FACULTAT DE QUÍMICA

Design of strategies for optimisation in Liquid Chromatography

Memoria para alcanzar el Grado de Doctor en Química dentro del Programa de Doctorado en Química (RD 1999/2011) presentada por:

José Antonio Navarro Huerta

Directores:
Dra. María Celia García Álvarez-Coque
Dr. José Ramón Torres Lapasió

Burjassot, Valencia, Abril 2021

Dña. MARÍA CELIA GARCÍA ÁLVAREZ COQUE, Catedrática de Universidad, D. JOSÉ RAMÓN TORRES LAPASIÓ, Profesor Titular de Universidad, adscritos al Departamento de Química Analítica de la Universidad de Valencia,

CERTIFICAN

Que la presente Memoria, "Design of strategies for optimisation in Liquid Chromatography", constituye la Tesis Doctoral de

D. JOSÉ ANTONIO NAVARRO HUERTA

Asimismo, certifican haber dirigido y supervisado tanto los distintos aspectos del trabajo, como su redacción.

Y para que conste a los efectos oportunos, firmamos la presente en Burjassot, a siete de abril de dos mil veintiuno.

María Celia García   José Ramón Torres Lapasió
   Álvarez-Coque

ACKNOWLEDGEMENTS

I would like to thank all those who have helped me to carry out, in one way or another, the research work presented here.

I also wish to express my gratitude to the FUSCHROM Group (Fundamental Studies in Chromatography) of the University of Valencia, for having made available to me all the necessary means to carry out my research work and the funds to attend to several research meetings. Above all, to my PhD supervisors, Professors María Celia García Álvarez-Coque and José Ramón Torres Lapasió. Professor García Álvarez-Coque has always helped me from the beginning, when I met her during the degree of Chemistry. She encouraged me to be enrolled in the Master of Experimental Techniques in the Department of Analytical Chemistry and the PhD. Program of the Faculty of Chemistry of the University of Valencia, to become an expert in liquid chromatography. During these years, she has taught me the importance of carrying out the work attentively to get good results, and that constant effort always has its reward. She has been an example to follow for her hard work, perseverance, dedication, and optimism, and I thank

stays more enjoyable: Paola Arena, Domenica, Adriana, Ivan, Barbara, Mario, Jassin, and Filippo, from Messina, and Honorine, Bastiaan, Gioele, Víctor, and Eulalia, from Geneva. I would like to thank especially to Luis Enrique for all the good time spent together, and having always been available to travel with me in Switzerland, despite the situation.

I want to express my sincere gratitude to my labmates and friends in the Department of Analytical Chemistry: Ester, Tamara, Nikita, Adrián, Joan and Noemí. Thanks for willingly helping me out with your abilities and knowledge, and all the good time we spent together.

I would like also to extent my appreciation to all the people who have contributed in some way to the development of my PhD. research: Enrique Javier, María Vergara, Carolina, Carmen, and other collegues from the Department of Analytical Chemistry. Also to Sergio López from the Department of Mathematics and Gamaliel from the Universidad Autónoma de México. I thank all of you for your valuable contributions to my PhD. Project.

Finally, I would like also to dedicate some words to all those people who, despite not having participated in the development of my research work, have been a fundamental pillar during this period, and my life:

To my friends Joan, Borja and Marina, for showing me that despite time and distance, they will always be willing to listen to me and advise me wisely. To my friends from Pinedo: Mónica, Anna, Guillem, Vicente, Marta, Miriam, Andrea, Leticia, Raquel, Blanca, and Elian, for the support and the experiences lived together. To my university friends: Libertad, Marta, Pau, and Esther, for having always been by my side and allowing me to share this experience with you.

To my family, particularly my parents, Montse and Toni; my brother, Javier; and my sister, Ángela, for your full support and love. They have always encouraged me to keep going, even in the most difficult situations. This PhD. Project would have not been possible without you.

To Luis Guillermo, who has been my partner during all this time. He has been the cornerstone of my life and has helped me to persist, never letting me give up. He has also contributed to the development of my PhD. Project with exciting discussions and his advice.

All of them, I thank very much.

ABBREVIATIONS

A: Left peak half-width

ACN: Acetonitrile

AIBN: 2,2-azobis[2-methylproprionitrile]

B: Right peak half-width

BEADS: Baseline Estimation And Denoising using Sparsity

BHT: 3,5-di-tert-butyl-4-hydroxy-toluene

Brij-35: Polyoxyethylene(23)lauryl ether

CMC: Critical micellar concentration

COF: Chromatographic objective function

CPA: Consecutive peaks approximation

CTBA: Cetyltrimethylammonium bromide

C4: Butylsiloxane

C18: Octadecylsiloxane

C8: Octylsiloxane

C30: Triacontylsiloxane

DAD: Diode array detection

DMF: Dimethylformamide

DNA: Deoxyribonucleic acid

DOE: Design of experiments

DVB: Divinylbenzene

EDMA: Ethylene dimethacrylate

ELS: Evaporative light scattering

GA: Genetic algorithms

HAcO: Acetic acid

HEMA: Hydroxyethyl methacrylate

HILIC: Hydrophilic interaction liquid chromatography

HMA: Hexyl methacrylate

HPLC: High performance liquid chromatography

HSLC: High submicellar liquid chromatography

i-PrOH: Isopropanol

i.d.: Internal diameter

IIC: Ion-interaction chromatography

IL: Ionic liquid

IPC: Ion-pair chromatography

k: Retention factor

k_{w}: Retention factor with water

LC: Liquid chromatography

LC×LC: Comprehensive two-dimensional liquid chromatography

LC-LC: Heart-cutting two-dimensional liquid chromatography

LDA: Linear discriminant analysis

LMA: Lauryl methacrylate

LOF: Lack of fit

LSER: Linear solvation energy relationship

LSS: Linear solvent strength

M: Molar concentration (mol/L)

MAA: Methacrylic acid

MELC: Microemulsion liquid chromatography

MeOH: Methanol

MLC: Micellar liquid chromatography

MS: Mass spectrometry

N: Plate number

NAC: N-Acetylcysteine

NaOH: Sodium hydroxide

NK: Neue-Kuss

NPLC: Normal-phase liquid chromatography

ODS: Octadecylsilane

OPA: o-Phthalaldehyde

P': Snyder's global polarity

P_C: Peak capacity

PFP: Pentafluorophenyl

pK_a: Acid dissociation constant

$P_{o/w}$: Octanol-water partition coefficient

R: Correlation coefficient

R_{Adj}: Adjusted correlation coefficient

RPLC: Reversed-phase liquid chromatography

SDS: Sodium dodecyl sulphate

SPSS: Statistical Package for the Social Sciences

SST: Solvent-selectivity triangle

TFA: Trifluoroacetic acid

THF: Tetrahydrofuran

t_{ext}: Extra-column time

t_G: Gradient time

t_g: Retention time in gradient elution

t_R: Retention time

t_0: Dead time

t_D: Dwell time

v/v: Volume/volume

w/w: Weight/weight

UV: Ultraviolet

^{1}D: First dimension

1D-LC: One-dimensional liquid chromatography

^{2}D: Second dimension

2D: Two-dimensional

2D-LC: Two-dimensional liquid chromatography

βAA: β-adrenoceptor antagonists

γ-MPS: 3-[trimethoxysilyl]propyl methacrylate

RESUMEN

La cromatografía líquida de fase inversa (RPLC, *reversed-phase liquid chromatography*) es la técnica más utilizada para el análisis de compuestos orgánicos, en un amplio intervalo de estructuras e hidrofobicidades, debido a su versatilidad, robustez y sensibilidad. Sin embargo, la selectividad y el tiempo de análisis dependen de una forma compleja de varios factores experimentales que interaccionan entre sí, como son la concentración de disolvente orgánico, el pH y la temperatura. Debido a la dificultad de la búsqueda de las condiciones experimentales que permitan la separación simultánea de todos los compuestos en una muestra, las optimizaciones basadas en ensayo y error resultan muy laboriosas, y en ocasiones, no son satisfactorias. Además, no ofrecen garantías de conducir al verdadero óptimo.

Las mejores condiciones de separación se deben hallar, preferiblemente, utilizando la información extraída de un conjunto reducido de experimentos cuidadosamente planificados, que cubran todo el espacio de interés para los factores experimentales. Los datos obtenidos se utilizan con el propósito de ajustar un modelo de retención para cada analito, que permite predecir los tiempos de retención para cualquier nueva condición arbitraria, dentro del dominio experimental y simular cromatogramas. Ello permite, finalmente, seleccionar las mejores condiciones mediante el uso de metodologías asistidas por ordenador, en las denominadas optimizaciones interpretativas. Los modelos ajustados también pueden proporcionar información sobre las interacciones establecidas dentro de la columna cromatográfica.

El trabajo de Tesis Doctoral realizado incluye estudios fundamentales para mejorar las metodologías de optimización interpretativa y su aplicación al análisis de fluidos fisiológicos y productos naturales (hojas y pulpa de olivo y hierbas medicinales). Se consideró la determinación de varios grupos de compuestos: alquilbencenos, sulfonamidas, antagonistas de los receptores

β-adrenérgicos, aminoácidos, fenoles y polifenoles, así como compuestos desconocidos en un amplio intervalo de polaridades contenidos en las hierbas medicinales analizadas. La mayoría de los análisis se realizaron con fases móviles de acetonitrilo-agua en elución isocrática y de gradiente, pero también se investigó el efecto de la presencia de equilibrios secundarios cuando se añade un tensioactivo a la fase móvil.

A lo largo del trabajo, se desarrollaron nuevas estrategias y herramientas, algunas de ellas sin antecedentes previos, lo que requirió la construcción de software diverso. El rendimiento de los nuevos desarrollos se comparó, cuando fue posible, con otros publicados anteriormente. El trabajo efectuado durante la Tesis Doctoral aparece expuesto en la Memoria en dos grandes apartados, que recogen diversos desarrollos relacionados con: (i) el incremento de la capacidad de modelización en cromatografía líquida, y (ii) la mejora del rendimiento en la separación de los picos en huellas dactilares cromatográficas. A continuación, se detallan los estudios realizados.

1. Incremento de la capacidad de modelización en cromatografía líquida

La fiabilidad de las estrategias interpretativas depende en gran medida de la exactitud de los modelos utilizados en la predicción de los tiempos de retención y perfiles de los picos cromatográficos, que se construyen a partir de la información obtenida de estándares de los analitos. La Memoria de Tesis Doctoral reúne varias contribuciones dedicadas a la optimización de los diseños experimentales empleados en la construcción de modelos. También contiene varias propuestas sobre su aplicación a la obtención de información sobre las interacciones que tienen lugar en el interior de una columna cromatográfica, la estimación de la capacidad de pico tanto en elución isocrática como en

gradiente, y la optimización de gradientes que utilizan eluyentes que contienen un tensioactivo en condiciones micelares o submicelares. Los aspectos más relevantes de cada propuesta se describen a continuación.

1.1. Obtención de información sobre las interacciones soluto-fase estacionaria

Se prepararon y ensayaron varias columnas monolíticas poliméricas, con un contenido variable de monómeros hidrofóbicos e hidrofílicos, utilizando compuestos apolares (alquilbencenos) y polares (sulfonamidas) como compuestos de prueba. Las columnas incluidas en el estudio fueron las siguientes: una columna formada con lauril-metacrilato (LMA), que le confiere un carácter hidrófobo dominante; una columna de polaridad intermedia con una mezcla de monómeros hidrófobos (LMA) e ionizables (ácido metacrílico, MAA, *methacrylic acid*); y una columna con un monómero más polar (metacrilato de hexilo, HMA, *hexyl methacrylate*), combinado con MAA.

Se seleccionó como fase estacionaria un monolito compuesto de HMA, MAA y dimetacrilato de etileno (EDMA*, ethylene dimethacrylate*), en base a la mejor resolución cromatográfica alcanzada y tiempos de análisis razonables, para los dos conjuntos de compuestos de prueba. A pesar de la presencia de grupos de ácido metacrílico de polaridad moderada en la columna monolítica de poli(HMA-co-MAA-co-EDMA), el orden de elución y la distribución regular de los tiempos de retención observados para los alquilbencenos (con carácter apolar) demostró la importancia de las interacciones hidrofóbicas. Por el contrario, el comportamiento de las sulfonamidas (con carácter polar) fue irregular, distribuyéndose los compuestos en tres grupos según su retención, mostrando coelución en la mayoría de las condiciones experimentales

ensayadas, con inversiones en los tiempos de retención a elevados contenidos de disolvente orgánico. Sin embargo, la resolución de las sulfonamidas mejoró muy significativamente, respecto a columnas monolíticas previas.

Se analizó el comportamiento cromatográfico de los compuestos de prueba, con la columna monolítica seleccionada, modelizando los tiempos de retención y los perfiles de los picos. Se estudió la exactitud de varios modelos de retención (Ecuaciones (2.1) a (2.12)), entre los que se incluyó un modelo que describe un mecanismo de retención mixto. Los parámetros ajustados para este modelo sugirieron que el mecanismo de retención se basaba principalmente en la adsorción, para los dos conjuntos de compuestos (alquilbencenos y sulfonamidas). Todos los modelos ensayados proporcionaron predicciones aceptables, con errores relativos a menudo inferiores al 1.0%. El rendimiento de los modelos para la columna monolítica fue similar o superior al encontrado con columnas de RPLC convencionales, cuando se analizan los mismos compuestos.

Se obtuvo información sobre el comportamiento de retención de las sulfonamidas con la columna monolítica, a partir de las correlaciones entre los parámetros del modelo logarítmico-cuadrático que incluye la transformación $P_{\mathrm{M}}^{\mathrm{N}}$ (S_1, S_2 y q en la Ecuación (2.7)), en lugar del contenido de disolvente orgánico del modelo de retención clásico. La elevada dispersión observada en las correlaciones entre los parámetros S_1 y S_2 del modelo (que cuantifican la fuerza eluyente de la fase móvil y la desviación del modelo de la linealidad), y la ordenada en el origen q (que cuantifica el nivel de retención de los solutos), indicó una variabilidad significativa en el comportamiento de retención de las distintas sulfonamidas, respecto al que experimentan los alquilbencenos. Esto se puede explicar por la existencia de diferentes proporciones de interacciones hidrofílicas e hidrofóbicas, en sulfonamidas con diferentes estructuras

moleculares, con los monómeros polares y apolares del monolito utilizado como fase estacionaria.

Las correlaciones de las semianchuras de los picos con los tiempos de retención, para los cromatogramas obtenidos con las columnas monolíticas y C18 convencionales, revelaron también la diversidad de interacciones para los alquilbencenos y sulfonamidas estudiados. La significativa dispersión observada en la correlación de las semianchuras derechas de los picos, para las sulfonamidas analizadas con la columna monolítica, indicó cinéticas particulares para diversos compuestos, lo que se debe interpretar de nuevo por la diversa participación de monómeros polares y apolares en la columna monolítica, cuando interactúan con las sulfonamidas.

1.2. *Búsqueda de diseños experimentales óptimos*

Los diseños experimentales isocráticos proporcionan la información más rica posible sobre el comportamiento de los solutos, para realizar el ajuste de modelos de retención con parámetros ofreciendo la máxima exactitud, con intervalos de confianza estrechos. Sin embargo, el uso de diseños isocráticos se ve obstaculizado por los largos tiempos de retención de los solutos más apolares, en mezclas con otros analitos, especialmente a bajos contenidos de disolvente orgánico. La solución habitual es utilizar diseños experimentales formados con gradientes de disolvente orgánico, en los que su concentración se incrementa gradualmente para reducir los tiempos de retención. Sin embargo, los diseños que incluyen experiencias de gradiente originan modelos de retención menos exactos y, en consecuencia, su rendimiento en la realización de predicciones es más deficiente.

Se exploró, como alternativa, el uso de diseños experimentales isocráticos, en los que se incluyen incrementos bruscos de disolvente orgánico (i.e., pulsos), en las fases móviles de menor fuerza eluyente. Las experiencias con pulsos se pueden considerar como un tipo de gradientes multi-isocráticos, que permiten obtener información cromatográfica para solutos apolares, eluidos con fases móviles que contienen un bajo contenido de disolvente orgánico. El efecto del pulso es trasladar en bloque los tiempos de retención de compuestos con elución tardía en elución isocrática, hacia tiempos más cortos. Los solutos más rápidos eluyen antes del pulso, y los solutos más retenidos tras el pulso en tiempos de retención aceptables.

Este tipo de diseño mixto se puede construir fácilmente, reemplazando las fases móviles isocráticas más lentas por otras que contienen uno o dos pulsos de corta duración, situadas a tiempos intermedios. La ubicación del pulso puede ser arbitraria, pero la mejor opción es situarlo en una región intermedia vacía del cromatograma. Debe tenerse en cuenta que las fases móviles que incorporan un pulso presentan un efecto importante sobre la selectividad y retención de los solutos que eluyen tras los pulsos. Por ello, la posición, duración e incremento en el contenido de disolvente orgánico del pulso debe adaptarse a cada muestra analizada.

La inclusión de pulsos no es práctica con fines de optimización, debido al mayor solapamiento de los picos, especialmente en la región del pulso, y a la fuerte caída en la eficacia de los picos que eluyen después del pulso. Sin embargo, se obtienen beneficios en la modelización de la retención.

Las predicciones de las condiciones de elución, para las fases móviles que contienen pulsos, se realizó utilizando la ecuación fundamental para la elución en gradiente. Se observó que los tiempos de retención calculados numéricamente mostraban desviaciones notables para los solutos eluidos cerca

del pulso, incluso cuando se utilizaba un modelo de retención con bajo error de predicción. Cuando se tuvo en cuenta el retardo intra-columna (i.e., tiempo necesario para que el frente del disolvente alcance al soluto desde la entrada de la columna), las predicciones mejoraron y los cromatogramas predichos coincidieron muy satisfactoriamente con los experimentales.

Cuando las predicciones realizadas a partir de diseños que contienen pulsos o gradientes se llevaron a cabo dentro del dominio experimental, la diferencia entre los tiempos predichos y experimentales fue inferior a 0.01 min. Los diseños con pulsos proporcionaron parámetros de los modelos de retención estudiados similares a los obtenidos con los diseños isocráticos, que como se ha comentado, son considerados los más exactos para realizar predicciones. Se verificó que los diseños con un solo pulso fueron los más exactos. Para las predicciones fuera del dominio experimental, la capacidad predictiva de los diseños que contienen pulsos también fue similar a la proporcionada por los diseños con experimentos puramente isocráticos.

En general, los diseños que contienen pulsos demostraron ser muy competitivos respecto a los diseños de gradientes, en términos de tiempo de análisis y consumo de disolvente. Aunque los diseños de gradiente con tiempo de gradiente variable ofrecieron tiempos de análisis más cortos y un menor consumo de disolvente orgánico, dieron lugar al mayor error en los parámetros de los modelos de retención y mayores desviaciones en los tiempos de retención extrapolados.

