utilizando PLSR fue realizado por Garza-Juárez et al. [11]; sin embargo, las limitaciones cuando se utiliza una sola longitud de onda se hicieron evidentes en los resultados del modelo matemático y en las predicciones, obteniendo un  $r^2$  de la validación cruzada de 0.80, un claro sesgo con tendencia a valores negativos y un deficiente resultado del modelo al predecir un conjunto externo de muestras. Además, la corrección de los desplazamientos del tiempo de retención se realizó manualmente, por lo que el preprocesamiento para la alineación era muy laborioso.

En el capítulo 5, se ha descrito una estrategia que implica el uso de varias longitudes de onda para la construcción de huellas dactilares aumentadas para plantas medicinales. Con el fin de mostrar la utilidad de este enfoque con múltiples longitudes de onda en la predicción de la actividad biológica de las fitomedicinas, se ha aplicado el procedimiento en cromatogramas de *Turnera diffusa*.

#### **6.1.4. Descripción del estudio.**

La actividad antioxidante podría ser una medida de la eficacia y de la calidad de la planta, si se considera que los efectos terapéuticos atribuidos a la damiana están relacionados, al menos en cierta medida, con dicha actividad. Las huellas dactilares cromatográficas construidas con múltiples longitudes de onda son una buena opción para predecir la actividad antioxidante de damiana y otras plantas medicinales, ya que representan la mayoría de los componentes de las plantas y porque es ventajoso manejarlas usando varios métodos quimiométricos. Así, en este trabajo se ha utilizado la estrategia de huellas dactilares aumentadas para mejorar la predicción de la actividad biológica de las plantas medicinales como parte de su control de calidad total. Se pretende mejorar el modelo de regresión PLSR y obtener predicciones más robustas para *T. diffusa*. Para facilitar la alineación de los cromatogramas se aplicó el algoritmo de alineación optimizada mediante correlación (COW).

#### **6.2. Metodología específica.**

##### **6.2.1. Descripción de las muestras.**

Con el fin de incluir en las muestras tantas fuentes de variabilidad como sea posible, independientemente del origen geográfico específico, se recolectaron cuarenta muestras de *Turnera diffusa* en diferentes regiones de México entre diciembre de 2005 y enero de 2009; las muestras fueron autenticadas y fue obtenido un comprobante de los especímenes depositados en el herbario de la Facultad de Ciencias Biológicas, Universidad Autónoma de Nuevo León (México). Las muestras de damiana se almacenaron protegidas de la luz intensa y de la humedad a temperatura de laboratorio (27 °C aprox.) y se dejaron secar.

##### **6.2.2. Reactivos.**

El metanol grado HPLC (MeOH) fue adquirido a Fisher Scientific (Fair Lawn, NJ, EE.UU.). El agua desionizada se obtuvo de Laboratorios de Monterrey S.A. de C.V.

(Monterrey, NL, México). El ácido trifluoroacético (TFA), el MeOH y el etanol (EtOH) grado reactivo fueron adquiridos a Fermont (Monterrey, NL, México). El 2,2-difenil-1-picrilo-hidrazil (DPPH) y la 2-(3,4-dihidroxifenil)-3,5,7-trihidroxichromen-4-ona (quercetina), utilizados para obtener la actividad de captación de los radicales por espectrofotometría, fueron adquiridos a Sigma (Monterrey, NL, México).

### **6.2.3. Preparación de los extractos.**

Las partes aéreas de las plantas secas (hojas y tallos) fueron trituradas y luego pasadas a través de un tamiz de 40 mesh. Cada muestra en polvo (1.00 g) se pesó con precisión y se extrajo tres veces con 5 mL de una solución de etanol-agua 9:1 (v / v) en cada vez, a 27 ° C, mediante un mezclador en vórtice durante 3 min y las respectivas soluciones fueron combinadas. El extracto final se filtró a través de un filtro de papel Whatman No. 40 y se evaporó hasta sequedad a presión reducida empleando un evaporador rotatorio a  $37 \pm 2$  °C. Antes del uso, se disolvieron 15 mg del extracto de la muestra, en 1 mL de MeOH y se filtró a través de un acrodisco de nylon de 0.45  $\mu$ m (Waters Corporation).

### **6.2.4. Instrumentación y procedimientos analíticos.**

Los análisis por CLAR se realizaron con un sistema Waters Alliance 2695, equipado con un detector UV-Vis de diodos en línea 2996, un inyector automático, un compartimiento de columna con temperatura controlada, un desgasificador de vacío y un ordenador equipado con el programa de control y tratamiento de datos Empower. La actividad de captación de radicales libres de los extractos de damiana se obtuvo usando un espectrofotómetro UV-Vis Beckman DU-7500.

#### **6.2.4.1. Procedimiento por CLAR-DAD.**

La separación se consiguió utilizando una columna C18 AccQ Tag de 3.9 x 150 mm con 4  $\mu$ m de tamaño de partícula (Waters Corporation) utilizada a 30 °C. La fase móvil consistió en (A) TFA al 0.1% en agua y (B) MeOH. Antes del uso, los componentes



de la fase móvil fueron desgasificados y filtrados a través de un filtro de nylon de 0.45  $\mu\text{m}$  (Waters Corporation, Milford, MA, EE.UU.). La secuencia del gradiente de elución fue realizada como sigue: inicio con 30% de B, incrementando el porcentaje de B desde 30 hasta 70% durante 30 min y manteniéndolo constante durante 5 min. Cada separación fue seguida por un período de equilibrio de 10 min a las condiciones iniciales (70% A, 30% B). Antes de la primera inyección, la columna fue preacondicionada durante 1 h con la fase móvil inicial. El flujo de la fase móvil se mantuvo a 0.4 mL  $\text{min}^{-1}$  y el volumen de inyección fue de 10  $\mu\text{L}$ . La longitud de onda del detector se estableció a 254 nm para seguir el perfil de elución. Los espectros UV se obtuvieron en el rango de 210 a 400 nm en intervalos de 1.2 nm y el tiempo de muestreo del detector fue 0.01667 min.

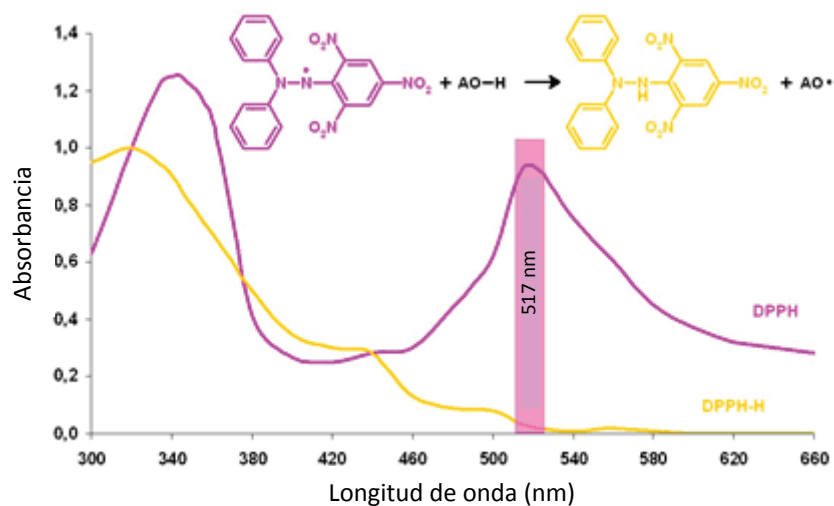
#### **6.2.4.2. Determinación de la actividad captadora del radical libre DPPH mediante espectrofotometría.**