Por otro lado, muchos analistas prefieren el uso de gradientes frente a las fases móviles isocráticas, no sólo para la realización de los análisis, sino también para la construcción de diseños experimentales con fines de modelización. Sin embargo, encontrar un diseño con una distribución óptima de gradientes no es sencillo. Con el fin de encontrar los mejores diseños

experimentales (formados por experiencias isocráticas o de gradiente), se desarrolló una metodología universal que permite evaluar su calidad.

La metodología desarrollada utiliza el principio de optimalidad G (*G-optimality principle*), que se basa en la teoría de propagación de errores, y relaciona las propiedades matemáticas de un modelo de retención con una determinada distribución de puntos en un diseño experimental. Se basa en la estimación de la varianza asociada a la predicción de tiempos de retención, utilizando una expresión que considera dos matrices jacobianas asociadas a experimentos de entrenamiento ($\mathbf{J}_{train}$) y de muestreo ($\mathbf{J}_{pred}$) (Ecuación (4.19)). Las matrices jacobianas implican el cálculo de derivadas parciales de los modelos de retención, para un gran conjunto de condiciones experimentales. Para elución en gradiente, el tiempo de computación puede ser inasumible, ya que requiere la predicción del tiempo de retención mediante la integración de la ecuación fundamental, lo que puede implicar cálculos masivos. En la Memoria de Tesis Doctoral, se muestra el desarrollo de una metodología práctica, que reduce el tiempo de computación apreciablemente, aprovechando desarrollos recientes realizados en el laboratorio del grupo investigador.

Se validó la metodología propuesta verificando la calidad de cinco diseños de entrenamiento, muy utilizados en RPLC para construir modelos para predecir la retención de 14 sulfonamidas de diversa polaridad, considerando diseños de muestreo en elución isocrática y de gradiente. Se comprobó que el modelo de retención propuesto por Neue-Kuss proporciona una mayor exactitud en las predicciones, en comparación al modelo lineal de fuerza eluyente del disolvente (LSS, *linear solvent strength*), con errores relativos de predicción por debajo del 0.7%. Se encontró que el modelo LSS, que se utiliza ampliamente para elución en gradiente, produce falta de ajuste, por lo que se descartó.

Para comparar el rendimiento de los diseños de entrenamiento, se hizo uso de gráficos en los que se representó el valor de las incertidumbres relativas en las predicciones, para las experiencias en los diseños de muestreo con cada compuesto analizado. Las incertidumbres se representaron para los diseños isocráticos frente a la composición de la fase móvil, con incrementos de acetonitrilo del 1%, y para los diseños de gradiente frente a la pendiente de la rampa de un gradiente lineal, con incrementos angulares constantes de 3º. Las incertidumbres relativas proporcionaron resultados más significativos e interpretables que las incertidumbres absolutas, que presentaron fuertes variaciones dependiendo de la retención de los solutos.

Un factor crítico para el cálculo de las derivadas de las matrices jacobianas es el nivel de exactitud en el cálculo del tiempo de retención en gradiente. Con un nivel de exactitud insuficiente, se obtienen gráficos de incertidumbre con curvas afectadas de mucho ruido, requiriéndose un nivel de exactitud de aproximadamente 10^{-15} para obtener curvas exentas de ruido. En la mayoría de los casos, para la elución en gradiente, se obtuvo un patrón en U característico, con incrementos en ambos extremos y errores más bajos en la zona intermedia. Para todos los diseños de entrenamiento estudiados, las regiones intermedias en los graficos de incertidumbre mostraron un cambio sistemático al disminuir la polaridad de los solutos. La magnitud de la incertidumbre mínima fue similar para las experiencias isocráticas y de gradiente. Sin embargo, los gradientes se predijeron generalmente con incertidumbres más bajas para cualquier diseño experimental, y fueron menos sensibles a la composición de la fase móvil que las predicciones isocráticas.

Se confirmó que el mejor diseño de entrenamiento, en la predicción de fases móviles isocráticas y gradientes, es el formado por un conjunto de experiencias isocráticas concentradas gradualmente hacia bajos contenidos de disolvente

orgánico (ISO1). Por el contrario, el rendimiento de los diseños de gradiente con tiempo de gradiente fijo y contenido final de disolvente orgánico variable (G1), o contenido final fijo y tiempo de gradiente variable (G2), fue insatisfactorio en la mayoría de situaciones, siendo sólo aceptable para los eluyentes más lentos y los solutos más rápidos. El diseño G3, que combina algunas características de los diseños G1 y G2, proporcionó un rendimiento razonablemente bueno para todos los compuestos de prueba, sólo superado por el diseño ISO1.

1.3. Estimación de la capacidad de pico en base a la simulación de picos cromatográficos

La capacidad de pico es un concepto clave en el análisis cromatográfico, que se refiere al número máximo de picos que idealmente se resuelven totalmente en una ventana de tiempo determinada. En RPLC, los cromatogramas tienden a distribuciones de picos desiguales, con solapamientos entre los picos y grandes espacios vacíos. Por ello, la capacidad de pico es un concepto meramente teórico. A pesar de ello, se considera útil para evaluar las posibilidades de una columna cromatográfica para lograr la resolución de los picos, y por ello, ha llamado mucho la atención.

Varios autores han propuesto algoritmos para estimar la capacidad de pico en condiciones isocráticas. Neue propuso también un algoritmo para realizar la estimación cuando se utiliza elución en gradiente. Sin embargo, estas estimaciones tienen varias limitaciones, como ser sólo aplicables a picos simétricos, en elución isocrática y utilizando gradientes lineales ignorando el tiempo de retardo (*delay time*) y la presencia de volúmenes extra-columnares. Además, se asume que el número de platos teóricos es constante. Para superar

estas limitaciones, se desarrolló una metodología basada en la simulación de cromatogramas formados por picos de compuestos ficticios, con el mismo tipo de comportamiento que los analitos de interés cuando se analizan con una columna determinada. Los picos de los compuestos ficticios se generan a partir de predicciones de los tiempos de retención y semianchuras de pico, y a continuación, se organizan para cumplir con la definición de capacidad de pico.

La predicción de la retención de los picos se realiza utilizando modelos ajustados a partir de la información obtenida de los estándares de un conjunto de compuestos estructuralmente relacionados, de polaridad variable. La propuesta se ilustra utilizando un conjunto de 15 sulfonamidas, analizadas con tres columnas en elución isocrática y aplicando gradientes lineales y multi-lineales. El proceso se inicia generando un gran número de picos ficticios con anchuras correspondientes a sus tiempos de retención. El comportamiento de retención se obtiene de la correlación de los parámetros en el modelo logarítmico-cuadrático que incluye la transformación $P_{\mathrm{M}}^{\mathrm{N}}$, ajustado con los estándares, mientras que las anchuras de pico se predicen a partir de la correlación de las semianchuras con los tiempos de retención. Una vez generados los picos, se adapta la retención hasta conseguir su conexión a la altura requerida, generalmente asumiendo una anchura de pico de 4σ libre de solapamiento.

La metodología propuesta, basada en la simulación de cromatogramas, quedó validada comprobando la buena concordancia al superponer los cromatogramas simulados con los reales, para la mezcla de sulfonamidas en las mismas condiciones de separación. Además, se observó que los valores de capacidad pico coincidían con los estimados con las ecuaciones clásicas, y posee la ventaja frente a algoritmos anteriores de ser aplicable a una variedad

de situaciones en las que éstos no se pueden aplicar, incluida la elución con gradientes multi-lineales complejos y la presencia de picos asimétricos.

La posibilidad de simular los cromatogramas permitió la optimización de las condiciones de elución, en multitud de condiciones, de acuerdo a los valores predichos de capacidad de pico. Para ello, se construyeron gráficos de Pareto en los que se representan las predicciones para condiciones isocráticas y utilizando gradientes lineales y multi-lineales (una solución se califica como óptimo de Pareto cuando una respuesta no puede mejorarse sin empeorar otra). Como era de esperar, las separaciones isocráticas presentaron los valores más bajos de capacidad de pico, mientras que los gradientes multi-lineales ofrecieron los valores más elevados, junto a un menor tiempo de análisis. Un sistema cromatográfico no puede proporcionar valores de capacidad de pico fuera de la región limitada por la tendencia isocrática y el límite superior de elución en gradiente.

Sin embargo, se encontró para el conjunto de sulfonamidas, que las condiciones de separación que conducían a la mejor resolución estaban lejos de las que proporcionaban la máxima capacidad de pico. Esto significa que una optimización basada en la capacidad de pico sólo puede ser significativa para muestras muy complejas. Para muestras en las que el número de compuestos es relativamente pequeño, se debe atender a los requisitos de resolución específicos para cada soluto.

1.4. *Optimización interpretativa en cromatografía líquida micelar con elución isocrática y de gradiente en dominios extendidos de disolvente orgánico*

Es posible analizar compuestos en un amplio intervalo de estructuras y polaridades, mediante RPLC. Sin embargo, los compuestos orgánicos ionizados y los aniones o metales inorgánicos, que poseen una elevada polaridad, muestran poca o ninguna retención. Otros analitos pueden presentar una retención excesivamente baja o elevada. Una forma de resolver estos problemas ha sido la preparación de nuevas fases estacionarias, pero una solución más sencilla es la adición de reactivos a la fase móvil, que incorporan al sistema cromatográfico una variedad de equilibrios secundarios con las fases estacionaria y móvil.

Entre las soluciones más utilizadas en RPLC para modificar la retención utilizando aditivos, se encuentra el uso de tensioactivos en concentraciones a las que forman micelas, lo que ha dado lugar a un modo cromatográfico al que se ha denominado cromatografía líquida micelar (MLC, *micellar liquid chromatography*). Esta técnica ha mostrado utilidad, especialmente, en el análisis de muestras fisiológicas que no requieren pre-tratamiento, ya que las proteínas se solubilizan en presencia del tensioactivo y eluyen cerca del tiempo muerto. La mayoría de los procedimientos descritos en MLC hacen uso del tensioactivo aniónico dodecilsulfato sódico (SDS, *sodium dodecyl sulphate*). Dado que, para la mayoría de los solutos, la fuerza eluyente de las disoluciones acuosas de SDS es baja, se debe añadir una cantidad relativamente pequeña de disolvente orgánico a la fase móvil para disminuir la retención. En un desarrollo más reciente, la concentración de disolvente orgánico en las disoluciones de tensioactivo se incrementa para obtener tiempos suficientemente cortos, para compuestos altamente retenidos con las columnas alquil-enlazadas. Este modo cromatográfico se ha denominado cromatografía

líquida submicelar alta (HSLC, *high submicellar liquid chromatography*), ya que no se forman micelas a pesar del uso de una concentración relativamente alta del tensioactivo.

Los procedimientos descritos en MLC se implementan generalmente en el modo isocrático, ya que el problema general de elución en RPLC (i.e., el aumento exponencial de la retención al disminuir la polaridad de los solutos) es menos problemático. Sin embargo, la elución en gradiente puede también ser útil para analizar, en tiempos más cortos, mezclas de compuestos en un amplio intervalo de polaridades. Los análisis de muestras fisiológicas se pueden realizar utilizando un gradiente que se inicie con una fase móvil que contenga micelas y un bajo contenido de disolvente orgánico, a fin de proporcionar una mejor protección a la columna frente a la precipitación de las proteínas. Una vez que las proteínas se eliminan de la columna, se puede aumentar la fuerza eluyente utilizando un gradiente positivo de disolvente orgánico para reducir los tiempos de retención de compuestos altamente retenidos. Esto da lugar a la transición del modo micelar al submicelar.

Para valorar la conveniencia del uso de gradientes frente a la elución isocrática en MLC, considerando un intervalo extendido de disolvente orgánico, se requería aún desarrollar un método de optimización interpretativo para elución en gradiente, basado en la descripción exacta de la retención. Para ello, se abordó el cribado de un conjunto de ocho compuestos básicos (antagonistas de los receptores β-adrenérgicos) en muestras de orina, realizando los análisis mediante inyección directa con columnas C8 o C18 y utilizando disoluciones acuosas de SDS con disolvente orgánico añadido. Se estudió el rendimiento de tres disolventes orgánicos (acetonitrilo, etanol y 1-propanol), a concentraciones variables de SDS. Con acetonitrilo, se consiguió la resolución completa, pero el tiempo de análisis fue excesivo. El etanol y el 1-propanol

ofrecieron un tiempo de análisis aceptable, pero la resolución máxima alcanzada con el etanol fue demasiado baja. Por lo tanto, se seleccionó el 1-propanol para realizar los análisis. Por otro lado, se comparó la exactitud que ofrecían nueve modelos (algunos de ellos propuestos previamente para MLC y HSLC), para predecir la retención utilizando las concentraciones de SDS y 1-propanol como variables. Se seleccionó la Ecuación (7.11), debido a su buena capacidad predictiva en dominios extendidos de disolvente orgánico, con errores relativos entre el 0.3 y 1.7%.

Cuando se analizan muestras fisiológicas mediante inyección directa, además de los fármacos administrados, los cromatogramas contienen un pico prominente correspondiente a un compuesto endógeno que eluye a tiempos de retención relativamente cortos, del que se desconocía su identidad. Este compuesto (cuya identidad desconocíamos) debía modelizarse para ser considerado en la optimización de la resolución. La información sobre su comportamiento de retención se consiguió a partir de los picos obtenidos al inyectar orina, manteniendo la concentración de 1-propanol suficientemente baja para evitar la precipitación de las proteínas. Debido al número limitado de experiencias disponibles para este compuesto, la Ecuación (7.5) condujo a mejores resultados para modelizar su retención.

Se realizó un estudio detallado para conocer la idoneidad de las columnas C8 y C18 en el análisis de los compuestos básicos, con inyección directa de la muestra de orina utilizando fases móviles isocráticas, y gradientes lineales o multi-lineales. La optimización de las condiciones de elución en modo isocrático proporcionó una buena resolución y un tiempo de análisis razonable (alrededor de 25 min), para ambas columnas, utilizando una concentración alta de SDS y un contenido de disolvente orgánico por debajo del 15%, lo que evitó la precipitación de las proteínas de la muestra. Se observó una buena

concordancia entre los cromatogramas predichos y experimentales, para ambas columnas.

Los gradientes lineales simples dieron lugar a una reducción significativa del tiempo de análisis, respecto a la elución isocrática. Se encontró que la inclusión de una etapa isocrática inicial con una baja concentración de disolvente orgánico era perjudicial para lograr una buena resolución. Se observaron problemas de línea de base con la columna C18, lo que producía desviaciones en la predicción de las señales. Por el contrario, la concordancia entre los cromatogramas predichos y experimentales fue excelente para la columna C8. Estos comportamientos pueden explicarse por la mayor capacidad de adsorción del tensioactivo sobre la columna C18, respecto a la columna C8, que es desorbido gradualmente por el disolvente orgánico a lo largo del gradiente.

En general, la implementación de gradientes multi-lineales con eluyentes que contienen tensioactivo, y cambios repentinos en las pendientes, origina una perturbación importante de la línea base, particularmente con la columna C18. Para la columna C8, los gradientes multi-lineales redujeron significativamente el tiempo de análisis manteniendo una buena resolución, y una buena concordancia entre los cromatogramas predichos y experimentales. Por lo tanto, es preferible el uso de gradientes lineales con la columna C8 para realizar estos análisis.

2. Optimización de la separación en huellas dactilares cromatográficas

Como se ha comentado, la búsqueda de las mejores condiciones de separación en cromatografía líquida se puede realizar utilizando la información obtenida con estándares de los analitos. Sin embargo, sigue siendo un desafío la obtención de información útil para muestras que contienen una gran cantidad de compuestos. La mayor dificultad corresponde a muestras para las que no se dispone de información previa sobre su composición química, al menos para algunos compuestos. También existe la posibilidad de que no se hallen disponibles los estándares de los compuestos analizados, necesarios para predecir las condiciones óptimas de separación con las estrategias interpretativas convencionales.

Independientemente de que se conozca o no la identidad de los compuestos que originan picos en un cromatograma, su separación mutua debe ser lo mayor posible, tanto para fines cualitativos como cuantitativos. Un caso extremo es la obtención de las denominadas huellas dactilares cromatográficas, donde la distribución y magnitud relativas de los picos son las características relevantes. En estas muestras, una mejor resolución puede ofrecer cromatogramas más informativos. La Memoria de Tesis Doctoral incluye propuestas para mejorar el procesamiento de las señales en cromatogramas complejos, la estimación de la resolución en huellas dactilares de hierbas medicinales mediante cromatografía líquida mono-dimensional, y la optimización de la separación de compuestos polifenólicos en huellas dactilares de extractos de hoja y pulpa de olivo, mediante cromatografía líquida bidimensional.

2.1. Sustracción de la línea base en cromatogramas complejos mediante un algoritmo basado en la discriminación de frecuencias

El procesamiento de las señales en cromatogramas de muestras complejas puede constituir un cuello de botella en la obtención de información significativa. Un problema importante que debe abordarse antes de tratar las señales es la sustracción de la línea base, que puede ser notablemente irregular, e idealmente debería realizarse sin supervisión. Una herramienta interesante, desarrollada recientemente para la sustracción de la línea base, es el algoritmo BEADS, que realiza una descomposición completa de los cromatogramas mediante el uso de filtros de frecuencia altamente eficientes, que separan las señales puras de los compuestos (descritas como señales dispersas), de la línea base (con una frecuencia baja) y el ruido (contribuciones de elevada frecuencia).

Sin embargo, el algoritmo inicialmente propuesto requiere, para procesar correctamente las señales, una selección cuidadosa de los parámetros de trabajo, especialmente la frecuencia de corte que es el parámetro más crítico. Dicha selección debe realizarse mediante prueba y error, dando lugar a un proceso demasiado lento e inestable. Por otro lado, la aplicación del BEADS original a cromatogramas que contienen picos de magnitud extremadamente distinta origina deformaciones en la línea base, que aparecen como pequeñas ondulaciones debajo de los picos principales, asociadas a las grandes diferencias de escala entre los componentes mayoritarios y las trazas. Además, la presencia de señales negativas en los cromatogramas afecta gravemente a la sustracción de la línea base.

Para mejorar el rendimiento y fiabilidad del algoritmo BEADS, la Memoria de Tesis Doctoral incluye la propuesta de modificaciones de diverso tipo, a lo que se ha denominado BEADS asistido, ya que la selección de los parámetros

de trabajo óptimos se simplifica en base al empleo de gráficos de autocorrelación auxiliares. Una característica importante del algoritmo BEADS modificado es la transformación logarítmica de las señales originales, que elimina las irregularidades observadas en la línea base debajo de los picos de mayor tamaño. La transformación logarítmica reduce el peso de estos picos, lo que conduce a líneas base suavizadas. Al realizar la transformación logarítmica de la señal, se obtuvieron gráficas escalonadas para cada parámetro de trabajo, cuyo valor óptimo se ubicó cerca del punto de inflexión.

El BEADS asistido puede adaptarse fácilmente a cualquier tipo de muestra, proporcionando una sustracción de la línea base satisfactoria para todas las muestras analizadas, independientemente de su complejidad. El algoritmo propuesto reduce la subjetividad en la selección de los parámetros de trabajo y proporciona resultados siempre fiables. La selección de la frecuencia de corte óptima, que constituye el límite entre la línea base y el resto de contribuciones (señales dispersas y ruido), es menos crítica en comparación con el algoritmo original. Los efectos de las señales negativas esporádicas, tras la substracción de la línea base, se corrigieron mediante la implementación de un proceso iterativo.

Cabe señalar que BEADS realiza un ajuste global de la línea base. Ello implica la pérdida de detalle en regiones particulares del cromatograma, respecto al ajuste local de la línea base (que sólo considera el entorno de un pico). Sin embargo, la magnitud de los errores obtenidos con el BEADS asistido fue muy aceptable. Debe, por último destacarse, que la aplicación de BEADS asistido no se limita únicamente a las señales cromatográficas.

2.2. Desarrollo de un criterio de resolución para caracterizar cromatogramas complejos cuando no se dispone de estándares

El objetivo de las estrategias de optimización interpretativa es la búsqueda de condiciones experimentales que originen la mejor resolución, en base a la predicción de tiempos de retención y perfiles de los picos de los analitos de interés, con los que se construyen cromatogramas simulados. La mayoría de los criterios de resolución utilizados para medir la calidad de una separación requieren de estándares de los analitos, para ajustar los modelos con los que se realizan las predicciones. Sin embargo, para algunas muestras, no hay estándares disponibles. Por lo tanto, se pensó en desarrollar una función de resolución global, válida para todas las situaciones (con o sin estándares).

La función propuesta se basa en la medida de la prominencia de pico, que es la fracción de área que excede la línea que une los valles que delimitan cada pico. El criterio de prominencia de pico se validó mediante la comparación de los resultados con los obtenidos con el criterio de pureza de pico, que mide el área de pico libre de solapamiento y proporciona estimaciones fiables de la resolución cromatográfica. El criterio de pureza de pico requiere un conocimiento exhaustivo de las señales individuales de cada analito, en cada condición del diseño experimental, a lo que sólo se puede acceder mediante simulación basada en la información proporcionada por estándares. Por el contrario, la prominencia de pico se puede medir directamente a partir de las señales en un cromatograma real, sin ningún conocimiento previo de los compuestos que contiene la muestra.

Para comparar los criterios de prominencia y pureza de pico, se obtuvieron los cromatogramas para un conjunto de aminoácidos derivatizados con o-ftalaldehído y N-acetilcisteína, en condiciones isocráticas y de gradiente. Con los datos obtenidos de los estándares para 10 condiciones de elución isocráticas,

se construyeron modelos de retención y semianchura de pico. Con estos modelos, se predijo la separación en alrededor de 1100 gradientes lineales y multi-lineales. Los aminoácidos derivatizados sólo se pudieron resolver a tiempos de análisis elevados, incluso utilizando gradientes multi-isocráticos y multi-lineales. Cuando se intentó reducir el tiempo de análisis, se produjo un solapamiento significativo para varios compuestos. Este comportamiento dio lugar a casos de estudio de interés para la evaluación de las funciones de resolución.

El estudio comparativo se llevó a cabo con la ayuda de gráficos de optimalidad de Pareto. Los gráficos se trazaron para ambos criterios de prominencia y pureza de pico, considerando las dos medidas de calidad opuestas a mejorar: la resolución cromatográfica y el tiempo de análisis. Se obtuvieron gráficos para varias situaciones simuladas: señales de diferente magnitud, inclusión de ruido instrumental, líneas base reales y presencia de compuestos desconocidos.

Se estudiaron tres funciones como candidatas para medir la prominencia de pico global (Ecuaciones (9.3) a (9.5)), que se compararon con la pureza de pico global expresada como la suma de los valores individuales (Ecuación (9.6)). La suma de las resoluciones individuales normalizadas (Ecuación (9.4)) resultó la mejor, ya que la proyección de los gradientes óptimos para el frente de Pareto, para esa función, coincidieron con los obtenidos para el frente de Pareto obtenido para la suma de las purezas de pico.

La mejor función de prominencia global se aplicó con éxito a la evaluación de la resolución de huellas dactilares cromatográficas de extractos de hierbas medicinales, que contenían un gran número de componentes cuya identidad se desconocía. El criterio de resolución propuesto posee la ventaja de poderse evaluar directamente a partir de los cromatogramas experimentales, sin requerir

etapas de modelización, predicción y simulación, utilizando información obtenida de estándares, como es el caso de la pureza pico.

2.3. Clasificación de extractos de hojas y pulpa de olivo mediante cromatografía líquida bidimensional

Los extractos de hojas y pulpa de olivo son mezclas complejas de cientos de compuestos diferentes. Entre ellos, los polifenoles han atraído mucha atención debido a sus efectos beneficios para la salud. Los análisis de polifenoles se suelen realizar mediante cromatografía líquida mono-dimensional. Sin embargo, la complejidad de las muestras hace que la resolución completa no sea posible. Por ello, se investigó la posibilidad de utilizar cromatografía líquida bidimensional en el modo LC×LC (*comprehensive two-dimensional liquid chromatography*), para realizar los análisis. Este modo cromatográfico combina dos columnas con diferentes mecanismos de separación para obtener una máxima resolución en el análisis de muestras complejas, dando lugar a cromatogramas en dos dimensiones.

Se evaluó la capacidad separadora de varias columnas (con diferente fase estacionaria, longitud, diámetro interno, así como distintos tamaños de poro y partícula), a fin de obtener el número máximo de picos visibles (i.e., capacidad de pico), en el análisis de huellas dactilares polifenólicas, haciendo uso de distintas condiciones de elución. A lo largo del estudio, se consideraron tres fases estacionarias en la primera dimensión (C18 convencional y C18 con grupos fenilo o pentafluorofenilo), y cinco en la segunda dimensión (C18, amido, ciano, fenilo y pentafluorofenilo). La separación en la primera dimensión se realizó con gradientes de metanol-agua, mientras que en la segunda dimensión, se hizo uso de gradientes de acetonitrilo-agua.

La optimización de la mejor combinación de columnas se inició utilizando columnas convencionales C18 y ciano en la primera y segunda dimensión, respectivamente, y un gradiente convencional, lo que dio lugar a un número demasiado bajo de picos visibles (29 para los extractos de hojas de olivo). El cambio de la columna ciano por una columna de pentafluorofenilo convencional, junto con la reducción del tiempo de modulación (tiempo de recogida del efluente de la primera dimensión antes de ser inyectado en la segunda dimensión) incrementaron el número de picos a 73. Finalmente, utilizando columnas submicro (C18 de 1.8 µm de diámetro interno en la primera dimensión, y pentafluorofenilo de 2.6 µm en la segunda), cambiando el orden de las columnas (pentafluorofenilo en la primera dimensión), y aplicando en la segunda dimensión un gradiente que desplazaba gradualmente los extremos del gradiente a valores más altos, se lograron huellas dactilares más informativas con 112 y 109 picos visibles para los extractos de hojas de olivo y pulpa, respectivamente.

El método LC×LC optimizado se aplicó con éxito a la confirmación de la presencia 26 picos comunes en los extractos de hojas de olivo y 29 en los de pulpa. Para estos compuestos, se seleccionaron los volúmenes relativos de los picos (menos sensibles al proceso de extracción que los volúmenes absolutos), con el fin de desarrollar un modelo de análisis discriminante lineal (LDA, *linear discriminant analysis*), capaz de distinguir la procedencia de los extractos. Se trazaron gráficos tridimensionales con las puntuaciones obtenidas a partir de la información proporcionada por los cromatogramas LC×LC de los extractos de hoja y pulpa de olivo, de acuerdo a las tres primeras funciones discriminantes. Los gráficos mostraron que todas las muestras pertenecientes a una clase determinada aparecían en grupos compactos. Los modelos LDA resultantes permitieron la correcta clasificación de siete cultivos de distinto

origen genético, para las hojas y pulpa de olivo de varias regiones españolas, obteniéndose una excelente separación entre categorías, con un alto nivel de confianza. Esto demuestra que los perfiles polifenólicos son característicos de cada cultivo.

INDEX

OBJECTIVES AND
DEVELOPMENT OF THE RESEARCH

High performance liquid chromatography (HPLC) is currently the most widely used analytical separation technique. Unfortunately, the efficiency is usually smaller than that achieved in gas chromatography, capillary electrophoresis and other electromigration techniques. This constrains the analysis of complex samples. Hence, the high effort dedicated to increase the efficiency and selectivity in HPLC, with significant improvements since the beginning of the technique. Despite the progress in the last decades with the development of increasingly more sophisticated instrumentation (including the introduction of ultra-high pressure pumps), and the advances in column technology (with the synthesis of new supports and stationary phases), there are still challenges to solve.

The PhD. work collected in this Project gathers some proposals of two types to improve the HPLC performance. The objectives are next briefly summarised:

Objective 1. *Increasing the modelling capability in liquid chromatography*

Interpretive optimisation strategies are based on the accurate description of the chromatographic behaviour (retention and peak profiles), using mathematical models. The predictive capability of the fitted models will depend on the quality of the information provided by the experimental design. Isocratic experiments are maximally informative, but they suffer from the important drawback of needing a long time for data acquisition, and are unpractical for highly hydrophobic solutes. For this reason, the use of gradient experimental designs are recommended. In this PhD. work, several studies are presented which have deepened in the analysis of experimental designs to enhance the prediction capability, as well as the development of new retention models to optimise the resolution in the presence of additives.

The following four objectives were addressed:

Objective 1.1: Investigate the application of interpretive optimisation strategies in nano-liquid chromatography using polymeric monolithic columns, and evaluate the chromatographic behaviour with different stationary phases prepared in the laboratory.

Objective 1.2: Analyse the predictive capability of diverse experimental designs and retention models, based on the quality information provided by isocratic, gradient and mixed experiments. The use of isocratic runs including transient increments of organic solvent (pulses) is proposed for modelling purposes.

Objective 1.3: Design a new methodology for the estimation of peak capacity, based on peak simulation, valid for a variety of situations including extra-column contribution effects, asymmetrical peaks and the use of multi-linear gradients.

Objective 1.4: Apply interpretive optimisation strategies in micellar liquid chromatography. A particular interest is on the development of new retention models that describe the chromatographic behaviour in extended organic solvent domains, and the optimisation of the elution conditions for the analysis of physiological fluids under isocratic and gradient elution.

Objective 2. *Improving the separation performance for chromatographic fingerprints*

The resolution of complex samples containing unknown compounds of different nature, or without standards, as is the case of chromatographic fingerprints, is still a challenge. Usually, samples with similar fingerprints also have similar properties. Therefore, this type of chromatogram has a potential interest to determine the identity, authenticity and consistency between batches of natural products of diverse nature. Strategies that allow obtaining fingerprints

as rich in information as possible are still neeeded. For this purpose, tools to process properly complex chromatograms and improve the separation performance should be developed.

The following three objectives were addressed:

Objective 2.1: Progress in data pre-processing of chromatographic fingerprints of natural products. We were particularly interested in the baseline subtraction of highly complex samples with little supervision.

Objective 2.2: Develop a chromatographic objective function valid to measure the separation quality when there are no standards available.

Objective 2.3: Improve the analysis of natural products, obtaining informative chromatographic fingerprints by comprehensive two-dimensional liquid chromatography.

The contents of the PhD. Project are divided in two parts:

Part 1. *Increasing the modelling capability in liquid chromatography*

Part 2. *Improving the separation performance for chromatographic fingerprints*

The work has implied a large experimental effort, designed to explore and extract information on the chromatographic behaviour of compounds of different nature. A great diversity of experimental conditions using aqueous-organic mobile phases with acetonitrile and methanol, and pure and hybrid micellar mobile phases, have been used along the studies. The work has implied an extensive data treatment, mainly related with the construction of models to predict the chromatographic retention, peak shape and resolution of mixtures of compounds, with optimisation purposes.

The large effort in several literature surveys on the different topics investigated in this work should be also highlighted. This has implied the search,

reading and organisation of a large amount of valuable information that is properly reflected in each part of the PhD. Project.

Supervisors and research laboratories

The research work leading to the PhD. degree in Chemistry was started in October 2016, once the Master degree on "Experimental Techniques in Chemistry", offered by the Departments of Analytical Chemistry and Inorganic Chemistry at the University of Valencia, was finished. The experimental work in this Project was developed in the Department of Analytical Chemistry at the University of Valencia, under the supervision of María Celia García Álvarez-Coque and José Ramón Torres Lapasió. Acknowledge should be also given to the valuable collaboration of José Manuel Herrero Martínez and Ernesto Francisco Simó Alfonso, in some fundamental and applied studies.

The PhD. period included two research stays of three-months abroad:

September to December 2018: under the supervision of Paola Dugo in the Department of CHIBIOFARAM at the University of Messina (Italy), working in the field of analytical method development. The aim of the research was to develop a new HPLC methodology with photodiode array and mass spectrometry detectors for the identification and quantification of polyphenols in mustard (Brassica Juncea) cultivars.

September to December 2020: under the supervision of Davy Guillarme in the Department of Analytical Science at the University of Geneva (Switzerland), working in the field of the analysis of protein biopharmaceuticals. The aim of the research was to develop an innovative strategy for the characterisation of monoclonal antibodies using ultra-short columns in hydrophilic liquid chromatography (HILIC) and ion-exchange liquid chromatography (IEX).

Publications

The publications included in this PhD. Project are the following (the journal impact factor, IF, and ranking in the category of Analytical Chemistry are given):

1. José Antonio Navarro Huerta, José Ramón Torres Lapasió, Sergio López Ureña, María Celia García Álvarez-Coque
   Assisted baseline subtraction in complex chromatograms using the BEADS algorithm.
   Journal of Chromatography A 1507 (2017) 1–10 (Chapter 8).
   IF (2017): 3.716 (Analytical Chemistry: 13/80)

2. María Celia García Álvarez-Coque, José Ramón Torres Lapasió, José Antonio Navarro Huerta
   Secondary equilibria in reversed-phase liquid chromatography.
   In Liquid Chromatography: Fundamentals and Instrumentation (edited by S. Fanali, P.R. Haddad, C.F. Poole, M. Riekkola), Elsevier, Amsterdam, Netherlands, 2nd ed., Vol. 1, 2017, pp. 125–146 (written by invitation) (Chapter 6).

3. Guillermo Ramis Ramos, María Celia García Álvarez-Coque, José Antonio Navarro Huerta,
   Solvent selection in liquid chromatography.
   In Liquid Chromatography: Fundamentals and Instrumentation (edited by S. Fanali, P.R. Haddad, C.F. Poole, M. Riekkola), Elsevier, Amsterdam, Netherlands, 2nd ed., Vol. 1, 2017, pp. 343–373 (written by invitation) (Chapter 1).

4. <u>José Antonio Navarro Huerta</u>, Tamara Álvarez Segura, José Ramón Torres Lapasió, María Celia García Álvarez-Coque

   Study of the performance of a resolution criterion to characterise complex chromatograms with unknowns or without standards.

   Analytical Methods 9 (2017) 4293–4303 (Chapter 9).

   IF (2017): 2.073 (Analytical Chemistry: 42/80)

5. <u>José Antonio Navarro Huerta</u>, José Ramón Torres Lapasió, María Celia García Álvarez-Coque

   Estimation of peak capacity based on peak simulation.

   Journal of Chromatography A 1574 (2018) 101–113 (Chapter 5).

   IF (2018): 3.858 (Analytical Chemistry: 15/84)

6. <u>José Antonio Navarro Huerta</u>, Adrián Gisbert Alonso, José Ramón Torres Lapasió, María Celia García Álvarez-Coque

   Benefits of solvent concentration pulses in retention time modelling of liquid chromatography.

   Journal of Chromatography A 1597 (2019) 76–88 (Chapter 3).

   IF (2019): 4.049 (Analytical Chemistry: 14/86)

7. <u>José Antonio Navarro Huerta</u>, Enrique Javier Carrasco Correa, José Ramón Torres Lapasió, José Manuel Herrero Martínez, María Celia García Álvarez-Coque

   Modelling retention and peak shape of small polar solutes analysed by nano-HPLC using methacrylate-based monolithic columns.

   Analytica Chimica Acta 1086 (2019) 142–155 (Chapter 2).

   IF (2019): 5.977 (Analytical Chemistry: 10/86)

8. <u>José Antonio Navarro Huerta</u>, Ángel Gamaliel Vargas García, José Ramón Torres Lapasió, María Celia García Álvarez-Coque

   Interpretive search of optimal isocratic and gradient separations in micellar liquid chromatography in extended organic solvent domains.

   Journal of Chromatography A 1616 (2020) 460784 (Chapter 7).

   IF (2019): 4.049 (Analytical Chemistry: 14/86)

9. María Vergara Barberán, <u>José Antonio Navarro Huerta</u>, José Ramón Torres Lapasió, Ernesto Francisco Simó Alfonso, María Celia García Álvarez-Coque

   Classification of olive leaves and pulp extracts by comprehensive two-dimensional liquid chromatography of polyphenolic fingerprints.

   Food Chemistry 320 (2020) 126630 (Chapter 10).

   IF (2019): 6.306 (Applied Chemistry: 5/71)

10. <u>José Antonio Navarro Huerta</u>, Adrián Gisbert Alonso, José Ramón Torres Lapasió, María Celia García Álvarez-Coque

    Testing experimental designs in liquid chromatography (I): Development and validation of a method for the comprehensive inspection of experimental designs.

    Journal of Chromatography A 1624 (2020) 461180 (Chapter 4).

    IF (2019): 4.049 (Analytical Chemistry: 14/86)

The work not included in this PhD. Project is:

1. Adrián Gisbert Alonso, <u>José Antonio Navarro Huerta</u>, José Ramón Torres Lapasió, María Celia García Álvarez-Coque

   Global retention models and their application to the prediction of chromatographic fingerprints.

   Journal of Chromatography A 1637 (2021) 461845.

   IF (2019): 4.049 (Analytical Chemistry: 14/86)

2. Adrián Gisbert Alonso, <u>José Antonio Navarro Huerta</u>, José Ramón Torres Lapasió, María Celia García Álvarez-Coque

   Testing experimental designs in liquid chromatography (II): Influence of the design geometry on the predictive performance of retention models.

   Journal of Chromatography A (in preparation, 2021).