Se llevó a cabo un ensayo de la actividad captadora del DPPH utilizando espectrofotometría de acuerdo a Leu et al. [12], con algunas modificaciones. Primero, el extracto se volvió a disolver en etanol (1 mg  $\text{mL}^{-1}$ ) y se utilizaron diferentes concentraciones (0.2 a 200  $\mu\text{g mL}^{-1}$ ) de cada extracto. En un volumen total de 1 mL, la mezcla de ensayo contenía 500  $\mu\text{L}$  del extracto y 500  $\mu\text{L}$  del DPPH (125  $\mu\text{M}$  en etanol). La mezcla de ensayo se agitó y se dejó reposar a temperatura ambiente en la oscuridad durante 30 min; a continuación, se midió la absorbancia a 517 nm. La quercetina se utilizó como control positivo. El fundamento del ensayo se ilustra en la Figura 6.1. La capacidad del radical DPPH para captar electrones se calculó como sigue:

$$\text{Actividad captadora del radical}(\%) = \frac{A-B}{A} \times 100 \quad (6.1)$$

donde A es la absorbancia del control negativo (DPPH más etanol), y B es la absorbancia de la muestra (DPPH, etanol más la muestra). Se realizó el gráfico de la concentración vs. el porcentaje de captación y la concentración efectiva del extracto que captó el 50% de los radicales DPPH ( $\text{EC}_{50}$ ) fue calculada por interpolación,

expresando la actividad como  $EC_{50}$ . Las actividades antioxidantes de las muestras y los respectivos cromatogramas fueron obtenidos simultáneamente, para evitar posibles variaciones en los resultados de la prueba del DPPH, debido al tiempo de almacenamiento de los extractos.



**Figura 6.1.** El radical estable DPPH (color violeta) capta un átomo de hidrógeno del antioxidante, formando DPPH-H (color amarillo).

### 6.2.5. Análisis multivariante.

Se utilizaron los datos de un estudio realizado por Garza-Juárez et al. [11]. Los datos CLAR-DAD de cuarenta muestras de damiana fueron exportados del programa Empower en formato ASCII e importados a MATLAB 7.9 (The MathWorks, Inc.); que fue utilizado para el manejo de las matrices cromatográficas y para realizar el escalado. Utilizando el programa PLS-Toolbox 6.2 (Eigenvector Research), se realizó la combinación de datos adyacentes, la corrección de la línea base, el desdoblamiento de la matriz de datos y la construcción de los modelos matemáticos.

La selección del cromatograma de referencia utilizado en la alineación, la optimización simplex para establecer la longitud del segmento y el grado de flexibilidad (slack size) y la alineación mediante COW fueron realizados utilizando los códigos para Matlab desarrollados por Skov [13] y por Tomasi [14].

Se llevó a cabo un PCA para verificar la utilidad de la alineación realizada mediante la evaluación de los resultados del modelo antes y después de efectuar la alineación.

Para construir el modelo PLSR, la matriz X se compone de las huellas dactilares aumentadas y el vector Y está construido con los valores de referencia de la actividad antioxidante ( $EC_{50}$ ) obtenidos por el ensayo con DPPH. Se utilizó la validación cruzada (CV) en la construcción de los modelos de calibración de PCA y PLSR. El número óptimo de factores del modelo se determinó con el criterio del valor mínimo de PRESS [15]. Los estadísticos calculados para los modelos de calibración incluyeron el RMSECV y coeficiente de determinación  $r^2$ .

#### **6.2.6. Construcción de las huellas dactilares con múltiples longitudes de onda.**

El procedimiento general del enfoque utilizado para la construcción de huellas dactilares aumentadas para plantas medicinales ha sido descrito en las secciones 5.1.4 y 5.2.6. La Figura 5.1 presenta los pasos de manera esquemática. La matriz de huellas dactilares aumentadas, ya corregida, es utilizada para construir los modelos de exploración y de regresión por medio del PCA y la PLSR. Tal como se mencionó en el capítulo 5, esta estrategia permite la inclusión de tantas longitudes de onda como seleccione el investigador; así, la mayoría de la información cromatográfica relevante puede ser considerada en la construcción de los modelos matemáticos y por lo tanto, este enfoque puede alcanzar un mejor desempeño que el obtenido usando una sola longitud de onda; especialmente en los resultados cuantitativos del PLSR al predecir una actividad biológica específica.

#### **6.2.7. Alineación de las huellas dactilares aumentadas mediante COW.**

En la sección 5.2.7 se describe el procedimiento para la optimización de los parámetros en la alineación de las huellas dactilares aumentadas.

En el caso de las huellas dactilares de damina, el espacio de optimización para obtener los valores de la longitud del segmento y el grado de flexibilidad fue definido como sigue: segmento de 45 a 135 y grado de flexibilidad de 4 a 40. El número de puntos de inicio en la red de búsqueda se ajustó a 5, el número máximo de pasos de optimización a 50 y la fracción de la máxima desviación del centro en la alineación COW al 15%.

### **6.3. Resultados y discusión.**

#### **6.3.1. Resultados del ensayo con DPPH y del método CLAR.**

Se determinó el porcentaje de reducción del radical DPPH exhibido por las diferentes concentraciones de una muestra dada y, posteriormente, fue calculado su  $EC_{50}$ ; cada muestra se midió por triplicado, obteniendo valores de RSD entre 0.1 y 18.1%.

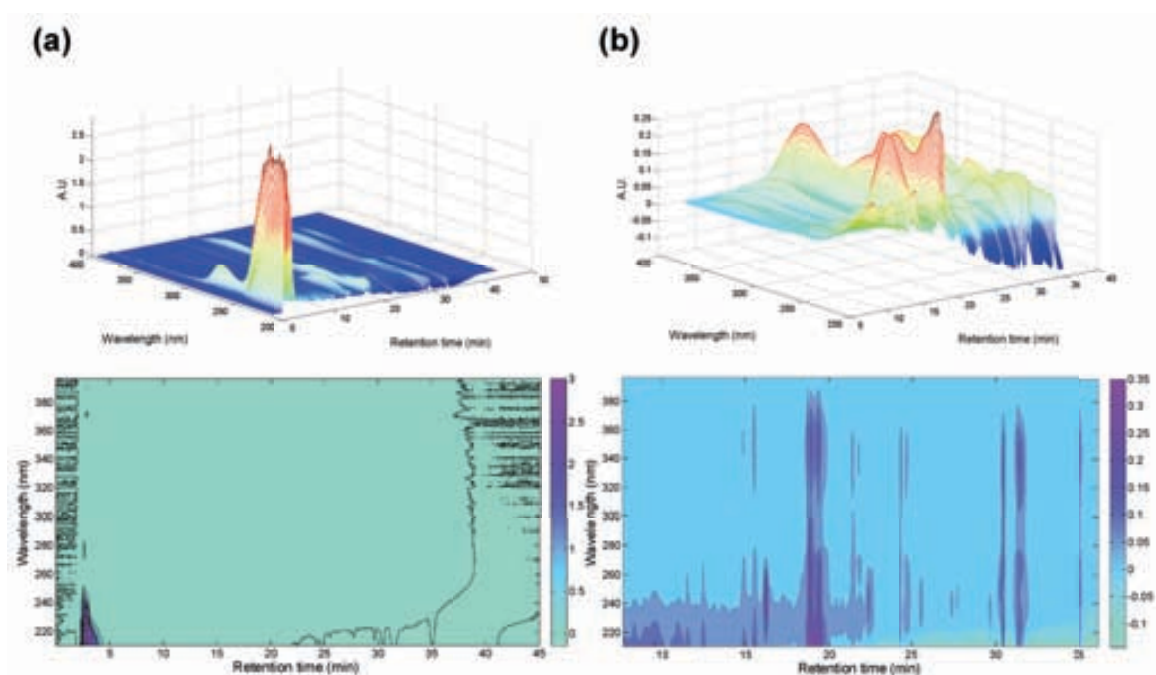
La precisión del método cromatográfico fue evaluada considerando los tiempos de retención relativos, las alturas de los picos relativas y las áreas relativas de 12 picos comunes a todos los cromatogramas, los cuales abarcaron toda la región cromatográfica, utilizando la metodología de [1]. Los resultados expresados como RSD estuvieron entre 0.1 y 2.8% para los tiempos de retención; 2.3 y 12.0% para las áreas relativas y entre 2.6 y 16.0% para las alturas relativas. Así, todos los resultados indicaron un comportamiento aceptable de los procedimientos analíticos utilizados.