   IF (2019): 4.049 (Analytical Chemistry: 14/86)

3. <u>José Antonio Navarro Huerta</u>, Amarande Murisier, Jennifer M. Nguyen, Matthew A. Lauber, Davy Guillarme, Szabolcs Fekete

   Ultra-short ion-exchange columns for the charge variant analysis of therapeutic proteins.

   Journal of Chromatography A (in preparation, 2021).

   IF (2019): 4.049 (Analytical Chemistry: 14/86)

Congress communications

The developed research has been also presented in 16 scientific conferences in the period June 2016 to November 2020 (10 communications in international conferences and 6 communications in national conferences, 23 were posters and 6 oral communications). The PhD. candidate has attended to 9 of these conferences: Workshop/November 2016 (Valencia), HPLC'2017 (Prague), Euroanalysis' 2017 (Stockholm), SEQA'2017 (Valencia), ISCC'2018 (Riva del Garda), ISSS'2018 (Jasná), Workshop/November 2018 (Valencia), Workshop/June 2019 (Valencia), and HPLC'2019 (Milan):

XVI Chemometrics in Analytical Chemistry (CAC'2016)
Barcelona (Spain), June 2016 (international)

1. María Celia García Álvarez-Coque, José Ramón Torres Lapasió, José Antonio Navarro Huerta, Tamara Álvarez Segura
   Measurement of resolution for highly complex muti-analyte samples
   (Poster P-002)

2. José Ramón Torres Lapasió, José Antonio Navarro Huerta, María Celia García Álvarez-Coque, Tamara Álvarez Segura
   Maximisation of the information in chromatographic fingerprints
   (Poster P-003)

3. José Ramón Torres Lapasió, José Antonio Navarro Huerta, Sergio López Ureña, María Celia García Álvarez-Coque
   Automatic baseline subtraction, peak detection and quantification of complex chromatograms
   (Poster P-014)

31st International Symposium on Chromatography (ISC'2016)
Cork (Ireland), August-September 2016 (international)

4. María Celia García Álvarez-Coque, José Ramón Torres Lapasió, Tamara Álvarez Segura, <u>José Antonio Navarro Huerta</u>
    Measurement of resolution and optimisation of chromatographic conditions in the absence of standards
    (Poster P01-001-017)

5. José Ramón Torres Lapasió, Juan José Baeza Baeza, <u>José Antonio Navarro Huerta</u>, Manuel David Peris Díaz, María Celia García Álvarez-Coque
    Modified Gaussian functions to model the peak shape and asymmetry of chromatographic peaks
    (Poster P02-002-058)

First Workshop on Separation Strategies in Chromatography
Burjassot, Valencia (Spain), November 2016 (national)

6. <u>José Antonio Navarro Huerta</u>, Tamara Álvarez Segura, José Ramón Torres Lapasió, María Celia García Álvarez-Coque
    A resolution criterion to characterise complex chromatograms with unknown compounds or without standards available: A validation study
    (Oral communication)

7. Sergio López Ureña, <u>José Antonio Navarro Huerta</u>, José Ramón Torres Lapasió, María Celia García Álvarez-Coque
    Assisted baseline subtraction in complex chromatograms using the BEADS algorithm
    (Oral communication)

45th International Symposium on High Performance Liquid Phase Separations and Related Techniques (HPLC'2017)
Prague (Czech Republic), June 2017 (international**)**

8. José Antonio Navarro Huerta, Sergio López Ureña, José Ramón Torres Lapasió, María Celia García Álvarez-Coque

    Assisted baseline subtraction in complex chromatograms using the BEADS algorithm

    (Poster FUN13-P06-Mo)

9. María Celia García Álvarez-Coque, José Antonio Navarro Huerta, Tamara Álvarez Segura, José Ramón Torres Lapasió

    Measurement of resolution in complex chromatograms aimed to optimise the separation

    (Oral communication FUN13-O2-Th)

XIX Europe's Analytical Chemistry Meeting (Euroanalysis'2017)
Stockholm (Sweden), August-September 2017 (international**)**

10. José Antonio Navarro Huerta, Tamara Álvarez Segura, José Ramón Torres Lapasió, María Celia García Álvarez-Coque

    Benefits of serially-coupled chromatographic columns

    (Poster 017 Tue-SEP)

11. José Antonio Navarro Huerta, José Ramón Torres Lapasió, María Celia García Álvarez-Coque

    Peak capacity for multi-linear gradients based on peak simulation

    (Poster 018 Tue-SEP)

XXI Meeting of the Spanish Society of Analytical Chemistry (SEQA'2017)
Valencia (Spain), September 2017 (national)

12. <u>José Antonio Navarro Huerta</u>, José Ramón Torres Lapasió, María José Ruiz Ángel, María Celia García Álvarez-Coque

    Interpretive optimization of gradients of organic solvent in micellar liquid chromatography

    (Poster PO-TS-09)

42nd International Symposium on Capillary Chromatography (ISCC'2018)
Riva del Garda (Italy), May 2018 (international)

13. <u>José Antonio Navarro Huerta</u>, Enrique Javier Carrasco Correa, José Ramón Torres Lapasió, José Manuel Herrero Martínez, María Celia García Álvarez-Coque

    Modelling retention of small polar solutes in methacrylate monolithic columns by nano-HPLC

    (Poster B.02)

24th International Symposium on Separation Sciences (ISSS'2018)
Jasná (Slovakia), June 2018 (international)

14. <u>José Antonio Navarro Huerta</u>, Adrián Gisbert Alonso, José Ramón Torres Lapasió, María Celia García Álvarez-Coque

    Prediction of peak width and asymmetry in linear and multi-linear gradient elution: Jandera's approximation versus numerical integration

    (Poster P058)

15. José Antonio Navarro Huerta, Tamara Álvarez Segura, Sergio López Ureña, José Ramón Torres Lapasió, María Celia García Álvarez-Coque

Comparison of search strategies for finding optimal multi-linear gradient

(Poster P109)

32nd International Symposium on Chromatography (ISC'2018) Cannes-Mandelieu (France), September 2018 (international)

16. José Ramón Torres Lapasió, Sergio López Ureña, José Antonio Navarro Huerta, María Celia García Álvarez-Coque

Design of experiments in gradient chromatography based on the minimization of prediction errors

(Poster PS-05-06)

17. José Ramón Torres Lapasió, José Antonio Navarro Huerta, María José Ruiz Ángel, María Celia García Álvarez-Coque

Analysis of β-adrenoceptor antagonists in urine samples using multi-linear gradient elution in micellar liquid chromatography

(Poster PS-07-07)

18. María Celia García Álvarez-Coque, José Antonio Navarro Huerta, José Ramón Torres Lapasió,

Prediction of peak capacity for isocratic and complex gradients based on peak simulation

(Oral communication S21-03)

XVIII Scientific Meeting of the Spanish Society of Chromatography and Related Techniques (SECyTA'2018)
Granada (Spain), October 2018 (national)

19. José Ramón Torres Lapasió, Adrián Gisbert Alonso, <u>José Antonio Navarro Huerta</u>, María Celia García Álvarez-Coque
    Pulses of organic solvent and their consequences to enhance isocratic and gradient predictions
    (Poster P-FCH-03)

3rd Workshop on Advances in Separation Techniques
Burjassot, Valencia (Spain), November 2018 (national)

20. <u>José Antonio Navarro Huerta</u>, José Ramón Torres Lapasió, María Celia García Álvarez-Coque
    Peak capacity in isocratic and gradient elution
    (Oral communication)

1st Workshop "Young Researchers in Chemistry"
Burjassot, June 2019 (national)

21. <u>José Antonio Navarro Huerta</u>, José Ramón Torres Lapasió, María Celia García Álvarez-Coque
    Estimation of peak capacity for isocratic and gradient elution based on peak simulation
    (Poster P3)

48th International Symposium on High Performance Liquid Phase Separations and Related Techniques (HPLC'2019)
Milan (Italy), June 2019 (international)

22. José Ramón Torres Lapasió, <u>José Antonio Navarro Huerta</u>, Adrián Gisbert Alonso, María Celia García Álvarez-Coque
    Solvent concentration pulses to improve retention modelling in liquid chromatography
    (Poster P475)

23. <u>José Antonio Navarro Huerta</u>, José Ramón Torres Lapasió, Sergio López Ureña, María Celia García Álvarez-Coque
    Evaluation of the information quality in gradient experimental designs
    (Poster P477)

26th International Symposium on Electroseparation and Liquid Phase-Separation Techniques (ITP'2019)
Toulouse (France), September 2019 (international)

24. José Ramón Torres Lapasió, <u>José Antonio Navarro Huerta</u>, María Celia García Álvarez-Coque
    Approach to predict peak capacity based on peak simulation: Extension of the concept to gradient elution
    (Poster P 50)

25. María Celia García Álvarez-Coque, José Antonio Navarro Huerta, José Ramón Torres Lapasió, Enrique Javier Carrasco Correa, José Manuel Herrero Martínez

    Performance of a nano-HPLC methacrylate monolithic column vs. a conventional microparticulate C18 column for the analysis of sulphonamides

    (Poster P 52)

4th Workshop on Advances in Separation Techniques, and 4th Workshop on Separation Strategies in Chromatography
Burjassot, Valencia (Spain), October 2019 (national)

26. María Vergara Barberán, José Antonio Navarro Huerta, José Ramón Torres Lapasió, Ernesto Francisco Simó Alfonso, María Celia García Álvarez-Coque

    Classification of olive leaf and pulp extracts by their genetic variety using polyphenolic fingerprints established by comprehensive two-dimensional liquid chromatography

    (Oral communication)

27th International Symposium on Electroseparation and Liquid Phase-Separation Techniques (ITP'2020), Virtual Conference
Nanjing (China), November 2020 (international)

27. José Ramón Torres Lapasió, Adrián Gisbert Alonso, José Antonio Navarro Huerta, María Celia García Álvarez-Coque

    Global retention models in HPLC for prediction of fingerprints under gradient elution

    (Poster)

28. José Ramón Torres Lapasió, Adrián Gisbert Alonso, <u>José Antonio Navarro Huerta</u>, María Celia García Álvarez-Coque

   Influence of the design geometry on the predictive capability of training gradient experiments

   (Poster)

29. María Vergara Barberán, <u>José Antonio Navarro Huerta</u>, José Ramón Torres Lapasió, Ernesto Francisco Simó Alfonso, María Celia García Álvarez-Coque

   Classification of olive leaf and pulp extracts of polyphenolic fingerprints by comprehensive two-dimensional liquid chromatography

   (Poster) (Best Poster Award, Peer Review, Second runner-up)

The research along the PhD. Project was funded by one autonomic and three national Research Projects:

1. Project CTQ2013-42558-P: *"Modulation of selectivity and efficiency in HPLC using multi-column strategies to enhance the resolution of complex samples"*, funded by Ministry of Economy and Competitiveness, January 2014–December 2016. Main researchers: María Celia García Álvarez-Coque and José Ramón Torres Lapasió.

2. Project PROMETEO/2016/128: *"Multi-column strategies to enhance the performance in the separation of complex samples by liquid chromatography"*, funded by Generalitat Valenciana (Direcció General d'Universitat, Investigació i Ciència), January 2016–December 2019. Main researcher: María Celia García Álvarez-Coque.

3. Project CTQ2016-75644-P: *"Design of methodologies to optimize the separation quality in liquid chromatography"*, funded by Ministry of Economy, Industry and Competitiveness, January 2017–December 2019. Main researchers: María Celia García Álvarez-Coque and José Ramón Torres Lapasió.

4. Project PID2019-106708GB-I00: "*Fundamental studies and design of strategies in one- and two-dimensional chromatography to enhance the separation performance for complex samples*", funded by Ministry of Science, Innovation and Universities, January 2020–December 2022. Main researchers: María Celia García Álvarez-Coque and José Ramón Torres Lapasió.

Along the PhD period, three grants were obtained:

1. Pre-doctoral grant, associated to Project PROMETEO/2016/128 (November 2016–December 2018).

2. UV-INV-PREDOC18F1-742530, funded by University of Valencia (January 2019–April 2021).

3. FPU17/02979, funded by Ministry of Science, Innovation and Universities (this grant was obtained at the same time as UV-INV-PREDOC18F1-742530 grant, but was dropped to enjoy the grant offered by the University of Valencia).

CHAPTER 1

INTRODUCTION

1.1. Elution strength

In liquid chromatography (LC), the elution strength is the ability of the mobile phase to sweep away the solutes retained on the stationary phase. It depends on the nature of the stationary phase and solutes, as well as on the mobile phase composition (i.e., nature and concentration of the solvents and additives), pH, and column temperature. Therefore, for a given stationary phase, the elution strength is not a property exclusively related to the solvent, since solutes undergo different elution strengths depending on their particular molecular structures. The elution strength of the mobile phase is a very practical concept in LC, commonly used to adjust the overall retention for a group of solutes inside the target retention region, optimally within the $1 < k < 5$ range, or at least $0.2 < k < 20$, k being the retention factor or relative retention:

$$k = \frac{t_R - t_0}{t_0} \tag{1.1}$$

where t_R is the retention time and t_0 the dead time (i.e., retention time of an unretained solute). For a given stationary phase and set of solutes, if the elution strength is too high, retention times will be too short, and consequently, the resolution will be poor. Conversely, if the elution strength is too low, retention times will be excessive, and consequently, the analysis time will be too long and, due to excessive dilution, the signal-to-noise ratio at the peak maxima of the most retained analytes will decrease significantly. Once the elution strength has been adjusted, the selectivity (i.e., elution order and peak distribution) can be optimised without modifying significantly the overall retention [1]. The optimisation criterion for selectivity is to resolve all the peak pairs of the target samples within a total analysis time as short as possible.

Table 1.1. Solvent properties.

Solvent	Normal boiling point (°C)[a]	Cutoff wavelength (nm)[a]	Viscosity at 20 °C (mPa·sec)[a]	Solubility parameter, (δ)[b]	Snyder global polarity, (P')[c]
Isooctane	99.2	200–210	0.50	7.0	−0.4
Diisopropyl ether	68.0	380	0.33	7.1	1.8
n-Heptane	98.4	200	0.42	~7.5	0.0
n-Hexane	68.7	200	0.31	~7.5	0.0
Triethylamine	89.5	235	0.38	7.5	1.8
Cyclohexane	80.7	200	0.98	8.2	0.0
Carbon tetrachloride	76.8	263	0.97	8.6	1.7
Ethyl acetate	77.1	256	0.46	8.9	4.3
Toluene	110.6	284	0.59	8.9	2.3
Tetrahydrofuran	66.0	212	0.55	9.1	4.2
Chloroform	61.2	245	0.58	9.2	4.4
Dichloromethane	40.0	232	0.44	9.6	4.3
Methyl ethyl ketone	79.6	329	0.42 (15 °C)	9.5	4.5
Acetone	56.3	330	0.30 (25 °C)	9.6	5.4
Carbon disulfide	46.0	220	0.36	10.0	1.1
1,4-Dioxane	101.3	215	1.44 (15 °C)	10.1	4.8
Pyridine	115.3	330	0.95	10.6	5.3

Table 1.1 (continued).

Solvent	Normal boiling point (°C)[a]	Cutoff wavelength (nm)[a]	Viscosity at 20 °C (mPa·sec)[a]	Solubility parameter, (δ)[b]	Snyder global polarity, (P')[c]
Isopropanol	82.3	205	2.86 (15 °C)	11.4	4.3
1-Butanol	117.7	215	2.95	11.6	3.9
2-Methoxyethanol	124.6	210	1.72	11.7	5.7
Dimethylformamide	153.0	268	0.92	11.8	6.4
Ethanol	78.3	205–210	1.20	12.0	5.2
Dimethylsulfoxide	189.0	286	2.20	12.0	6.5
Acetonitrile	81.6	190	0.34	12.1	6.2
1-Propanol	97.2	210	2.26	12.2	3.9
Acetic acid	117.9	210	1.31 (15 °C)	13.0	6.2
Methanol	64.7	205	0.55	14.5	6.6
Formamide	210.5	210	3.50	19.2	7.3
Water	100.0	<190	1.00	23.5	9.0

[a] Refs. [2,3]. [b] According to Hildebrand, Refs. [2–5]. [c] Refs. [6,7].

In addition to water, many organic solvents can be used to prepare the mobile phase (see some examples in Table 1.1). Also, it is possible to use mixtures of solvents in different ratios to modify the solvent properties (e.g., the elution strength and selectivity). This can make solvent selection for a given purpose a puzzling task, unless suitable guidelines are followed. This chapter summarises the most common strategies used by skilled chromatographers. Although mostly developed and used for reversed-phase liquid chromatography (RPLC) [8], the guidelines should be useful for normal-phase liquid chromatography (NPLC) as well [9], including the aqueous-compatible normal mode known as hydrophilic interaction liquid chromatography (HILIC). The elution strength can be either maintained constant (isocratic elution), or gradually increased (gradient elution). In both approaches, the elution strength can be tuned to get the desired resolution and analysis time.

1.2. Columns and solvents in RPLC, NPLC, and HILIC

In RPLC, the stationary phase is non-polar or weakly polar. The most common choice is octadecyl-silica (C18). The retention of highly hydrophobic solutes is reduced by using octyl- (C8) or butyl-silica (C4), and reversely, to increase the retention of some solutes, highly hydrophobic stationary phases such as triacontyl-silica (C30) are used. Other bonded phases such as pentafluorophenylpropyl-silica or biphenyl-silica offer different selectivity. The mobile phase is prepared with water, to which a miscible organic solvent (the "modifier") is added to reduce the polarity and increase the elution strength. As the mixture progressively resembles the stationary phase, it competes better for desorption of non-polar solutes, which are strongly associated with the stationary phase. In principle, a wide range of water-miscible organic solvents

may be used as modifiers (Table 1.1); however, only three are usual in RPLC: acetonitrile (ACN), methanol (MeOH), and tetrahydrofuran (THF), especially the first. Solute elution occurs according to the decreasing polarities: the most hydrophilic solutes (which prefer the polar mobile phase) elute the first, while the most hydrophobic (which prefer the stationary phase) elute the last.

In NPLC, the stationary phase is polar. In order of increasing polarity, the most common stationary phases are cyanopropyl-silica, hydride silica, underivatised silica, diolpropyl-silica, and aminopropyl-silica. The mobile phase should be non-polar and consists of an alkane mixed with a miscible polar solvent (the "modifier") to increase the elution strength. As the mixture more closely resembles the polar stationary phase, retention is reduced. Hexane is still largely used; however, because of concern about its long-term toxicity, it is being progressively substituted with isoheptane or the slightly more viscous *n*-heptane or cyclohexane. In addition, due to concern about the environmental impact of alkanes, sustainable or "green chemistry" solvents have been proposed as substitutes. These are mostly terpenes of vegetal origin as limonene, *p*-cymene and α-pinene. Among the suitable modifiers (Table 1.1), the most common are chloroform (the worst choice from the viewpoint of green chemistry), ethyl acetate, dichloromethane, and isopropanol. Solutes elute in the order of increasing polarity: the most hydrophobic solutes elute the first, followed by the more polar solutes, which interact stronger with the stationary phase.

A water-rich layer adsorbed onto a polar stationary phase, such as underivatised silica or a silica-bonded polyol, ionic and zwitterionic stationary phase, is used in HILIC. Water-ACN mixtures (water is now the "modifier") are most frequently used as mobile phases. Instead of ACN, other water-

miscible solvents used in HILIC are acetone, isopropanol, ethanol, 1,4-dioxane, dimethylformamide and MeOH.

1.3. Assessment of the elution strength

Two types of scales have been essentially used to estimate the capability of solvents to interact with their own and with other molecules: the solvatochromic scales, based solely on the solvent properties, and the eluotropic scales, which measure solvent properties in the presence of a reference stationary phase. For the first type, polarity scales based on spectroscopic measurements (spectral shifts in the absorption bands of some reference solutes), energy measurements, or theoretical descriptors have been proposed [5,10]. All these polarity scales can be used to estimate the elution strength of a solvent or a solvent mixture, and thus predict the retention for a given analyte.

Retention results from the many different intermolecular interaction mechanisms established between analytes and both the stationary and mobile phases. However, an extremely rough but rather useful simplification in LC is to refer to the elution strength of the mobile phase, independently from the nature of the solutes. This provides an idea about the global capability of the solvent mixture to push any heterogeneous group of analytes down the system. Fortunately, the elution strength is differently experienced by different analytes, which makes separation and selectivity tuning possible. Other decisions that should be taken in modelling or predicting retention are how many solvent interactions will be handled and how they will be measured. This is equivalent to selecting a polarity scale or a set of polarity descriptors, which estimate the interactions between the solvent molecules, while assuming that the strength of solute-solvent interactions for solutes of any kind is reasonably represented by the internal forces among the solvent molecules. As far as this assumption is

true, any attempt of modelling and predicting retention on the sole basis of the descriptors of solvent properties will be successful. For instance, a highly associated solvent as water is assumed to strongly interact with polar solutes, whereas poorly associated solvents such as alkanes are assumed to weakly interact with all types of solutes.

1.3.1. The Hildebrand solubility parameter and other global polarity estimators

The simplest choice of using a single descriptor of polarity (i.e., a global polarity of a solvent or solvent mixture), will be discussed first. The Hildebrand solubility parameter is a global measurement of the interactions that hold the solvent molecules together and, thus, provides a quantitative polarity scale for solvents by handling a single parameter [10,11]:

$$\delta = \left(-\frac{E}{v} \right)^{1/2} \tag{1.2}$$

where E is the cohesive energy of a mole of solvent, and v the molar volume. The minus sign responds to the fact that the cohesion process is exothermic. As observed in Table 1.1, water is at the bottom of the scale, and its large δ-value is typical of a highly associated solvent. Other polar solvents occupy intermediate positions, and alkanes appear at the top of the scale, with a δ-value typical of solvents with weak internal interactions. From the data, it follows that, for mixtures containing the same amount of modifier, the elution strength increases in the following order: MeOH < ACN < isopropanol << THF in RPLC, and ethyl acetate < chloroform < dichloromethane < isopropanol in NPLC. In RPLC, this order roughly coincides with the elution strength found for mixtures of water with a given amount of modifier. Similarly for NPLC, the

order coincides with that observed using mixtures of an alkane with a given amount of a miscible modifier.

Obviously, not all solvent mixtures are possible. In RPLC and NPLC, only solvents that are miscible with water or heptane, respectively, are used. As a rule, solvents are completely miscible if they are in the same third of the Hildebrand polarity scale (Table 1.1). Therefore, all solvents in the upper-third, bottom-third, or centre-third are completely miscible with each other. A particular case is the dichloromethane/1,4-dioxane pair. These solvents have the same global polarity parameter but are not miscible; dichloromethane is totally miscible with alkanes, whereas 1,4-dioxane mixes with water in all proportions. This reveals the limitations of global polarity parameters, where the contributions of the molecular interactions of different types are not individually considered. Thus, water is incapable of accepting protons from dichloromethane, but 1,4-dioxane readily accepts protons from water. Also, a few "universal" solvents, such as ACN, THF, and isopropanol, are miscible with almost all solvents including heptane and water.

The addition of surfactants at sufficiently high concentration increases the miscibility of certain solvents. This has been useful for the development of micellar LC, where some organic solvents, such as butanol and pentanol, are used at concentrations higher than those miscible in aqueous solution, expanding the range of possible mixtures in RPLC [12]. However, if surfactants are present in the mobile phase, and depending on the nature and proportion of the mixture components, either true solutions, thermodynamically stable and transparent microemulsions, or unstable translucent emulsions may result. In contrast, solvent immiscibility provides the basis for countercurrent chromatography. In this technique, the separation is based on the different

relative solubilities of the solutes in two immiscible solvents, one playing the role of the stationary phase, and the other the role of the mobile phase [13].

Another way of globally measuring intermolecular interactions is the relative retention of solvents by adsorption on silica, ε^o. On this strongly polar solid phase, alcohols show strong interaction ($\varepsilon^o = 0.6$–0.7), whereas alkanes interact weakly ($\varepsilon^o = 0.01$). This polarity descriptor is eluotropic, since it is established using a reference stationary phase. Other global eluotropic polarity scales are obtained by measuring adsorption on other solid surfaces, such as alumina. The discrepancies among the different solvatochromic and eluotropic scales are inevitable, due to the limitations inherent in the use of a single global polarity parameter or uniparametric approach; however, the discrepancies do not disappear by using a multi-parametric approach relaying on a few solvent descriptors, as they also depend on the way they are defined and measured.

1.3.2. Global polarity for solvent mixtures

In RPLC, the polarity of a mixture of solvents is usually estimated as follows:

$$\delta_M = \sum_j \delta_j \, \varphi_j \tag{1.3}$$

where δ_j and φ_j are the Hildebrand solubility parameter and volumetric fraction of solvent j in the mixture, respectively (of course, any other polarity scale, whether solvatochromic or eluotropic based on a single descriptor, can be used for these calculations). For instance, for the MeOH-water mixtures used in RPLC:

$$\delta_M = 14.5 \, \varphi_{MeOH} + 23.5 \, (1 - \varphi_{MeOH}) \tag{1.4}$$

The variation of the global polarity of a mixture (and, consequently, of the elution strength) with mobile phase composition is approximately linear for RPLC using modifier concentrations below 30% (v/v). Non-linear relationships, as those provided later in this chapter, should be expected outside this limit. In NPLC, non-linearity begins at lower modifier contents. Thus, the effect of minute amounts of a polar solvent in an alkane can be much larger than the effect of further adding larger amounts. However, keeping in mind these limitations, Equation (1.3) is useful to estimate the composition of isoeluotropic mixtures in RPLC, as will be next explained.

1.3.3. *Application field of the chromatographic modes as deduced from the Schoenmakers' rule*

Two conditions should be fulfilled to elute solutes within the target retention region:

1. Ideally, the solute polarity (δ_X) should be not far from the mean value of the stationary phase (δ_S) and mobile phase (δ_M) polarities:

$$\delta_X \approx \frac{\delta_S + \delta_M}{2} \qquad (1.5)$$

   Otherwise, solutes will show an excessive preference for one of the phases. With gradient elution, δ_M changes with time. This means that each solute should fulfil Equation (1.5) during its main elution stage, when the analyte is progressing along the column.

2. The polarities of both phases should differ significantly, which is required for a group of solutes of a wide polarity range to fulfil Equation (1.5). If $\delta_M \approx \delta_S$, then δ_X for most solutes would not be in between δ_M and δ_S.

These two conditions are summarised in the rule proposed by Schoenmakers et al. [11], which states that the retention factors are within the optimal target region when:

$$(\delta_M + \delta_S - 2\delta_X)\,(\delta_M - \delta_S) \approx 0 \qquad\qquad (1.6)$$

The second parenthesis should be as large as possible, so that all solutes in a mixture can fulfil Equation (1.5).

Assuming a linear behaviour, the rule can be expressed graphically as shown in Figure 1.1. According to the scheme in Figure 1.1a, solutes with $\delta_X \approx 15.5$ (rather polar) are properly eluted with water ($\delta_M = 23.5$) on a C18 stationary phase ($\delta_S = 7.0$), and a miscible organic solvent should be added to elute less polar solutes. With 100% ACN, solutes with $\delta_X \approx 10$ (rather low polarity) are properly eluted. Therefore, within the limits of the predictions based on the Hildebrand solubility parameter and the assumption of linearity, solutes in the $10 > \delta_X > 15.5$ range are properly eluted using a 0–100% ACN gradient. Less polar solutes, going down to $\delta_X \approx 8.5$, are eluted by substituting ACN with THF.

The polarity range of solutes properly eluted from a silica column with alkane-isopropanol mixtures in NPLC is depicted in Figure 1.1b. As observed, the solute polarity range is approximately $11.5 < \delta_X < 13.5$, which is inscribed within the range covered by RPLC. Therefore, all analytes eluted by NPLC can be also eluted with optimal retention factors using RPLC. However, this does not mean that NPLC and RPLC have the same or a similar chromatographic value. Thus, hydrophobic samples as mineral and vegetable oils that can be directly injected on an NPLC system are not compatible with most RPLC mobile phases. Furthermore, NPLC and RPLC can provide rather different values of selectivity and efficiency depending on the nature of the solutes.
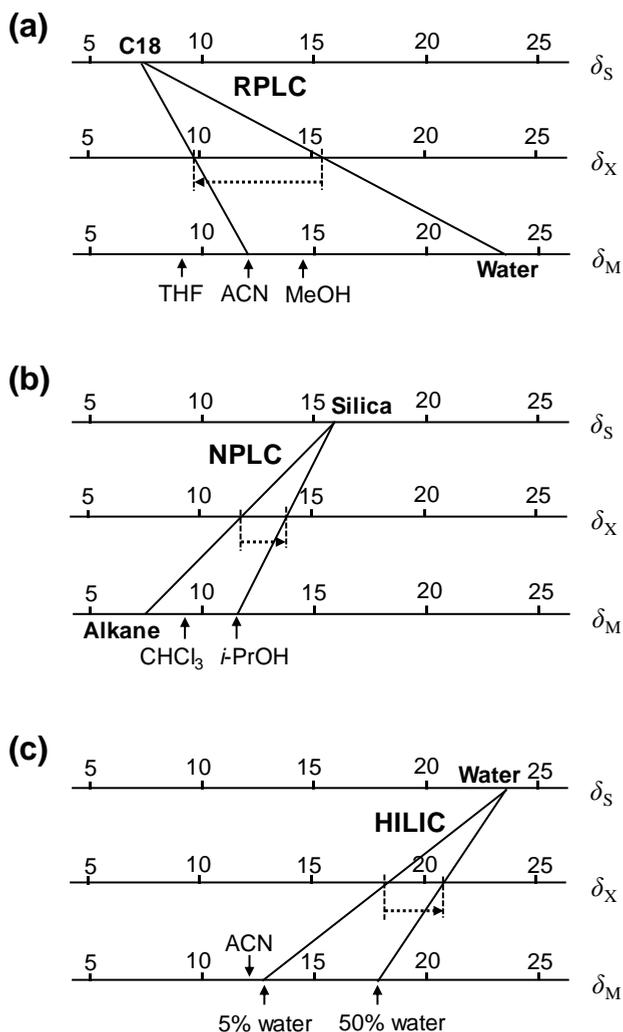
Figure 1.1. Graphical expression of the Schoenmakers' rule. Within the limits of predictions based on the Hildebrand solubility parameter, range of global polarity of solutes that are properly eluted when a wide elution gradient is applied for: (a) RPLC with C18 and ACN-water; (b) NPLC with underivatised silica and isopropanol-heptane; (c) HILIC with a water layer and water-ACN. The Hildebrand global polarity of the stationary phase, solute and mobile phase are represented on the δ_S, δ_X, and δ_M scales, respectively.

Finally, in HILIC, where solutes are retained on a water layer ($\delta_S \approx 23.5$, Figure 1.1c), highly polar solutes in the $18 < \delta_X < 21$ range (mainly ions, polyions, or zwitterions) are eluted with water-ACN mixtures by increasing water from 5% to 50%. However, a problem with HILIC is that the samples and polar analytes should be soluble in the organic-rich mobile phases that are required, mainly at the beginning of the gradient.

1.4. Isoeluotropic mixtures

Fine tuning of the polarity through discrete or continuous changes of the mobile phase composition in the isocratic and gradient elution modes, respectively, is mainly achieved by adjusting the modifier concentration in the solvent mixture. On the other hand, the selectivity is controlled by changing the solvent nature, and for some solutes, by also modifying the mobile phase pH [14], or column temperature [15,16]. For ionic analytes, the concentration of an ion-pairing salt is also an important factor. The selectivity depends mainly on the specific interactions of solutes with the stationary and mobile phases [17,18], that is, on the profile of the contributions to the global polarity of solutes and phases.

A basic question in selectivity optimisation is how to modify the nature of a solvent mixture without altering the selected elution strength. Mixtures with the same elution strength but prepared with different modifiers are called isoeluotropic mixtures. For binary mixtures of MeOH, ACN, or THF with water, from Equation (1.3), and using the Hildebrand parameter as a measure of global polarity assuming a linear behaviour,

$$
\begin{aligned}
\delta_{\text{MeOH}}\ \varphi_{\text{MeOH}} + \delta_{\text{H2O}}\ (1 - \varphi_{\text{MeOH}}) &= \delta_{\text{ACN}}\ \varphi_{\text{ACN}} + \delta_{\text{H2O}}\ (1 - \varphi_{\text{ACN}}) \\
&= \delta_{\text{THF}}\ \varphi_{\text{THF}} + \delta_{\text{H2O}}\ (1 - \varphi_{\text{THF}})
\end{aligned}
\tag{1.7}
$$

By substituting the polarity values given in Table 1.1,

$$\varphi_{MeOH} = 1.27 \, \varphi_{ACN} = 1.60 \, \varphi_{THF} \tag{1.8}$$

Hence, the elution strength of an aqueous mobile phase with 20% MeOH is approximately the same as for 15.7% ACN or 12.5% THF. Since THF is the most hydrophobic solvent, the same elution strength is achieved with a smaller percentage of organic solvent. As indicated previously, the predictions of elution strength depart from linearity at large modifier concentrations. To address this problem, non-linear relationships and nomograms, such as that shown in Figure 1.2, can be used. On this nomogram, all possible isoeluotropic binary mixtures constituted by water and either ACN, MeOH, or THF can be estimated. ACN is generally stronger than MeOH, and THF appreciably stronger than ACN. Note that the scale for ACN is linear, making it necessary to draw non-linear scales for MeOH and THF. However, due to the limitations inherent in the global polarity parameters, predictions are rough and depend largely on the solute properties.
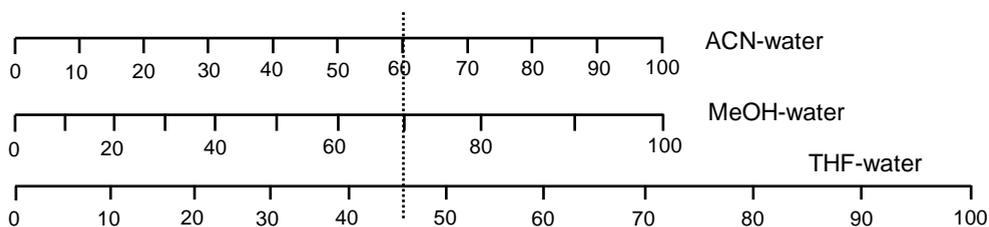
Figure 1.2. Nomogram showing isoeluotropic binary mixtures in RPLC. The compositions are obtained by connecting the solvent scales with a vertical line. The example indicates that aqueous binary mixtures having 60% ACN, 70% MeOH, or 46% THF are isoeluotropic. Adapted from Sigma-Aldrich.com/Supelco 2009-2010 chromatography products catalog, p. 38.

1.5. Solvent-selectivity triangles

1.5.1. The Snyder's solvent-selectivity triangle

Mobile phase selectivity is understood as a consequence of the particular profile of the contributions of solvent-solvent intermolecular interactions to the global polarity. Six types of interactions are considered to contribute to the Hildebrand solubility parameter [10]: interactions between permanent dipoles, between induced dipoles, between permanent and induced dipoles, hydrogen ion donation (acidity), hydrogen ion acceptance (basicity), and electrostatic interactions. However, as commented below, these are not the only possible interactions. Owing to the different contributions, if solutes with exactly the same global polarity but structural differences are separated by chromatography, retention times will be close but still different. We could add "fortunately different", because otherwise selectivity optimisation would not be possible.

To deal with more than three parameters, multivariate statistics is required, where the solvents in the multivariate space are projected on the reduced space of the first principal components [2]. However, in the strategy proposed by Snyder in 1974 [6,19], electrostatic interactions are neglected and some of the most akin interactions (among permanent and induced dipoles) are summarised in a single property called dipolarity (i.e., polarity and polarisability). Accordingly, mobile phase selectivity was characterised by only three parameters: acidity, basicity, and dipolarity. This made possible plotting solvent properties on a triangular diagram, called the Snyder's solvent-selectivity triangle (SST), where each corner represents one of the properties (Figure 1.3) [20].
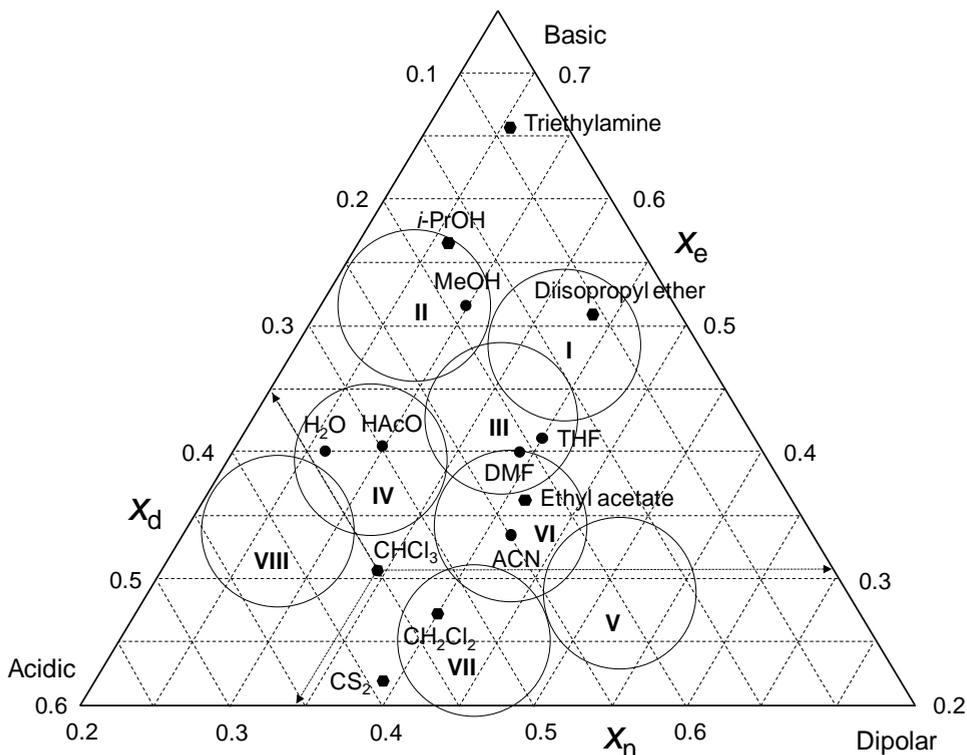
Figure 1.3. Snyder's solvent-selectivity triangle, indicating the eight solvent families (large circles). The location of several solvents, including those most commonly used in RPLC and NPLC, is indicated (DMF, dimethylformamide; HAcO, acetic acid; *i*-PrOH, isopropanol). The arrows starting from chloroform illustrate how to read the scales.