#### **6.3.2. Construcción de las huellas dactilares aumentadas.**

Los cromatogramas fueron muy complejos, presentando unos 46 picos. Para cada muestra, se obtuvo una matriz de datos cromatográficos con dimensiones 2700 x 158 (tiempo de retención x longitudes de onda).

### 6.3.2.1. Selección de las longitudes de onda de trabajo.

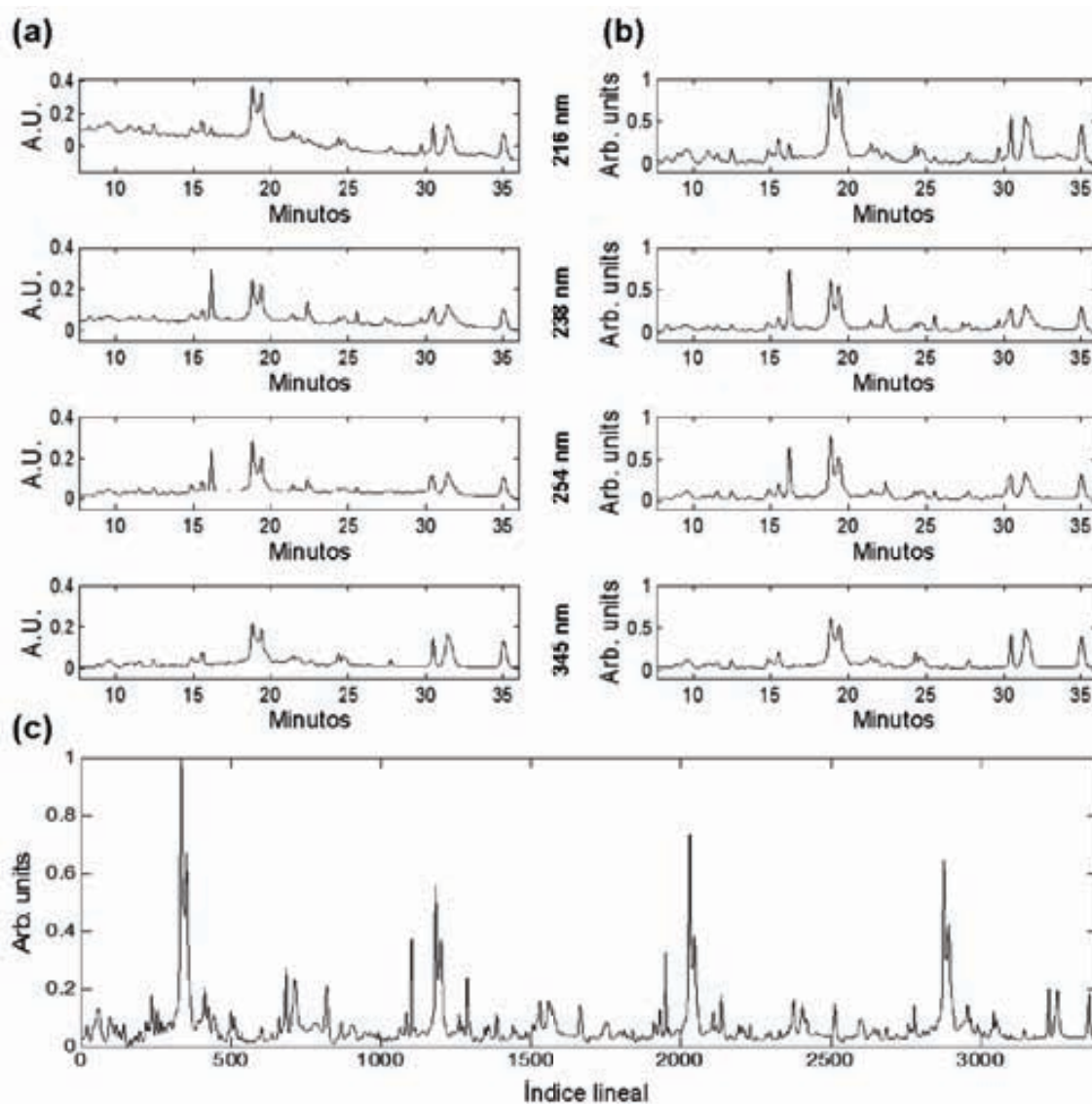
La Figura 6.2a muestra un gráfico 3D de un cromatograma sin tratar de una muestra de damiana y su correspondiente gráfico de contorno. Usando el gráfico de contorno de los datos cromatográficos originales, era difícil obtener información debido a la alta intensidad de una señal ancha y ruidosa a tiempos de retención inferiores a 7.66 minutos y que corresponde al frente de elución, que contiene varios compuestos no resueltos; por lo tanto esta zona de tiempos de retención fue eliminada de los cálculos posteriores. Además, los tiempos de retención por encima de 35.86 minutos fueron eliminados porque sólo había señal de fondo (línea base). De esta manera, en el intervalo de tiempo de retención seleccionado, fueron conservados los picos mejor resueltos y más estables en términos de área y se evitó la inclusión de ruido o de señales sin información en los modelos matemáticos, quedando ahora descrita cada muestra por una matriz de dimensiones (1692 x 158).



**Figura 6.2.** Arriba, cromatograma en 3D (min x nm x absorbancia [A.U.]) de una muestra de damiana y abajo su respectivo gráfico de contorno; (a) cromatograma completo y (b) después de la selección del intervalo de tiempo de retención (retention time).