The solvent properties were estimated using three probes: ethanol (e), 1,4-dioxane (d), and nitromethane (n), which is a simplification of the six-probe system formerly proposed by Rohrschneider to represent solvent properties. By using these three probes, the intended properties are: "hydrogen ion donor" (ethanol), "hydrogen ion acceptor" (1,4-dioxane), and "polar or polarisable"

(nitromethane). In fact, none of the three probes represents these characteristics uniquely: ethanol is predominantly a hydrogen ion donor but also a weak acceptor and is moderately dipolar; 1,4-dioxane is a good hydrogen ion acceptor, weakly dipolar and a non-hydrogen ion donor; and nitromethane is strongly dipolar but also both weakly acidic and weakly basic. Although far from ideal, the selected probes led to a useful classification of solvents.

Solvents were characterised according to their capacity to interact with the three probes, which was estimated from gas-liquid partition equilibria. Snyder's global polarity, P' (Table 1.1), was defined as the sum of the three contributions:

$$P' = \log k_{e}^{'} + \log k_{d}^{'} + \log k_{n}^{'} \tag{1.9}$$

where $k_{e}^{'}$, $k_{d}^{'}$, and $k_{n}^{'}$ are the gas-liquid partition coefficients for the probes, which were determined from their equilibrium concentrations in a sealed vial, containing a fixed volume of the solvent to be characterised. The partition coefficients were defined as the ratio of the solute concentration in the solvent and in the vial void volume, after making two corrections to eliminate the effect of the solvent volume and the non-specific contributions (C–H weak permanent or induced dipole interactions, obtained with n-octane). Finally, to eliminate the differences among the global polarities of the solvents, normalisation was performed:

$$1 = \frac{\log k_{e}^{'}}{P'} + \frac{\log k_{m}^{'}}{P'} + \frac{\log k_{n}^{'}}{P'} = x_e + x_d + x_n \tag{1.10}$$

where x_e represents the basic character, x_d is the acidic character, and x_n is the dipolar character of the solvent (Table 1.2). Using this approach, the character of a solvent is defined by the balance or profile of these three normalised

parameters, independently from its global polarity. It is therefore assumed that a solvent that preferably retains ethanol or 1,4-dioxane rather than nitromethane should have a predominantly basic and acidic character, respectively; and a solvent that preferably retains nitromethane rather than the other two probes has a polar character or is readily polarisable rather than a proton donor or acceptor.

The x_e, x_d, and x_n data for a large number of solvents are plotted on the SST (Figure 1.3). Solvents are grouped according to their properties in eight families: (I) aliphatic ethers and amines; (II) aliphatic alcohols; (III) pyridine and THF; (IV) glycols and acetic acid; (V) dichloromethane and dichloroethane; (VI) aliphatic ketones, esters, 1,4-dioxane, and nitriles; (VII) aromatic hydrocarbons and nitrocompounds; and (VIII) phenols and water. The scales should be read counterclockwise: x_e is represented on the right side (the higher on the scale, the stronger is the basic character of the solvent), x_d is on the left side (the lower on the scale, the stronger is the acidic character), and x_n is on its base (with the solvent dipolarity increasing to the right).

The diagram shows that the most common solvents in RPLC provide different selectivity, since they have rather different profiles of the three properties defined in the SST. Thus, water is a strong hydrogen ion donor and acceptor (it is situated at half-height in the SST), but a weak dipole (it is on the left). ACN is less acidic than water but appreciably more dipolar. MeOH is appreciably more basic (higher in the diagram), more dipolar than water, and less dipolar than ACN. Finally, THF has both acidic and basic character, but it is more dipolar than water.

Table 1.2. Normalised selectivity factors

Solvent[a]	Derived from gas-liquid partition data of Rohrschneider's probes[b]			Derived from Kamlet-Taft solvatochromic parameters[c]		
	x_d	x_e	x_n	α	β	π^*
Diisopropyl ether	0.10	0.51	0.39	0.00	0.64	0.36
Hexane	–[d]	–[d]	–[d]	0.00	0.00	0.00
Carbon disulfide	0.39	0.22	0.39	0.00	0.10	0.90
Triethylamine	0.07	0.61	0.32	0.00	0.84	0.16
Carbon tetrachloride	0.38	0.30	0.32	0.00	–[d]	0.59
Ethyl acetate	0.23	0.34	0.43	0.00	0.45	0.55
Toluene	0.28	0.25	0.47	0.00	0.17	0.83
Tetrahydrofuran	0.20	0.38	0.42	0.00	0.49	0.51
Chloroform	0.35	0.31	0.34	0.43	0.00	0.57
Dichloromethane	0.33	0.27	0.40	0.20	0.00	0.82
Methyl ethyl ketone	0.22	0.35	0.43	–[d]	–[d]	–[d]
Acetone	0.23	0.35	0.42	0.06	0.38	0.56
Carbon disulfide	0.39	0.22	0.39	0.00	0.10	0.90
1,4-Dioxane	0.24	0.36	0.40	0.00	0.40	0.60
Pyridine	0.22	0.41	0.36	0.00	0.42	0.58
Isopropanol	0.19	0.55	0.27	0.35	0.43	0.22
1-Butanol	0.19	0.59	0.25	0.37	0.41	0.22

Table 1.2 (continued).

Solvent[a]	Derived from gas-liquid partition data of Rohrschneider's probes[b]			Derived from Kamlet-Taft solvatochromic parameters[c]		
	x_d	x_e	x_n	α	β	π^*
2-Methoxyethanol	0.24	0.38	0.38	–[d]	–[d]	–[d]
Dimethylformamide	0.21	0.39	0.40	0.00	0.44	0.56
Ethanol	0.19	0.52	0.29	0.39	0.36	0.25
Dimethylsulfoxide	0.27	0.35	0.38	0.00	0.43	0.57
Acetonitrile	0.27	0.31	0.42	0.15	0.25	0.60
1-Propanol	0.19	0.54	0.27	0.36	0.40	0.24
Acetic acid	0.31	0.39	0.30	0.54	0.15	0.31
Methanol	0.22	0.48	0.31	0.43	0.29	0.28
Formamide	0.33	0.38	0.30	0.33	0.21	0.46
Water	0.37	0.37	0.25	0.43	0.18	0.45

[a] Solvents ordered according to Table 1.1. [b] Large values of x_d, x_e, and x_n denote good hydrogen ion donor, good hydrogen ion acceptor, and large permanent or induced dipole moments, respectively [5–7]. [c] α, β and π^* represent solvent ability to interact as hydrogen ion donor, hydrogen ion acceptor, and by polar and polarization effects, respectively [8,9]. [d] Not available.

The SST scales should not be interpreted as "percentages" of the intended properties, since solvent properties were obtained from solutes with a mixed character, and therefore, the vertices do not represent "pure" properties. For example, a strongly basic solvent such as triethylamine is not located close to the upper vertex due to its basicity but because it strongly retains ethanol and weakly retains 1,4-dioxane and nitromethane. Ideally, if the SST scales would correspond to pure properties (each vertex representing 100% acidity, 100% basicity, and 100% dipolarity), mixtures of three hypothetical solvents, each one located at each vertex, would provide a whole universe of possibilities. However, such solvents do not exist. Furthermore, real solvents located close to the SST vertices are not mutually miscible or are not compatible with common stationary phases. ACN, MeOH, and THF are at intermediate locations in the SST, being excellent choices to achieve a wide range of properties in RPLC. Not surprisingly, these solvents were already popular by the time the SST was developed.

1.5.2. Prediction of the character of solvent mixtures

The SST allows predicting whether the elution strength will increase or decrease for certain solutes when one modifier is replaced by another. For example, substituting a MeOH-water mixture with an isoeluotropic ACN-water mixture will reduce the ability of the mobile phase to accept hydrogen ions, so the elution strength will be reduced for acidic solutes. Simultaneously, the dipolar character of the mobile phase will increase so that dipolar and polarisable compounds will elute earlier. This reasoning can be of help in solute identification. Thus, if a solute elutes earlier when a MeOH-water mixture is substituted with an isoeluotropic ACN-water mixture, then the solute should have a basic or a dipolar character or both.

As shown in the SST of Figure 1.4, the character of all possible mixtures of water, ACN, MeOH, and THF is delimited by straight-lines connecting the four solvents. This figure illustrates how wide the selectivity range in RPLC is. The character of isoeluotropic mixtures of the four solvents, at increasing elution strength, is indicated by the three small *a*, *b* and *c* triangles. The location of these isoeluotropic mixtures on the SST was established according to their compositions obtained from the nomogram of Figure 1.2. A linear variation of the properties with modifier concentration was also assumed. The small triangles *a*, *b* and *c* of Figure 1.4 illustrate how the character of a mixture of solvents is modified by varying its composition, while maintaining a constant elution strength, as estimated by the Hildebrand solubility parameter, δ.

1.5.3. A solvatochromic solvent-selectivity triangle

The essential conclusion of the Snyder's SST and other alternative diagrams also based on solvatochromic properties, independently from the approach used to construct them, is that, to explore the full range of possibilities during mobile phase selectivity optimisation, solvents having both mutual miscibility and, at the same time, maximal differences in their properties should be selected. Another application of the diagrams is the visualisation of the possibility of substituting a solvent by an equivalent one with improved non-chromatographic characteristics, such as price, availability, or better conformation to the principles of green chemistry. Finally, the diagrams are also useful to predict the miscibility of solvents and the solubility of the solutes in a number of alternative solvents with similar properties. In addition to the Snyder's pioneering work, other solvent descriptors and the diagrams derived from them could be also useful in providing more clarifying and complementary criteria for solvent classification, comparison and selection.
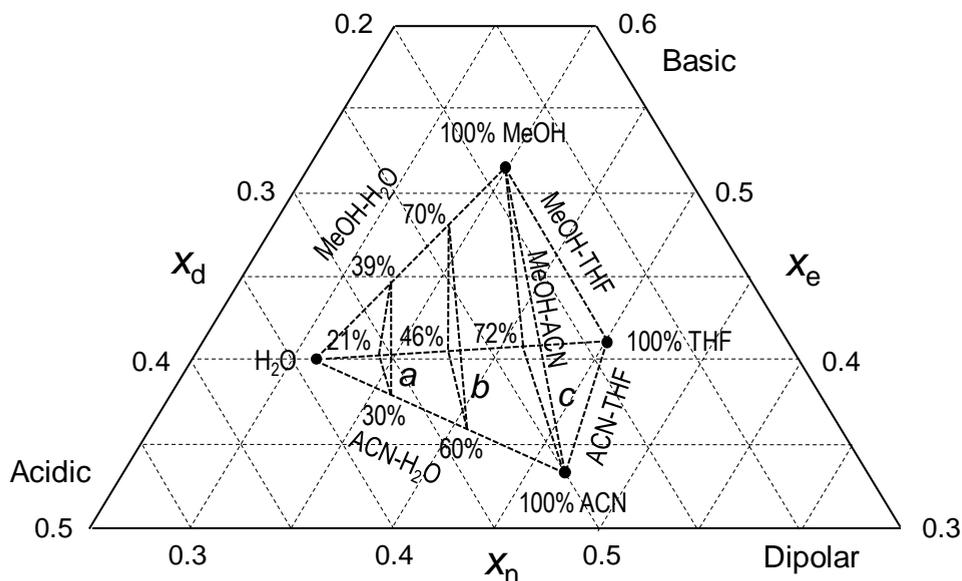
Figure 1.4. Snyder's solvent-selectivity triangle indicating the character of mixtures of water, ACN, MeOH, and THF. The small triangles *a*, *b* and *c* describe isoeluotropic mixtures at increasing elution strength. In *a*, the lowest vertex corresponds to 30:70 ACN-water, the upper vertex to 39:61 MeOH-water, and the left vertex to 21:79 THF-water. Other points on the sides of the small triangle *a* correspond to ternary mixtures, and points inscribed in triangle *a* correspond to quaternary mixtures. Similarly, the small triangles *b* and *c* correspond to isoeluotropic mixtures with respect to 60:40 ACN-water and 100% ACN, respectively.

According to the "mixed" character of the probes used to construct the SST, x_e reflects, in fact, a composite of hydrogen bond basicity, hydrogen bond acidity, and dipolarity; x_d reflects a composite of solvent acidity and dipolarity; and x_n reflects predominantly solvent dipolarity with small contributions from hydrogen bond basicity and acidity. In 1989, Rutan and Carr [7,20,23] substituted the gas-liquid partition coefficients obtained with Rohrschneider's probes by the Kamlet-Taft "solvatochromic parameters" (Table 1.2). These parameters, mainly derived from spectroscopic measurements, separately estimate the hydrogen bond donor (α), hydrogen bond acceptor (β), and dipolarity/polarisability (π^*) properties of solvents as contributors to the global solvent polarity. Solvatochromic parameters are averages over results obtained with several probes. Thus, it is normally assumed that they provide more "pure" measurements of the addressed properties than gas-liquid partition coefficients derived from only three probes. However, reconstruction of the SST using normalised solvatochromic parameters was rather disappointing, since many solvents laid on a line joining the basic and dipolar summits of the triangle, and thus, solvent discrimination was rather poor [20].

1.5.4. Other solvent descriptors and alternative diagrams for solvent classification and comparison

An alternative to the use of the Snyder probes and the Kamlet-Taft solvatochromic parameters are the Hansen parameters [24,25]. These are derived from the Hildebrand solubility parameter, which is split into three contributions:

$$\delta^2 = \delta_d{}^2 + \delta_p{}^2 + \delta_h{}^2 \qquad\qquad (1.11)$$

each one representing the dispersive forces (δ_d), the polarity (δ_p), and the hydrogen bonding (δ_h) (both donor and acceptor). By using the Hansen parameters, an alternative SST to that of Snyder, also showing a good dispersion of solvents according to its character, was constructed.

A somewhat more complex but widely accepted solvent classification system is that based on the five linear solvation energy relationships (LSERs) or Abraham descriptors [26–32]. The solvation parameter model describes five interactions by means of five descriptors related to the compound properties: *E* (the excess molar refraction, related to the presence of *n*- and π-electrons resulting in charge transfer, π–π interactions and dipole-induced dipole interactions); *S* (standing for the presence of dipoles and polarisability); *A* and *B* (describing hydrogen bond acidity and basicity, respectively); and *V* (the McGowan's volume, related to dispersive interaction and cavity energy formation). Representation procedures other than triangles should be used to deal with five descriptors. A possibility is to use projections after a principal component rotation. However, by using principal components, the chemical significance of the axes is lost. An alternative is the use of spider diagrams [33], as that given in Figure 1.5. With this representation technique, a number of parameters above three can be projected on a plane with little loss of information. Careful selection of the order of the axes is essential to minimise the loss of information due to the reduction of the number of dimensions. Thus, those descriptors that are the most positively correlated (for instance *E* and *S* for the LSERs descriptors) should be juxtaposed, in opposition to those that are negatively correlated, while the least correlated ones should be placed as orthogonal as possible. However, as in any other projection technique, compensation of descriptors making rather different solvents to lie in close positions on the spider diagram is possible.
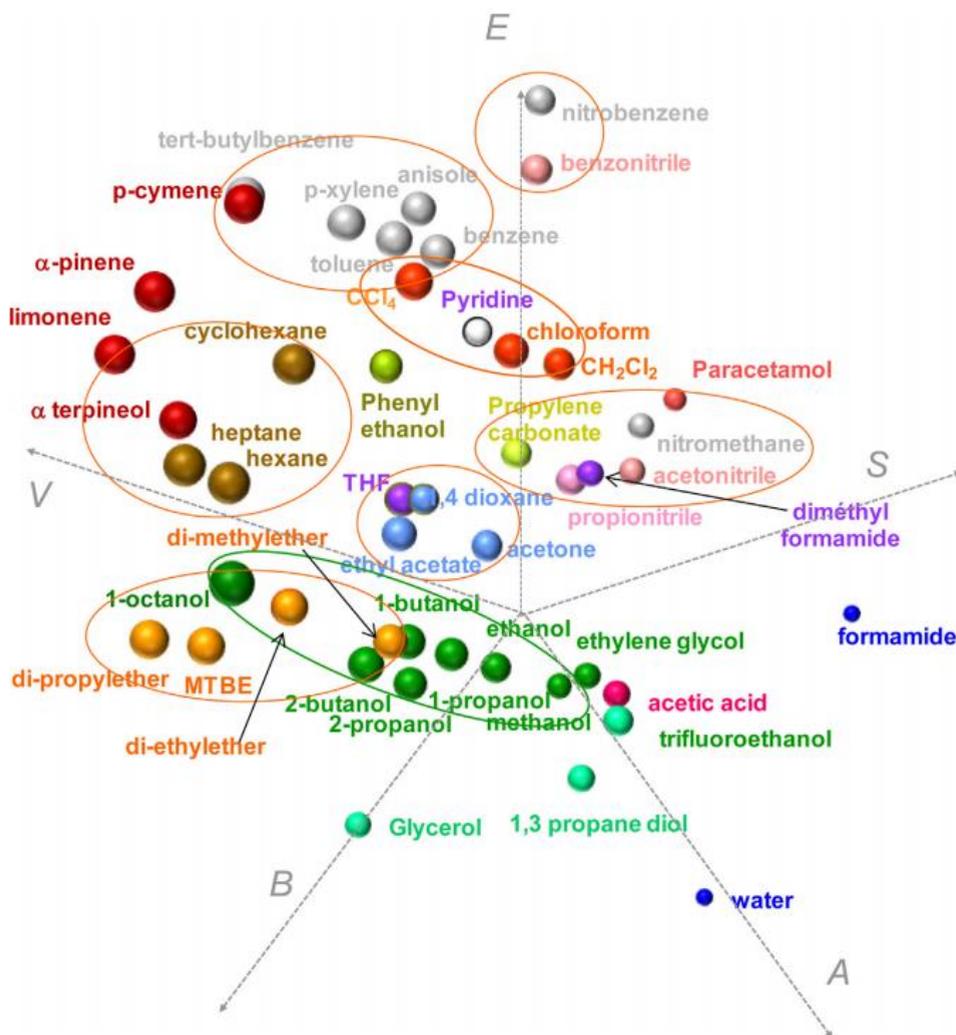
Figure 1.5. Spider diagram based on the Abraham descriptors *E, S, A, B, V*. The point size is proportional to the *V/U* ratio. Reproduced with permission from Ref. [33].