La Figura 6.2b muestra el cromatograma en el intervalo de tiempo de retención seleccionado. A partir del gráfico de contorno es posible observar que a longitudes de onda concretas hay diferentes picos con diversas intensidades. La mayoría de los

constituyentes químicos detectados en los extractos de damiana mostraron las mayores respuestas en cuatro longitudes de onda: 216, 238, 254 y 345 nm; en consecuencia, los cromatogramas a estas longitudes de onda se utilizaron para construir la huella dactilar aumentada de cada muestra, nótese que no todos los picos están presentes en todas las longitudes de onda, véase la Figura 6.3a. De esta manera se consiguió incluir un gran número de picos con altas intensidades, mientras que las longitudes de onda con información redundante fueron excluidas.



**Figura 6.3.** Cromatogramas de una muestra de *T. diffusa* a las longitudes de onda empleadas en la construcción de la huella dactilar aumentada.; (a) antes del pretratamiento de datos y (b) después de la combinación de datos, la corrección de línea base y el escalado. A.U. corresponde a unidades de absorbancia y Arb. units indica unidades arbitrarias. (c) Aspecto típico de una huella dactilar aumentada de *T. diffusa*.

### 6.3.2.2. Efecto de los pretratamientos.

Se realizó la traspuesta de la matriz de datos que describe cada muestra y se co-adicionó la dimensión del tiempo de retención. En este caso las variables adyacentes fueron combinadas de dos en dos, conservando el valor medio; obteniendo una nueva matriz para representar la muestra de (158 x 847, [longitudes de onda x tiempo de retención]). Se extrajeron las cuatro longitudes de onda de trabajo y las dimensiones finales de la matriz fueron (4 x 847).

La Fig. 6.3a muestra las variaciones presentadas por los cromatogramas de una muestra de *T. diffusa* a las longitudes de onda seleccionadas para construir las huellas dactilares aumentadas. La más evidente de estas variaciones fue el cambio en la pendiente de la línea base; que también fue evidente en el cromatograma 3D de la Fig. 6.2b. Aparece una ligera pendiente a valores negativos a 238, 254 y 345 nm y una mayor tendencia a valores negativos a 216 nm. En tales condiciones, no era viable desdoblar la matriz cromatográfica para construir el vector de la huella dactilar aumentada, de modo que, los datos tuvieron que ser corregidos. Se eliminó la señal de fondo por medio de WLS utilizando un polinomio de tercer grado. Los principios del método WLS han sido expuestos en la sección 1.6.1.5 y su aplicación descrita en el Capítulo 5; se pueden hallar más detalles sobre el método en [16]. La variabilidad en la escala entre las muestras de damiana se minimizó aplicando para cada variable un rango de escala de tal manera que, en los vectores con la línea base corregida, el valor más alto fuera +1 y el más bajo 0. La Fig. 6.3b muestra los resultados de la aplicación de los pretratamientos a los vectores que forman la matriz cromatográfica. La combinación de los datos mantiene la información cromatográfica, es decir, no afecta a las características del cromatograma, y reduce el número de datos de la matriz en la dimensión del tiempo de retención. La corrección de la línea base eliminó con éxito las tendencias mostradas por la señal de fondo de los cromatogramas, mientras que al llevar a la unidad la escala de las huellas dactilares aumentadas permitió mantener las proporciones de intensidad entre las cuatro longitudes de onda utilizadas para construir la huella dactilar aumentada, evitando así discontinuidades en los valores de absorbancia cuando las cuatro longitudes de onda fueron desplegadas.

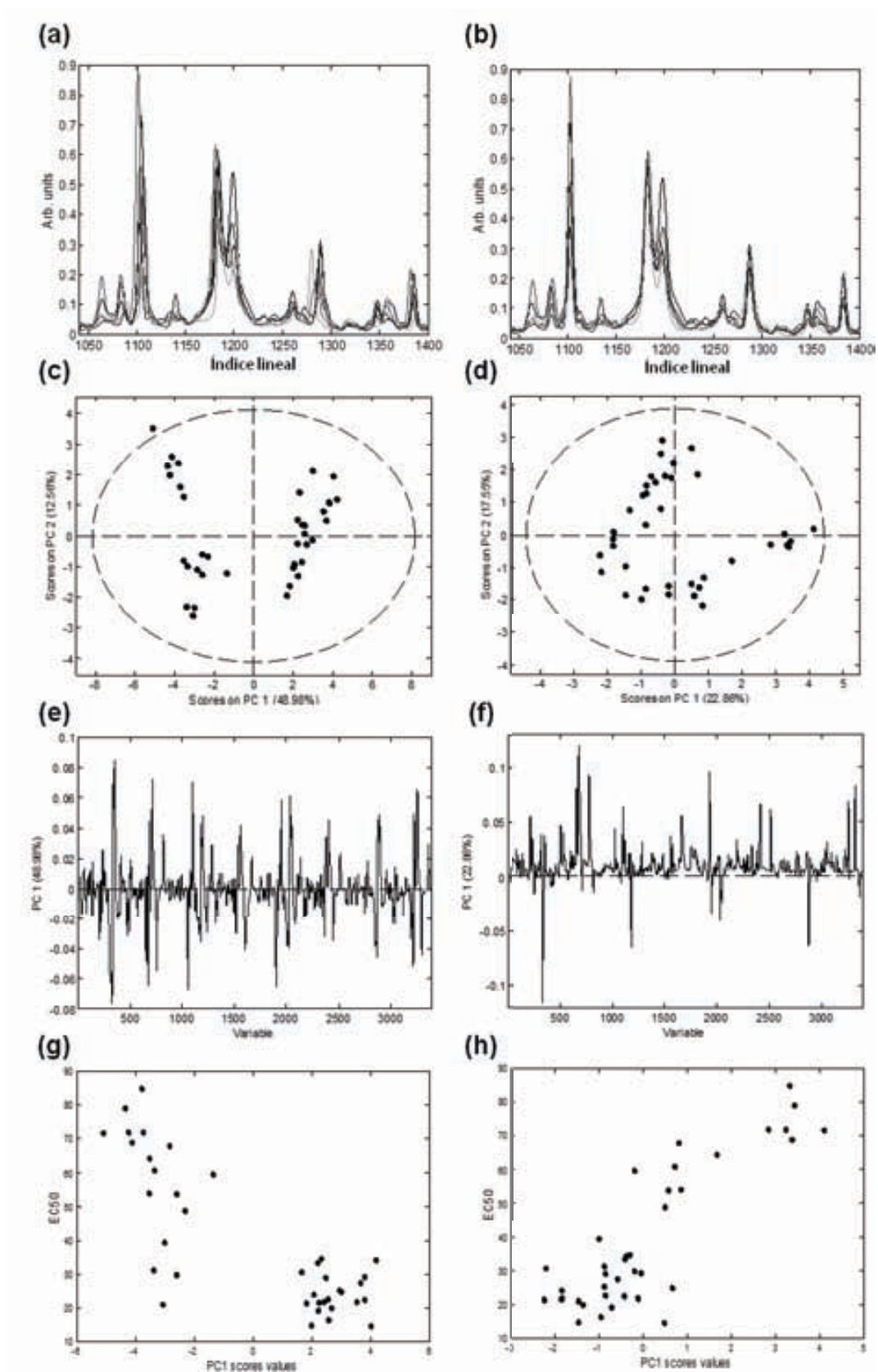
Finalmente., la matriz corregida se desplegó a lo largo de la dimensión del tiempo de retención para generar un único vector de datos con dimensiones de 1 x 3388, que es la huella dactilar aumentada de la muestra; Figura 6.3c. Como era de esperar, la huella dactilar aumentada presentó más conjuntos de picos que los exhibidos a una sola longitud de onda. Debido al despliegue de la matriz se pierde la dimensión del tiempo de retención, la cual es sustituida con un índice lineal que fácilmente se puede traducir otra vez a unidades de tiempo de retención. Las huellas dactilares aumentadas de todas las muestras se ordenaron en una única matriz para la posterior alineación y la construcción de los modelos.

### 6.3.2.3. Alineación.

La corrección del desplazamiento del tiempo de retención fue el último pretratamiento de datos realizado; los parámetros de entrada fueron la huella dactilar aumentada utilizada como referencia, la longitud del segmento y el grado de flexibilidad. La optimización simplex empleada para encontrar los parámetros de trabajo para la alineación proporcionó una longitud de segmento de 45 y un grado de flexibilidad de 23; las Fig. 6.4a y Fig. 6.4b presentan las huellas dactilares aumentadas de varias muestras antes y después del alineamiento utilizando estos parámetros. Los cambios del tiempo de retención aparecen corregidos correctamente y los picos y características de los cromatogramas no fueron distorsionadas por el uso del algoritmo COW.

El gráfico de scores del PCA realizado antes de la alineación (Fig. 6.4c) muestra dos grupos principales de muestras separadas por la PC1; el grupo en valores negativos de score en PC1 está también dividido en dos subgrupos. El análisis de los loadings de este modelo sugirió que el gráfico de scores observado se debe principalmente a los cambios del tiempo de retención, ya que un gráfico de los loadings presentó un aspecto distorsionado y exhibió formas similares a las obtenidas cuando se calcula una derivada; Fig. 6.4e.





**Figura 6.4.** (a) Sección de varias huellas dactilares aumentadas antes de COW y (b) las mismas muestras después de la alineación. (c) y (d) Gráficos de score de PC1 vs. PC2 de un PCA global antes y después de la alineación, respectivamente. (e) y (f) Gráficos de loadings, antes y después de efectuar COW. (g) Representación de los valores de score de PC1 vs. la actividad antioxidante (EC<sub>50</sub>) para cada muestra antes de la alineación y (h) después de efectuarla.

La Fig. 6.4d es el gráfico de scores obtenido después de realizar la corrección de los desplazamientos del tiempo de retención; la división de los dos subgrupos ya no está presente y el gráfico de loadings tiene más semejanza a un cromatograma (Fig. 6.4f), lo que sugiere que se corrigieron con éxito los cambios del tiempo de retención.

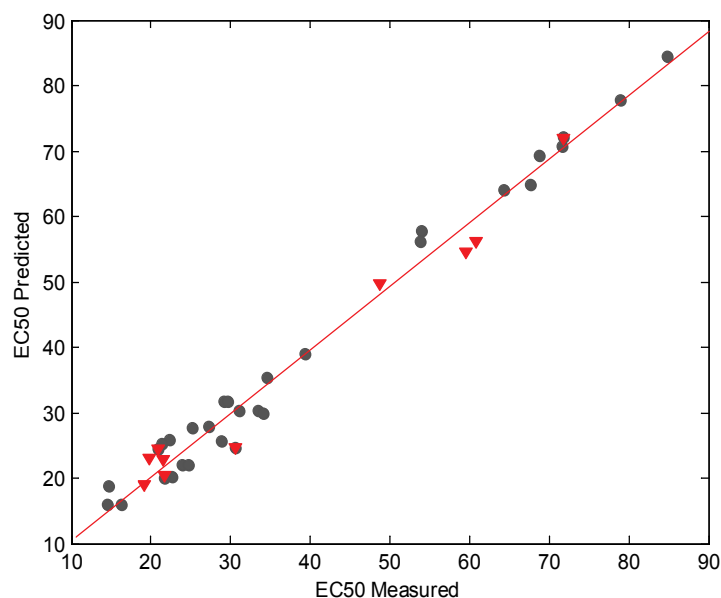
Con el fin de averiguar si existe algún tipo de relación entre las componentes principales y la actividad antioxidante, se representaron gráficamente los scores de cada componente principal calculado por el PCA frente a la actividad antioxidante. La proyección de las muestras en el eje PC1 está relacionada con los valores de  $EC_{50}$ ; al representar los scores de la PC1 vs. la actividad antioxidante, después de la alineación, fue evidente una relación lineal, véase la Fig. 6.4h. Sin embargo, el mismo gráfico antes de la alineación no muestra ninguna relación aparente; Fig. 6.4g. No hubo ninguna relación evidente entre los otros componentes principales y los valores de  $EC_{50}$ .

### **6.3.3. Predicción de la actividad antioxidante ( $EC_{50}$ ) por PLSR.**

Se construyó un modelo PLSR para relacionar la actividad antioxidante y los datos cromatográficos a partir de las huellas dactilares aumentadas de 30 muestras. Para evaluar el comportamiento del modelo, se utilizaron 10 muestras no empleadas para la calibración como un conjunto de predicción externa; los valores de estas muestras se extienden por todo el intervalo de calibración de la actividad antioxidante.

La matriz X de las huellas dactilares aumentadas tenía dimensiones (30 x 3388) y su respectivo vector Y de actividades (30 x 1). Los datos de X fueron centrados (mean centering) y se realizó una validación cruzada, aplicando el método de persianas venecianas, con 6 divisiones de los datos [17]. Fue elegido un modelo con cuatro variables latentes (LV) aplicando el criterio del mínimo PRESS, consiguiendo una varianza explicada del 52.5% para las variables X y de 98.4% para Y; se obtuvo un RMSECV de 6.02 y un  $r^2$  de la CV de 0.928. En ningún caso fueron detectadas muestras anómalas durante la construcción del modelo.

La Figura 6.5 muestra el gráfico de los valores de EC<sub>50</sub> medidos frente a los valores de EC<sub>50</sub> calculados, incluyendo los conjuntos de calibración y de predicción, con una pendiente de 0.972 (s = 0.022) y una ordenada en el origen de 0.88 (s = 0.94), indicando que los valores calculados son coincidentes con los medidos.



**Figura 6.5.** (●) Calibración; (▼) Valor calculado. Representación de los valores de referencia vs. los valores calculados de la actividad antioxidante (EC<sub>50</sub>) de cada muestra; se presenta la línea de regresión.