On the spider diagram of Figure 1.5, obtained from the LSER descriptors, water is located at the bottom right, showing its high acidity (A is large) and weak hydrophobicity (V is low). Alcohols, acetic acid and formamide are located close to water. Nitriles (like ACN) display higher dipole interactions and are located at the right-hand side of the plot, above the-alcohols. Alkanes, with high hydrophobicity, are naturally at the opposite of the figure, on the left, close to the V axis. Aromatic solvents are at the top of the diagram, around the E axis. THF, 1,4-dioxane, acetone and ethyl acetate are located in the same group, at the centre of the diagram.

The Abraham descriptors are very useful in explaining the selectivity differences between the three solvents more frequently used in RPLC. Thus, MeOH is the best donor and acceptor of hydrogen bonds, ACN displays the greatest dipolar interactions, and THF, having the greatest McGowan's volume, favours the solubility of most organic compounds through dispersive interactions, explaining its high eluting strength in RPLC.

Finally, the Abraham descriptors also provide a useful global polarity scale defined as:

$$U_X = (E_X^2 + S_X^2 + A_X^2 + B_X^2 + V_X^2)^{1/2} \tag{1.12}$$

where the equation is written for a given solute, X. This global parameter can be used to estimate the elution strength of solvent mixtures, as done above in Equation (1.3) using the Hildebrand parameter. In Figure 1.5, the size of the symbol representing each solvent was made proportional to V/U.

1.6. Practical guidelines for optimisation of mobile phase composition

1.6.1. Selection of the chromatographic mode

The optimisation of the modifier type and volume fraction in the mobile phase is frequently performed on a trial and error basis. Next, some guidelines to rationalise and speed up this process are given. After selecting the chromatographic mode (e.g., RPLC, NPLC, or HILIC), and deciding between isocratic or gradient elution, the elution strength should be adjusted, and finally, the selectivity optimised until all peak pairs of interest are resolved. To select the chromatographic mode, two criteria are attended:

(i) Solute nature. If the solute molecules contain extensive hydrophobic regions in "external" structural parts, they are retained on the hydrophobic RPLC stationary phases. In contrast, if the influence of ionic or polar groups (e.g., $-COOH$, $-OH$, or $-NH_2$) predominates, the solute experiences poor retention and requires polar stationary phases typical in NPLC. A good solution to increase retention of permanent ionic analytes is ion pairing [34]. In this technique, a salt is added to the mobile phase. Retention is enhanced by mixed mechanisms involving association of ions of opposite charge in the hydro-organic mobile phase, and by ion-exchange on the surface of the stationary phase, where the added salt is adsorbed. Since permanent ions and other highly polar solutes are not compatible with NPLC mobile phases, HILIC could be another correct choice. However, a frequent limitation in HILIC is the poor solubility of ionic analytes in the rich organic solvent mobile phases that are required.

(ii) Sample compatibility with the mobile phase. Direct injection of samples soluble in water or in hydro-organic mixtures (e.g., serum, urine, and other aqueous samples or aqueous extracts) require RPLC or HILIC. If

HILIC is selected, the elution strength should be decreased by evaporation of water in the sample, followed by redissolution in a rich ACN mixture, or by dilution with ACN at the cost of a poorer limit of detection. For hydrophobic samples (oils, greases, hydrocarbons, or extracts in heptane, dichloromethane, or other hydrophobic solvents), NPLC is needed. Extracts in solvents that provide high elution strength, such as ethyl acetate in NPLC, or isopropanol in both RPLC and NPLC, should be avoided. It is often possible to change the solvent initially used to extract the sample. For instance, an aqueous sample can be extracted with heptane or dichloromethane, a vegetable oil can be extracted with an aqueous buffer or MeOH, and compounds of interest in an environmental aqueous sample can be concentrated on a solid phase, followed by elution with an appropriate solvent. Within the limits of the analyte's solubility or stability, it is possible to change the solvent nature by evaporation and dilution to make the medium compatible with a given chromatographic mode. Within this context, centrifugal evaporators that allow the removal and substitution of the solvent using vacuum but without boiling thus to prevent analyte losses, are most useful.

1.6.2. Description of the retention using the modifier content as a factor

Solute retention is most commonly controlled by the modifier concentration in the mobile phase. In order to predict the optimal chromatographic conditions, it is convenient to know the retention behaviour as the organic solvent content is varied. In RPLC, the retention for a solute X can be expressed in terms of the solubility parameters according to [35]:

$$\ln k_{\mathrm{X}} = \frac{v_{\mathrm{X}}}{RT}\Big[(\delta_{\mathrm{M}} - \delta_{\mathrm{X}})^2 - (\delta_{\mathrm{S}} - \delta_{\mathrm{X}})^2\Big] + \ln \frac{n_{\mathrm{S}}}{n_{\mathrm{M}}} \tag{1.13}$$

where R is the gas constant, T the absolute temperature, v_{X} the solute molar volume, and n_{M} and n_{S} the moles of mobile phase and stationary phase in the column. For a binary mixture of water and organic solvent, the mobile phase polarity can be calculated as a function of the modifier volume fraction (Equation (1.3)). By substituting Equation (1.13) in Equation (1.3), for binary mixtures, a general-purpose parametric equation is obtained, which is commonly used to characterise the retention [36]:

$$\log k = c_0 + c_1\,\varphi + c_2\,\varphi^2 \tag{1.14}$$

In narrow modifier concentration ranges, the quadratic relationship can be simplified to a linear one, which is very often used.

Surface adsorption in NPLC is better described by non-logarithmic and logarithmic empirical models [37]:

$$\frac{1}{k} = (c_0 + c_1\,\varphi)^n \tag{1.15}$$

$$\log k = c_0 + c_1\,\log \varphi \tag{1.16}$$

where φ is again the concentration of the stronger solvent (here the more polar) in a binary mobile phase. Equation (1.15) has been also found highly satisfactory for RPLC (where φ would be the less polar solvent).

Retention in HILIC is more complex. Equations that combine both partitioning and adsorption phenomena have been suggested [38], such as:

$$\log k = c_0 + c_1\,x + c_2\,\log x \tag{1.17}$$

where x is the fraction of water in the mobile phase. The applicability of the model can be expanded to higher solvent strength regions as follows:

$$\log k = c_0 + c_1 \log x + c_2 (\log x)^2 \tag{1.18}$$

$$\log k = c_0 + c_1 x + c_2 \log (1 + c_3 x) \tag{1.19}$$

1.6.3. Systematic trial and error mobile phase optimisation for isocratic elution

Isocratic elution can be selected if the polarities of the compounds in the sample are similar. In contrast, if the polarities span a wide range then, gradient elution is needed. For an unknown problem, it is preferable to start the optimisation in the gradient elution mode. However, we focus first on the simpler development of an isocratic method.

Usually, in RPLC, a C18 stationary phase is tried first. If no previous information about solute polarities is available, starting with a mobile phase of high elution strength, such as 95% ACN, is advisable. This ensures elution of most compounds in the sample, although many may elute close to the dead time. If the retention of one or more solutes is still too high ($k > 20$), NPLC is probably preferable. Other options are changing the C18 column for C8 or C4 columns, or using a higher column temperature. Less retentive stationary phases, such as C2 or C1, are not recommended, owing to their low stability. Next, the retention of solutes eluting close to the dead time should be increased by using progressively smaller modifier concentrations (e.g., 60, 40, and 20%). At this stage, gradient elution is probably necessary if the solutes of interest cannot be moved to the target range of the retention factor, with any of the modifier concentrations tried.

An analogous strategy can be followed by using NPLC: initially, a polar column (e.g., bare silica or propyl-cyan silica) and a mobile phase with high elution strength are selected. However, the chromatographer should be aware that, in NPLC, a few parts percent of a polar modifier added to the alkane in the mobile phase can cause dramatic effects on retention. For instance, a smaller increase in retention can be produced by decreasing the ethyl acetate concentration from 40% to 2% than from 2% to 0%. This is because, contrary to RPLC where the "strong" solvent is water and not the modifier, in NPLC, the "strong" solvent, which mainly determines the solvating properties of the mixture, is the modifier. Therefore, in NPLC with moderate modifier concentrations, most solutes probably elute close to the dead time. In the absence of excessively retained solutes, the elution strength should be progressively reduced by decreasing the amount of modifier until appropriate retention times are obtained. Similarly, for HILIC, aqueous mixtures containing up to 50% water can be initially tried, followed by the stepwise reduction of the water concentration. The retention mechanism is rather different with hydride silica columns, where the solutes are mainly retained by accepting protons from those covering the stationary phase surface. Elution is promoted by substituting a weak solvent, as ACN, by MeOH, which is a much stronger proton acceptor. Thus, MeOH displaces the analytes from their union sites on the hydride silica stationary phase.

In the three most usual chromatographic modes (i.e., RPLC, NPLC, and HILIC), the selectivity can be further optimised to improve the resolution between all peak pairs. For this purpose, solvent mixtures of similar elution strength, another pH or column temperature, or if necessary, a different stationary phase, can be tried. Here, we will discuss the selection of an isoeluotropic mixture. This may be based on solute properties guided by the

polarity scales described above with the help of any of the triangular or spider diagrams that can be derived. For example, in the RPLC elution of two solutes with the same retention but with different acidity, the more acidic solute elutes earlier if ACN is replaced by MeOH. However, often solute properties are not known or the interpretation of the possible solute-solvent interactions in multi-functional solutes is not straightforward. Therefore, the selectivity is most frequently optimised in an empirical fashion.

In RPLC, by following an empirical experimental scheme, the first modifier to be tested is ACN, due to its low viscosity and short ultraviolet (UV) cut-off wavelength (190 nm) (Table 1.1), which allow a low back-pressure and a UV detection window capable of detecting many absorbing compounds, even if they are poorly conjugated. If the separation is not satisfactory, the second option is MeOH. The viscosity of MeOH-water mixtures is much higher than for ACN-water mixtures, with a maximum at 40% MeOH, which due to the large back-pressures, makes them unsuitable for working at high flow rates with long packed columns, or small particle sizes. Also, the cut-off wavelength of MeOH is higher (205 nm). The third option, THF, has a still higher viscosity, a cut-off wavelength of 212 nm, and requires long equilibration times. Therefore, not surprisingly, these solvents are always tried in the same order: ACN, MeOH, and THF. This is indicated by the A–B–C vertices of the method development triangle (Figure 1.6).
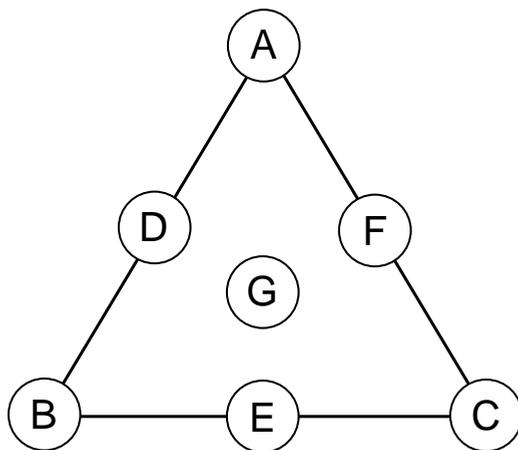
Figure 1.6. Method development triangle. A, B and C represent isoeluotropic binary mixtures of water with ACN, MeOH, and THF, respectively; D–F are isoeluotropic ternary mixtures (e.g., point D is an ACN-MeOH-water mixture, where half of the first modifier has been substituted by an isoeluotropic amount of the second modifier). The central point G is the ACN-MeOH-THF-water isoeluotropic quaternary mixture, where two thirds of the first modifier have been substituted by isoeluotropic amounts of the two other modifiers.

If one of the three isoeluotropic mixtures is successful, the problem is over. If some peaks remain unresolved, ternary or even quaternary isoeluotropic mixtures may be tried. For this purpose, the order of the D–G mixtures in Figure 1.6 is usually followed. After selecting the optimal isoeluotropic mixture, its composition can be slightly changed until all the peaks of interest are satisfactorily resolved. Let us consider a 70:30 ACN-water mixture, for which all peaks for a given sample are in the target range of k values. If the resolution between some peak pairs is unsatisfactory, following the scheme in Figure 1.6 and the nomogram in Figure 1.2, the mobile phase to try next is

78:22 MeOH-water (point B in Figure 1.6). If required, we continue with 52:48 THF-water (point C), 35:39:26 ACN-MeOH-water (point D), 39:26:35 MeOH-THF-water (point E), and so on. Mixtures D and E were calculated by substituting half of the ACN content of the A mixture by its equivalent amount of MeOH or THF, respectively. This trial and error method is more common in practice than the use of considerations based on polarity descriptors, owing to its simplicity, and because it requires no knowledge of solute properties. However, when the problem remains unresolved, either the polarity descriptors or a computer-assisted interpretive optimisation (see Section 1.6.5) is of help. Similarly, selectivity optimisation in NPLC and HILIC can be conveniently carried out by systematically substituting the modifier by other miscible solvents exhibiting a different profile of its descriptors, thus, laying down in a different location on any SST or selectivity spider diagram.

1.6.4. Systematic trial and error mobile phase optimisation for gradient elution

When analysing samples with solutes covering a wide range of polarities, a gradient of elution strength is needed to get both an adequate retention of the first peaks in the chromatogram, and progressively expedite the elution of the most retained solutes. For this purpose, at least two solvent mixtures with different elution strength (mixtures A and B, with B stronger) should be combined. The gradient is normally started at the time of sample injection, although full control on the actual gradient conditions is lost if the delay time, or time required for the gradient to arrive to the column, is not taken into consideration. During the gradient time, t_G (the time the gradient is run), the flow of B and A are increased and decreased, respectively, keeping the sum of the two flows constant, until only B is pumped. To reduce the baseline noise due to fluctuations in the mixture composition, which can be particularly large

with quaternary pumps, A and B mixtures containing at least 5% of the minor solvent, should be used.

In gradient elution, starting with mobile phases with low elution strength, strongly retained analytes migrate very slowly, so that this range of mobile phase compositions does not contribute significantly to their elution. As the elution strength increases along the gradient, the analytes are "accelerated" through the column. A graphical image of the effect is described by: "a solute sits at the head of a column until a strong enough solvent comes along to push it through the column leaving the other solutes behind, then it travels to the column outlet fairly quickly" [39]. The point at which this occurs depends on the strength of solute interaction with the mobile phase and stationary phase. Therefore, solutes in gradient RPLC seldom experience the whole range of mobile phase compositions. The fraction of the solvent composition range that actually affects solute migration has been called "significant solvent concentration range" [40]. Thus, in addition to the chromatographic separation mechanisms, gradient elution also works as a fractional extraction, making the analytes to progress along the column when they are extracted from the stationary phase. In this sense, the elution strength plays in LC an analogous role as temperature in gas chromatography where fractional distillation is a significant separation mechanism.

For the first trial on an unknown sample, a broad gradient with a small slope is recommended to ensure the elution of all solutes (e.g., in RPLC, from 5 to 100% ACN). The ratio $\Delta t/t_G$, where Δt is the difference between the retention times of the first and last peaks of interest in the chromatogram, provides a criterion for deciding whether the sample can be separated isocratically or gradient elution is required. If $\Delta t/t_G < 0.25$, the sample can be isocratically eluted within the k target region by using a mobile phase composition close to

that running when the midpoint in Δt was reached. In contrast, $\Delta t/t_G > 0.25$ means that the solutes elute in a wide k range and isocratic elution is not practical. In this case, the new gradient should be focused between the mobile phase composition at the time of the first eluting peak (start of Δt; new mixture A) and the time for the last peak (end of Δt, new phase B). If the sample contains other components that are more retained than the analytes, then, a final gradient step at a high elution strength should be executed thus to clean up the column. This will prevent cross-contamination between successive injections.

If some peak pairs remain unresolved, the composition of mixtures A and B should be modified without altering significantly their respective elution strengths. In RPLC, this can be achieved by substituting ACN with MeOH or THF, or by using isoeluotropic ternary or quaternary mixtures, as discussed for isocratic elution. When all solutes are satisfactorily resolved, the gradient time can be further reduced without losing resolution. The easiest way is to increase the gradient slope as much as tolerated by the resolution of the least resolved peak pair. Another option is using a segmented or multi-linear gradient, that is, a gradient whose slope changes according to the peak distribution: the slope is smaller in time regions of poorly resolved peaks and steeper in regions without peaks. Non-linear gradients with concave or convex profiles are also occasionally applied when dealing with multicomponent samples requiring extra resolution. Gradients include often isocratic hold periods, at the beginning and/or the end of the runs, or inserted between linear or non-linear gradient segments. Reverse gradients (with decreasing modifier concentration) can be useful in some cases (e.g., to elute amphiphilic analytes whose solubility increases by increasing both the polar and the less polar component of the mobile phase).

In addition to elution strength gradients, it is possible to establish selectivity gradients by increasing the mobile phase acidity, basicity, dipolarity, or any other polarity descriptor, at either constant or increasing elution strength. Therefore, in principle, there are four possibilities:

(i) Isocratic isoselective elution where the mobile phase composition is constant.

(ii) Isocratic elution with a selectivity gradient, obtained by modifying the solvent mixture in such a way that the polarity descriptors, for instance acidity, basicity or dipolarity are varied while a global polarity descriptor is maintained invariable. This entails the continuous modification of the coordinates of the mixtures used on an SST or a selectivity spider diagram, with the restriction of not modifying δ_X (Hildebrand solubility) or U_X (Abraham global polarity, see Equation (1.12)). For example, on the Snyder's SST a selectivity gradient is obtained by following any line along the sides of the a, b or c small triangles in Figure 1.4 that correspond to isoeluotropic mixtures. Obviously, any translation along the triangle surface implies a change in selectivity.

(iii) Isoselective gradient elution where the elution strength is increased but the selectivity is not modified. Isoselective gradients are implemented by using A and B mixtures corresponding to the same profile of normalised polarity descriptors (e.g., to the same point on a given selectivity diagram), but where solvent mixture B has a higher global polarity than solvent mixture A. Then, as the B/A ratio increases, the global polarity of the mixture increases but without a substantial modification in selectivity.