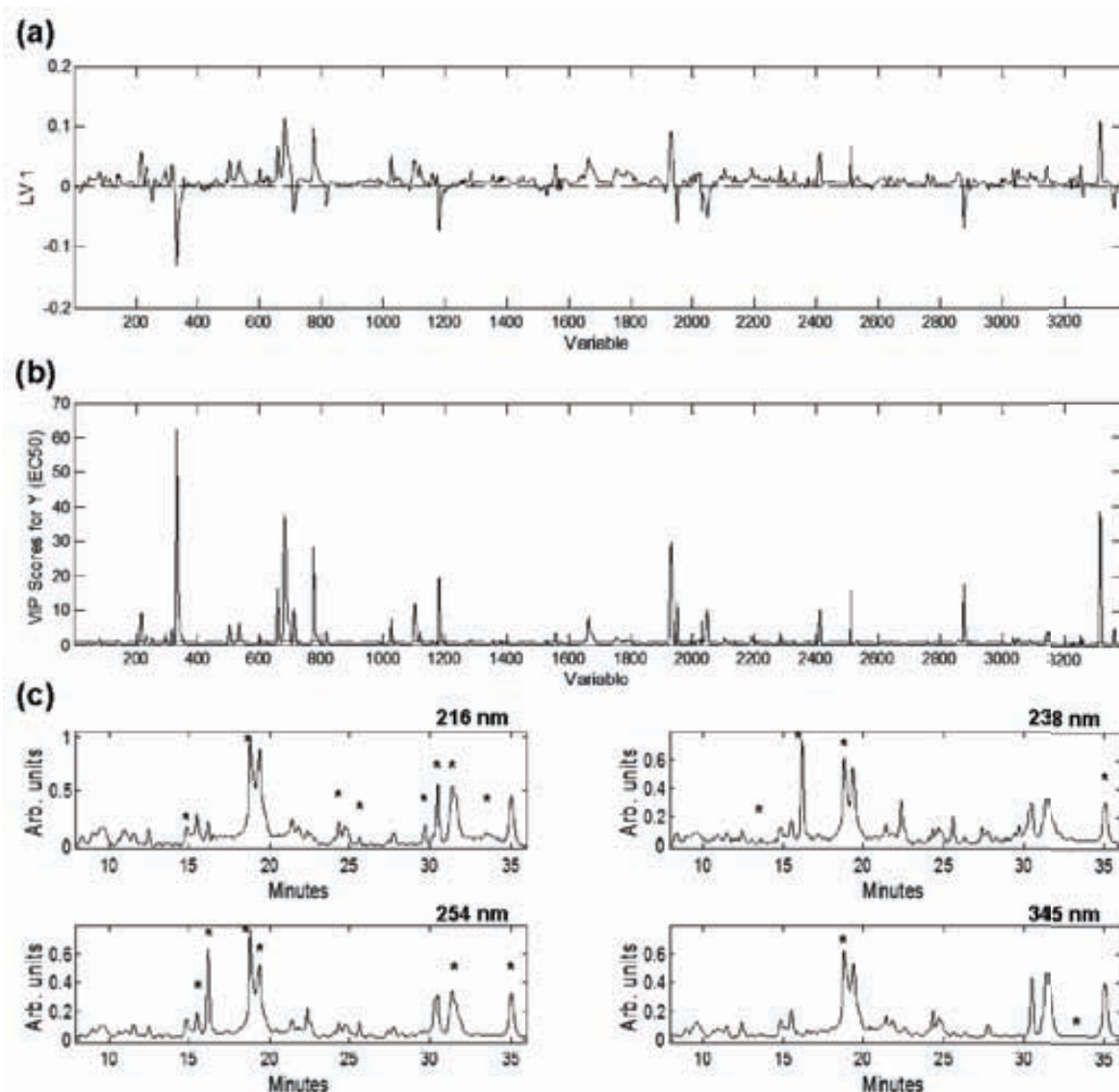
El error relativo estándar de predicción (RSEP) se calculó como sigue:

$$\%RSEP = \sqrt{\frac{\sum (y_{calc} - y_{exp})^2}{\sum y_{exp}^2}} \times 100 \quad (6.2)$$

donde  $y_{calc}$  es el valor de EC<sub>50</sub> calculado por PLSR,  $y_{exp}$  el valor de actividad encontrado con el ensayo del DPPH; la suma se extiende a todas las muestras de predicción [18]. El valor de RSEP obtenido utilizando las huellas dactilares aumentadas fue 7.8%, mientras que el valor calculado a partir de los datos publicados en la referencia [11], utilizando cromatogramas a 254 nm, fue 20.5%.

La primera variable latente explica el 91.3% de la varianza de Y ( $EC_{50}$ ), lo que sugiere que la LV1 contiene la mayor información para predecir la actividad antioxidante. Las Figuras 6.6a y 6.6b muestran el gráfico de loadings de la LV1 y un gráfico de la puntuación de la importancia de las variables en la proyección (variable importance in projection (VIP) scores) para  $EC_{50}$ , respectivamente. El VIP estima la importancia de cada variable utilizada en un modelo PLS en la proyección y se utiliza a menudo para la selección de variables. Las variables que tienen una contribución importante en el loading y un valor alto de VIP pertenecen a los compuestos que están muy relacionadas con la actividad antioxidante. Sin embargo, no puede ser deducido directamente si estas variables aumentan o disminuyen el valor calculado de  $EC_{50}$ , ya que los valores negativos de loadings no significan necesariamente que las variables en cuestión tengan una influencia negativa en el resultado calculado, de modo que no es posible asignar un papel antagonista o sinérgico a los picos relacionados a dichas variables.

Después de reconstruir los cromatogramas a partir de las respectivas huellas dactilares aumentadas, fue posible asignar las variables importantes encontradas en la regresión ( $VIP > 5$ ) a tiempos de retención específicos en cada una de las longitudes de onda. Estos tiempos de retención se correspondieron con el máximo de los picos. La Figura 6.6c muestra los cuatro cromatogramas de una muestra; todos los picos encontrados como relevantes estudiando las Figuras 6.6a y 6.6b fueron marcados, independientemente de si el pico está presente o no en esta muestra en particular. Recientemente, Pérez-Meseguer et al. [7] purificaron a partir de extractos de damiana un compuesto que mostró la mejor actividad antioxidante en un ensayo de CCF a 254 nm; se identificó como 8-C- $\beta$ -[6-desoxi-2-O-( $\alpha$ -1-ramnopiranosil)-xilo-hexopiranos-3-ulosido]. Esta es la única fuente natural de donde se ha reportado este compuesto; estuvo presente en 41 muestras analizadas de damiana nativa y es uno de los señalados en este trabajo como importante en la misma longitud de onda (en el tiempo de retención de 19.40 min), pero hay otros compuestos influyendo en la predicción del valor  $EC_{50}$ . Cualquier compuesto marcado como "importante", en cualquier longitud de onda, contribuye significativamente a la regresión, por lo que merece mayor investigación.



**Figura 6.6.** (a) Gráfico de los loadings en LV1 de las variables cromatograficas utilizadas. (b) importancia de las variables en la proyección (VIP) para EC<sub>50</sub>; se consideraron importantes para el modelo las variables con un valor superior a 5. (c) En los cromatogramas que integran una huella dactilar aumentada de una muestra de *T. diffusa*, se señalan con un asterisco los picos relevantes para la predicción del valor de EC<sub>50</sub>.

#### 6.4. Conclusiones del capítulo.

El enfoque de toma de huellas dactilares cromatográficas aumentadas utilizado en este trabajo, junto con PLSR, produce muy buenos resultados en la predicción de la actividad antioxidante a partir de los datos cromatográficos complejos de las muestras de damiana. Los cromatogramas de cuatro longitudes de onda (216, 238, 254 y 345 nm) fueron compilados en un vector de datos único para construir la huella dactilar aumentada.

El gráfico de contorno es una herramienta útil para la selección de las longitudes de onda de trabajo adecuadas para la construcción de las huellas dactilares aumentadas; los pretratamientos apropiados pueden eliminar las variaciones no deseadas de línea base.

La alineación de cromatogramas complejos compilados a partir de varias longitudes de onda fue posible empleando alineación optimizada mediante correlación (COW), evitando la introducción de ruido y de artefactos matemáticos.

El modelo PLSR construido explica una gran cantidad de la varianza de X e Y (52.5%, 98.4%; respectivamente); con 4 LVs se obtuvo un valor de 6.02 para el RMSECV. Para evaluar la fiabilidad, el modelo se aplicó a un conjunto de predicción externa, obteniendo un RSEP del 7.8%; casi tres veces inferior al obtenido utilizando sólo la longitud de onda de 254 nm.

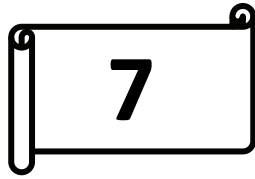
Un estudio de la puntuación de la importancia de las variables de regresión en la proyección y de los loadings de LV1 (con mucho la variable latente más importante para explicar los valores de Y) hace que sea posible encontrar qué picos cromatográficos son más importantes para la regresión en las longitudes de onda elegidas, lo cual es una pista para encontrar compuestos que están relacionados con la actividad antioxidante.

## Referencias.

- (1) Garza-Juárez, A.; Waksman-De-Torres, N.; Ramírez-Durón, R.; Salazar Cavazos, M. Development and validation of fingerprints of *Turnera diffusa* extracts obtained by use of high-performance liquid chromatography with diode array detection and chemometric methods. *Acta Chromatographica* 2009, 21, 217-235.
- (2) Alcaraz-Meléndez, L.; Real-Cosío, S.; Robert, M. L. Morphological comparison of damiana (*Turnera diffusa*, Willd.) regenerated in vitro from leaves cultured in solidified medium and liquid cultures. *Scientia Horticulturae* 2002, 96, 293-301.
- (3) Alcaraz-Meléndez, L.; Delgado-Rodríguez, J.; Real-Cosío, S. Analysis of essential oils from wild and micropropagated plants of damiana (*Turnera diffusa*). *Fitoterapia* 2004, 75, 696-701.
- (4) Salazar, R.; Pozos, M. E.; Cordero, P.; Perez, J.; Salinas, M. C.; Waksman, N. Determination of the Antioxidant Activity of Plants from Northeast Mexico. *Pharm. Biol. (N. Y., NY, U. S. )* 2008, 46, 166-170.
- (5) Dudonné, S.; Vitrac, X.; Coutière, P.; Woillez, M.; Mérillon, J. Comparative Study of Antioxidant Properties and Total Phenolic Content of 30 Plant Extracts of Industrial Interest Using DPPH, ABTS, FRAP, SOD, and ORAC Assays. *J. Agric. Food Chem.* 2009, 57, 1768-1774.
- (6) BLOKHINA, O.; VIROLAINEN, E.; FAGERSTEDT, K. V. Antioxidants, Oxidative Damage and Oxygen Deprivation Stress: a Review. *Annals of Botany* 2003, 91, 179-194.
- (7) Perez-Meseguer, J.; Garza-Juarez, A.; Salazar-Aranda, R.; Salazar-Cavazos, M. L.; de la Torre Rodriguez, Y. C.; Rivas-Galindo, V.; Waksman de Torres, N. Development and validation of an HPLC-DAD analytical procedure for quality control of damiana (*Turnera diffusa*), using an antioxidant marker isolated from the plant. *J. AOAC Int.* 2010, 93, 1161-1168.
- (8) Ramirez-Durón, R.; Cenicerros-Almaguer, L.; Salazar-Aranda, R.; Salazar-Cavazos, Ma.de L.; Waksman de Torres, N. Evaluation of thin-layer chromatography methods for quality control of commercial products containing *Aesculus hippocastanum*, *Turnera diffusa*, *Matricaria recutita*, *Passiflora incarnata*, and *Tilia occidentalis*. *J. AOAC Int.* 2007, 90, 920-924.
- (9) Dumarey, M.; van Nederkassel, A. M.; Deconinck, E.; Vander Heyden, Y. Exploration of linear multivariate calibration techniques to predict the total antioxidant capacity of green tea from chromatographic fingerprints. *Journal of Chromatography A* 2008, 1192, 81-88.
- (10) van Nederkassel, A. M.; Daszykowski, M.; Massart, D. L.; Vander Heyden, Y. Prediction of total green tea antioxidant capacity from chromatograms by multivariate modeling. *Journal of Chromatography A* 2005, 1096, 177-186.
- (11) Garza-Juarez, A.; de la Luz Salazar-Cavazos, M.; Salazar-Aranda, R.; Perez-Meseguer, J.; Waksman de Torres, N. Correlation between Chromatographic Fingerprint and Antioxidant Activity of *Turnera diffusa* (Damiana). *Planta Med.* 2011, 77, 958-963.
- (12) Leu, S. J.; Lin, Y. P.; Lin, R. D.; Wen, C. L.; Cheng, K. T.; Hsu, F. L.; Lee, M. H. Phenolic constituents of *Malus doumeri* var. *formosana* in the field of skin care. *Biol. Pharm. Bull.* 2006, 29, 740-745.
- (13) Skov, T.; van den Berg, F.; Tomasi, G.; Bro, R. Automated alignment of chromatographic data. *Journal of Chemometrics* 2006, 20, 484-497.

- (14) Tomasi, G.; van den Berg, F.; Andersson, C. Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. *Journal of Chemometrics* 2004, 18, 231-241.
- (15) Mia, H. In *Robust Calibration; Practical Guide To Chemometrics*, Second Edition; CRC Press: 2006; pp 167-216.
- (16) Daszykowski, M.; Walczak, B. Use and abuse of chemometrics in chromatography. *TrAC Trends in Analytical Chemistry* 2006, 25, 1081-1096.
- (17) Eigenvector Documentation, Using Cross-Validation,  
[http://wiki.eigenvector.com/index.php?title=Using\\_Cross-Validation](http://wiki.eigenvector.com/index.php?title=Using_Cross-Validation), con acceso en Junio de 2012.
- (18) Otto, M.; Wegscheider, W. Spectrophotometric multicomponent analysis applied to trace metal determinations. *Anal. Chem.* 1985, 57, 63-69.





## **Conclusiones Generales.**

---



Se han desarrollado y validado métodos para el control de calidad de cuatro plantas medicinales con amplio uso comercial, para las cuales existen informes de confusiones, adulteraciones o presentación de reacciones adversas en los consumidores.

Además de los criterios empleados tradicionalmente para comparar los resultados de los modelos de clasificación, los cuales se limitan a indicar el número de muestras clasificadas de forma correcta o incorrecta, se han utilizado otros que permiten obtener mayor información cuando se analizan los resultados. Se han establecido criterios de medición claros para la comparación de la efectividad de los diferentes modelos de clasificación construidos. En particular, la tasa general de éxito en la clasificación, al ser un parámetro que evalúa la efectividad global de los modelos, es un criterio sumamente útil.

Se han ensayado los modelos de clasificación propuestos con mezclas formadas por el producto de interés y diferentes proporciones y tipos de adulterantes, consiguiendo obtener información sobre la capacidad de detección de adulteraciones de los distintos métodos utilizados y la proporción mínima de adulterante discriminada. Esto representa un avance frente a la forma convencional de emplear los métodos de clasificación, donde regularmente se acota el estudio a la diferenciación entre especies.

La espectroscopia NIR ha demostrado ser una técnica útil para realizar la identificación de las plantas medicinales. En el desarrollo de un método de clasificación con esta espectroscopia se deben incluir muestras que sean representativas del material herbal a analizar y que aporten variabilidad a los modelos matemáticos, de esta manera se pueden obtener buenas predicciones de las muestras desconocidas. Para las plantas medicinales analizadas con esta metodología (*E. senticosus* y *P. ginseng*), las bandas espectrales procedentes del agua, no aportan una variabilidad útil para la construcción de los modelos de clasificación, por lo que esta información debe ser removida para obtener modelos con un resultado satisfactorio. El porcentaje de

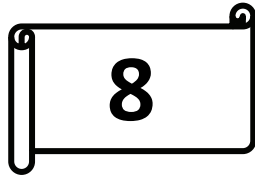
adulterante detectado en el descubrimiento de muestras adulteradas fue dependiente tanto de la especie de planta presente en la mezcla como del modelo de clasificación utilizado. Para las dos especies estudiadas con espectroscopia NIR, DA proporciona una capacidad de detección de adulteraciones menor que empleando SIMCA o PLS-DA.

La cromatografía de líquidos de alta resolución acoplada a un detector de diodos es una herramienta muy versátil para el control de calidad de las plantas medicinales, ya que proporciona más información (cromatogramas y espectros) que la espectroscopia de infrarrojo cercano (sólo espectros). Aprovechando esta característica de CLAR, fue posible incrementar la información de las huellas dactilares combinando cromatogramas registrados a diferentes longitudes de onda. El enfoque de múltiples longitudes de onda representa un incremento en la información proporcionada por la variabilidad química de la muestra, debido al incremento en el número de picos y a la minimización de la variabilidad no informativa de los datos. Por ello, es una buena alternativa para mejorar los resultados tanto de una clasificación compleja, como para la predicción de la actividad biológica, tal como fue observado para los casos de *V. officinalis* y *T. diffusa*. Para obtener buenos resultados, ya sea usando múltiples longitudes de onda como a una única longitud de onda, es imperativo corregir las variaciones presentadas de un análisis a otro mediante la aplicación de pretratamientos. En el caso de la alineación mediante COW, la correcta selección de los parámetros (vector de referencia, longitud de segmento y grado de flexibilidad) es indispensable para evitar afectar las características de los cromatogramas, previniendo la introducción de artefactos matemáticos.

La aplicación de quimiometría es fundamental para potenciar las herramientas de control de calidad de las plantas medicinales. La técnica quimiométrica a usar depende no sólo del propósito para el cual fue estructurado el algoritmo, también depende del material herbal sometido a análisis. Este hecho quedo reflejado en la clasificación del *P. ginseng*; para esta raíz medicinal, el mejor método para controlar su calidad fue SIMCA, mientras que para *E. senticosus* y *V. officinalis* la técnica más adecuada fue PLS-DA. Los diferentes pretratamientos disponibles y la posibilidad de editar las matrices de datos potencian la calidad y robustez de los resultados

obtenidos, tal como fue observado con la importante mejora en la predicción de la actividad antioxidante ( $EC_{50}$ ) de *T. diffusa* cuando fueron empleados los pretratamientos adecuados de los datos y múltiples longitudes de onda en la construcción del vector usado como huella dactilar cromatográfica.





## **Perspectivas.**

---





Este trabajo abre la posibilidad de desarrollar una estrategia para realizar el control de calidad total de los productos a base de plantas medicinales. El mismo perfil cromatográfico o espectroscópico empleado para lograr la identificación de la especie correspondiente, puede ser utilizado en la predicción de sus propiedades terapéuticas. Sin embargo, para lograr esta aplicación es necesario desarrollar métodos que permitan medir objetivamente la propiedad de interés; lo cual no siempre resulta sencillo, en especial en las determinaciones *in vivo*. En el caso de las propiedades que puedan ser evaluadas *in vitro*, debe realizarse un exhaustivo estudio fitoquímico con el fin de poder relacionar el perfil químico (componentes específicos) con la propiedad medicinal de interés; este es el caso de *T. diffusa*. Esta forma múltiple de aplicar las huellas dactilares, permitiría asegurar la llegada al usuario de un producto que no sólo es seguro, al recibir la planta correcta, sino también eficaz, ya que su actividad terapéutica ha sido evaluada. Asimismo, estas estrategias pueden ser aplicadas en la detección de señales analíticas relacionadas con una acción sinérgica o antagonista de la actividad biológica en estudio presentada por los múltiples constituyentes de un extracto; por ejemplo, picos específicos en una huella dactilar cromatográfica.

En esta tesis se utilizaron múltiples longitudes de onda para la construcción de un vector cromatográfico que fue empleado como huella dactilar aumentada; obteniendo buenos resultados. No obstante, el uso simultáneo de la información de los cromatogramas y los espectros contenida en una matriz de datos en tres dimensiones, es un área de oportunidad para el desarrollo de métodos de control de calidad de productos fitoterapéuticos a través de métodos quimiométricos multidimensionales.

La continuación de esta línea de investigación se enfrenta al reto de la identificación de especies de plantas con perfiles químicos aún más semejantes que los encontrados en esta Tesis, que probablemente necesiten de técnicas quimiométricas de clasificación más avanzadas como máquinas de soporte vectorial (SVM, support vector machine), redes neuronales y otros.



**Anexo.**

---



**Los trabajos realizados en esta tesis han dado lugar a los siguientes artículos:**

1. Application of near infrared spectral fingerprinting and pattern recognition techniques for fast identification of *Eleutherococcus senticosus*.  
J. Ricardo Lucio-Gutiérrez, J. Coello, S. MasPOCH.  
Departament de Química, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Barcelona, Spain.  
*Food Research International* 44 (2011) 557–565.
2. Enhanced chromatographic fingerprinting of herb materials by multi-wavelength selection and chemometrics.  
J. Ricardo Lucio-Gutiérrez, J. Coello, S. MasPOCH.  
Departament de Química, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Barcelona, Spain.  
*Analytica Chimica Acta* 710 (2012) 40– 49.
3. Multi-wavelength high-performance liquid chromatographic fingerprints and chemometrics to predict the antioxidant activity of *Turnera diffusa* as part of its quality control.  
J. Ricardo Lucio-Gutiérrez<sup>a</sup>, Aurora Garza-Juárez<sup>b</sup>, J. Coello<sup>a</sup>, S. MasPOCH<sup>a</sup>, M.L. Salazar-Cavazos<sup>b</sup>, Ricardo Salazar-Aranda<sup>b</sup>, Noemi Waksman de Torres<sup>b</sup>.  
<sup>a</sup> Departament de Química, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Barcelona, Spain.  
<sup>b</sup> Departamento de Química Analítica, Facultad de Medicina, Universidad Autónoma de Nuevo León, C.P. 64460, Monterrey, Nuevo León, Mexico.  
*Journal of Chromatography A*, 1235 (2012) 68– 76.
4. Expeditious identification and semi-quantification of *Panax ginseng* using near infrared spectral fingerprints and multivariate analysis.  
J. Ricardo Lucio-Gutiérrez, J. Coello, S. MasPOCH.  
Departament de Química, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Barcelona, Spain.  
(Enviado).