(iv) Double gradient elution where both elution strength and selectivity are modified. These are the most common gradients: when the ACN or MeOH content is increased in a mixture with water, not only the global

elution strength increases, but also the polarity descriptors are varied, thus making the coordinates in any SST or selectivity spider diagram also to change. Double RPLC gradients can be programmed by progressively decreasing the water flow while simultaneously increasing the flow for one or even two modifiers at different rates. In this way, the elution strength is increased, and simultaneously, the selectivity is continuously modified in the desired direction (higher acidity, basicity, dipolarity, etc.).

1.6.5. Computer-assisted interpretive optimisation

Finding the best mobile phase composition or gradient to obtain good peak resolution within a short analysis time is not easy. In spite of being particularly slow and inefficient, the trial and error strategies explained previously (or other less systematic ones) are still frequent. Many solute mixtures, however, are so complex that the protocol can be too long and, often, the best (or at least acceptable) conditions are not found. Fortunately, method development can be expedited with more reliable results by applying computer-assisted interpretive strategies [41–45].

The optimisation process includes two steps: system modelling using data from experimental chromatograms, and resolution prediction through computer-simulated chromatograms. In the first step, to fit equations or train algorithms that allow the prediction of retention, a number of experiments as reduced and informative as possible are carried out. Incidentally, in addition to relative retention times, other properties that summarise a chromatogram, such as peak width and asymmetry, are also inferred from the experiments. The aim is to develop models capable of predicting the separation at any new arbitrary condition [46]. Next, based on the models, the separation quality is predicted

for a large number of separation conditions, to find that giving the maximal (or at least an appropriate) resolution of all the peak pairs. In practice, this is done by simulating the sample separation inside a prefixed factorial space, and calculating a numerical value that qualifies the chromatograms, ideally according to the analyst's appraisal of resolution. In addition to resolution, properties such as short analysis time, minimal solvent consumption, or desirable peak profiles (i.e., high efficiencies and low asymmetries) can be optimised.

To assist an interpretive optimisation, several software packages, such as DryLab [47], ChromSword [48], Osiris [49], PREOPT-W [50], and MICHROM [51], have been commercialised. The user can also develop his or her own software with the aid of a spreadsheet or a high-efficiency programming environment, such as MATLAB or R.

1.6.6. Use of combined mobile phases or gradients to achieve full resolution

Conventional HPLC presents major challenges in the analysis of complex samples. When a separation fails, the usual choice is introducing a drastic change in the chromatographic system (column, solvent, pH, temperature and/or use of additives). However, the possibilities of HPLC may be also expanded through other strategies that combine mobile phases or gradients.

Thus, the use of one or more pulses of a weak eluent (e.g., 200 μL water or 500 μL buffer solution on an RPLC system), strategically inserted to alter abruptly the local mobile phase composition, may improve the resolution between poorly separated peaks but with little or no effect on the already resolved neighbouring peaks [52]. This may be very practical when full resolution has been achieved for most analytes. Another approach, termed solvent modulation, consists of introducing individual solvent zones of constant

composition (usually two, A and B, such as 90% and 100% MeOH, or 75% MeOH and 60% ACN), in a varying or repeating sequence into the LC column [53]. The applied sequence is established by the length ratio of the solvent zones A and B within one cycle, and the number of cycles carried out along the elution. Because the solvent zones are separated from one another spatially and temporally, non-ideal solvent-solvent interactions are effectively eliminated, and the overall solute retention is just a linear combination of the retention times in the individual solvent zones. The advantage is that the effect on the chromatogram of changing the length of the zones is easy and accurately predicted. The approach has also been applied in gradient elution, in the so called "relay gradients", which is a special type of segmented gradient where the nature of the modifiers is abruptly changed between segments.

On the other hand, it is not rare to analyse a sample using two different columns or the same column, and two different isocratic or gradient conditions, to separate different target analytes. The possibilities of this approach can be maximally exploited if the two solvent systems are optimised to be complementary [54]: a separation condition focuses on the resolution of some compounds in the sample, while the other analytes remain unresolved, but are optimally resolved in a second (or subsequent) condition(s). When the results of the optimal complementary separation conditions are considered altogether, all analytes are maximally resolved.

The approach using parallel columns may involve different separation modes, such as RPLC and HILIC, to deal with samples comprising analytes in a wide range of polarities. However, for high throughput analyses, performing separate chromatographic runs with different columns is unpractical; thereby the interest in coupling in series RPLC and HILIC columns. However, despite both chromatographic modes use the same solvents, diametrically opposed

concentrations are needed: HILIC needs a high organic solvent content, while RPLC needs a high amount of water. The solvent strength incompatibility between RPLC and HILIC is, however, solved by increasing the ACN content in the eluate from the RPLC column (aimed to separate low polarity solutes) by on-line mixing with ACN to meet the solvent requirements of the HILIC column (aimed to separate highly polar solutes) [55]. Another option is the direct connection of RPLC and HILIC, using a single gradient program starting at a high organic solvent content compatible with both RPLC and HILIC [56].

More sophisticated configurations connect the two columns through valve setups and involve two chromatographic pumps that allow the operation with different solvent systems in a two-dimensional (2D) fashion [57]. The principle of operation is to carry out the off-line or on-line transference of specific fractions of the eluent from the outlet of the first column (which represents the first dimension) to the inlet of the second column (the second dimension). In comprehensive 2D-LC (LC×LC), the whole eluate from the first dimension is chopped into small segments that are continuously separated in the second dimension. Instead of this, in heart-cutting 2D-LC (LC-LC), only selected segments of the first dimension eluate, presumably those containing target unresolved analytes, are transferred to the second dimension for further separation. This is technically much simpler than LC×LC, since the segments can be parked for a time on the head of the column or different columns, until the system is ready to proceed with the elution in the second dimension. Optimisation of the elution conditions and data treatment is also much simpler in LC-LC than in LC×LC. For both approaches, the advantage of exploiting different retention mechanisms, and the freedom to manipulate independently the mobile phase gradient in each column, yield a considerable increase in peak

capacity. Chromatographic optimisation of 2D-LC is non-trivial, but can open enormously the range of resolutions.

1.7. Additional considerations for solvent selection

There may be several reasons to choose a given solvent other than the elution strength and selectivity, or the limits established by solvent viscosity and cut-off wavelength (Table 1.1) [58,59]. Thus, below 220 nm, the baseline drift caused by the differential solvent absorbance can be sufficient to prevent the practical use of certain solvents, such as MeOH or THF. In its turn, MeOH is less expensive and less toxic than ACN, and its higher polarity reduces the risk of buffer precipitation.

In general, solvents producing high backgrounds or baseline drift with the selected detector cannot be used. In this regard, the continuous modification of the concentration of a minor component in the mobile phase might be far more significant in gradient methods than in isocratic approaches. This occurs, for instance, when an absorbing solvent is used with UV detection or when one of the components of the mixture contains a conducting buffer with conductimetric detection, and in all instances with refractometric detection. Also, lot-to-lot variability of solvents can affect UV detection, particularly when working near the cut-off wavelength. A wider range of solvents is compatible with evaporative light scattering, corona-charged aerosol, mass spectrometric and ion-mobility spectrometric detectors; however, non-volatile buffers and low volatility solvents cannot be used with these detectors.

Other desired features are solvent stability, reduced reactivity, and ability to dissolve a wide range of solutes. Thus, THF has the drawback of its relative instability. However, using other ethers instead of THF can be problematic, due to their limited miscibility with water. Analytes can also be affected by

reactivity with certain solvents. For example, higher alcohols (e.g., isopropanol) tend to be less denaturing to biomolecules than MeOH. In fact, one of the reasons that made ACN a popular choice for LC is its ability to dissolve a wide range of compounds with minimal chemical change. Care should be also taken with bacterial growth, which is a source of unexpected and unexplained chromatographic peaks, promoted by certain reagents added to aqueous mobile phases.

Unavailability or legal restrictions should be also attended. For instance, from late 2008 to early 2009, the production of ACN came down giving rise to an important increase in its price. There is also a concern that many volatile organic solvents are toxic or hazardous to human health or the environment (e.g., chlorinated solvents deplete the ozone layer). Therefore, legislation restricting the use of certain solvents can affect their choice or impel finding alternatives for established methods in analytical laboratories.

To reduce solvent consumption and its environmental impact, columns with a narrower internal diameter and/or smaller particle size can be used. Also, solvent recycling technologies can be a solution. All these reduced consumption patterns are supported by commitments to "greener" strategies in an effort to minimise pollution and wastes and increase sustainability. As commented above, several "green" solvents of vegetal origin, mainly terpenes, have been recommended to substitute alkanes. Ethanol and solketal are green alternatives to ACN and MeOH, but with the drawback of their larger viscosity. Also, ethanol is subjected to restrictions in some countries to avoid illegal diversion to human consumption. Acetone is a good green alternative, but the cut-off for UV-Vis detection is large, around 330 nm.

The organic solvent required in RPLC for a given separation can be reduced by using high column temperatures. Commercial equipment for control and programming of column temperature up to 200 ºC, with mobile phase preheating and post-column cooling, as well as bonded-silica columns capable of routinely supporting high temperatures are now available [60]. Preheating is necessary to avoid the loss of efficiency produced by radial gradients within the column. Post-column cooling is also required to prevent boiling of the mobile phase when pressure falls down.

Water becomes less polar at high temperature. This increases its elution strength. From room temperature to 200 ºC, a 5 ºC increase is equivalent to approximately a 1% and 1.3% increase in ACN and MeOH, respectively. This allows the development of water-based greener, environmentally friendly RPLC methods, although at the cost of the additional energy needed to maintain the oven temperatures and preheating and cooling systems [61,62]. Selectivity changes achieved by increasing the temperature are complementary with respect to those produced by modifying the mobile phase composition. These changes are mainly due to a different polarity of the solvent mixture, also depending largely on the solute molecules (derived from entropic, steric, conformational, and ionisation effects). Unfortunately, the elution strength of water is still relatively low below 200 ºC, which in most cases hinders total replacement of organic solvents by water. Further reduction of water polarity can be achieved at temperatures over 200 ºC, but commercial equipment is not available and the choice of suitable stationary phases, capable of standing the harsh conditions, is rapidly reduced.

1.8. References

[1] J.W. Dolan, Selectivity in reversed-phase LC separations, Part I: Solvent-type selectivity, *LCGC North America* 28 (2010) 1022–1027.

[2] M. Chastrette, M. Rajzmann, M. Chanon, K.F. Purcell, Approach to a general classification of solvents using a multivariate statistical treatment of quantitative solvent parameters, *J. Am. Chem. Soc.* 107 (1985) 1–11.

[3] T.J. Bruno, P.D.N. Svoronos, *Handbook of basic tables for chemical analysis,* 2nd ed., CRC Press, Boca Raton, FL, 2003.

[4] B.L. Karger, L.R. Snyder, C. Eon, An expanded solubility parameter treatment for classification and use of chromatographic solvents and adsorbents: parameters for dispersion, dipole and hydrogen bonding interactions, *J. Chromatogr.* 125 (1976) 71–88.

[5] C.K. Poole, *The Essence of Chromatography*, Elsevier, Amsterdam, 2003, p. 369.

[6] L.R. Snyder, Classification of the solvent properties of common liquids, *J. Chromatogr.* 92 (1974) 223–230.

[7] S.C. Rutan, P.W. Carr, W.J. Cheong, J.H. Park, L.R. Snyder, Re-evaluation of the solvent triangle and comparison to solvatochromic based scales of solvent strength and selectivity, *J. Chromatogr.* 463 (1989) 21–37.

[8] M.C. García Álvarez-Coque, G. Ramis Ramos, J.J. Baeza Baeza, Reversed phase liquid chromatography, in *Analytical Separation Science Series*, Vol. 1 (edited by J.L. Anderson, A. Stalcup, A. Berthod, V. Pino), Wiley, New York, 2015, pp. 159–198.

[9] A.A. Younes, C. Galea, D. Mangelings, Y. Vander Heyden, Normal-phase and polar organic solvents chromatography in *Analytical Separation Science Series*, Vol. 1 (edited by J.L. Anderson, A. Stalcup, A. Berthod, V. Pino), Wiley, New York, 2015, pp. 227–244.

[10] J.H. Hildebrand, R.L. Scott, *The solubility of non-electrolytes*, 3rd ed., Dover Publication, New York, 1964.

[11] P.J. Schoenmakers, H.A.H. Billiet, L. de Galan, The solubility parameter as a tool in understanding liquid chromatography, *Chromatographia* 15 (1982) 205–214.

[12] S. López Grío, M.C. García Álvarez-Coque, W.L. Hinze, F.H. Quina, A. Berthod, Effect of a variety of organic additives on retention and efficiency in micellar liquid chromatography, *Anal. Chem.* 72 (2000) 4826–4835.

[13] T.P. Abbott, R. Kleiman, Solvent selection guide for counter-current chromatography, *J. Chromatogr.* 538 (1991) 109–118.

[14] M. Rosés, Determination of the pH of binary mobile phases for reversed phase liquid chromatography, *J. Chromatogr. A* 1037 (2004) 283–298.

[15] R.G. Wolcott, J.W. Dolan, L.R. Snyder, S.R. Bakalyar, M.A. Arnold, J.A. Nichols, Control of column temperature in reversed-phase liquid chromatography, *J. Chromatogr. A* 892 (2000) 211–230.

[16] S. Pous Torres, J.R. Torres Lapasió, J.J. Baeza Baeza, M.C. García Álvarez-Coque, Combined effect of solvent content, temperature and pH on the chromatographic behaviour of ionisable compounds, *J. Chromatogr. A* 1163 (2007) 49–62.

[17] J.L. Rafferty, J.I. Siepmann, M.R. Schure, Understanding the retention mechanism in reversed-phase liquid chromatography: Insights from molecular simulation, *Adv. Chromatogr.* 48 (2010) 1–55.

[18] M.R. Schure, J.L. Rafferty, L. Zhang, J.I. Siepmann, How reversed-phase liquid chromatography works, *LCGC North America* 31 (2013) 630–637.

[19] A.R. Johnson, M.F. Vitha, Chromatographic selectivity triangles, *J. Chromatogr. A* 1218 (2011) 556–586.

[20] L.R. Snyder, P.W. Carr, S.C. Rutan, Solvatochromically based solvent-selectivity triangle, *J. Chromatogr. A* 656 (1993) 537–547.

[21] A. De Juan, G. Fonrodona, E. Casassas, Solvent classification based on solvatochromic parameters: a comparison with the Snyder approach, *Trends Anal. Chem.* 16 (1997) 52–62.

[22] L.R. Snyder, Changing reversed-phase high performance liquid chromatography selectivity. Which variables should be tried first?, *J. Chromatogr. B* 689 (1997) 105–115.

[23] J. Li, Y. Zhang, P.W. Carr, Novel triangle scheme for classification of gas chromatographic phases based on solvatochromic linear solvation energy relationships, *Anal. Chem.* 64 (1992) 210–218.

[24] C.M. Hansen, The three dimensional solubility parameters. Key to paint component affinities. I. Solvents, plasticizers, polymers and resins, *J. Paint Technol.* 39 (1967) 104–117.

[25] C.M. Hansen, A.L. Smith, Using Hansen solubility parameters to correlate solubility, *Carbon* 42 (2004) 1591–1597.

[26] M.J. Kamlet, R.W. Taft, P.W. Carr, M.H. Abraham, Linear solvation energy relationships. Part 9. Correlation of gas/liquid partition coefficients with the solvatochromic parameters, π^*, α and β, *J. Chem. Soc. Faraday Trans. I* 78 (1982) 1689–1704.

[27] M.J. Kamlet, J.L.M. Abboud, M.H. Abraham, R.W. Taft, Linear solvation energy relationships. 23. A comprehensive collection of the solvatochromic parameters, π^*, α, and β, and some methods for simplifying the generalised solvatochromic equation, *J. Org. Chem.* 48 (1983) 2877–2887.

[28] M.H. Abraham, Scales of solutes hydrogen-bonding: their construction and application to physicochemical and biochemical processes, *Chem. Soc. Rev.* 22 (1993) 73–83.

[29] M.H. Abraham, A. Ibrahim, A.M. Zissimos, Determination of solute descriptors from chromatographic measurement, *J. Chromatogr. A* 1037 (2004) 29–47.

[30] J. Jover, R. Bosque, J. Sales, Determination of Abraham solute parameters from molecular structure, *J. Chem. Inf. Comput. Sci.* 44 (2004) 1098–1106.

[31] M.J. Ruiz Ángel, S. Carda Broch, M.C. García Álvarez-Coque, A. Berthod, Effect of ionization and the nature of the mobile phase in quantitative structure-retention relationship studies, *J. Chromatogr. A* 1063 (2005) 25–34.

[32] M. Vitha, P.W. Carr, The chemical interpretation and practice of linear solvation energy relationships in chromatography, *J. Chromatogr. A* 1112 (2006) 143–194.

[33] E. Lessellier, ∑pider diagram: A universal and versatile approach for system comparison and classification. Application to solvent properties, *J. Chromatogr. A* 1389 (2015) 49–64.

[34] M.C. García Álvarez-Coque, G. Ramis Ramos, M.J. Ruiz Ángel, Liquid chromatography, ion pair, in *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering Series* (edited by J. Reedijk), Elsevier, Waltham, MA, 2015.

[35] P.J. Schoenmakers, H.A.H. Billiet, L. de Galan, Description of solute retention over the full range of mobile phase compositions in reversed-phase liquid chromatography, *J. Chromatogr.* 282 (1983) 107−121.

[36] P. Nikitas, A. Pappa-Louisi, Retention models for isocratic and gradient elution in reversed-phase liquid chromatography, *J. Chromatogr. A* 1216 (2009) 1737–1755.

[37] P. Jandera, M. Kučerová, Prediction of retention in gradient-elution normal-phase high performance liquid chromatography with binary solvent gradients, *J. Chromatogr. A* 759 (1997) 13–25.

[38] M.R. Euerby, J. Hulse, P. Petersson, A. Vazhentsev, K. Kassam, Retention modelling in hydrophilic interaction chromatography, *Anal. Bioanal. Chem.* 407 (2015) 9135–9152.

[39] J.W. Dolan, The hazards of adjusting gradients, *LCGC North America* 20 (2002) 940–946.

[40] G. Vivó Truyols, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Estimation of significant solvent concentration ranges and its application to the enhancement of the accuracy of gradient predictions, *J. Chromatogr. A* 1057 (2004) 31–39.

[41] L.R. Snyder, J.L. Glajch, J.J. Kirkland, Theoretical basis for systematic optimisation of mobile phase selectivity in liquid-solid chromatography, *J. Chromatogr.* 218 (1981) 299–326.

[42]   S. Nyiredy, B. Meier, C.A.J. Erdelmeier, O. Sticher, Prisma: a geometrical design for solvent optimization in HPLC, *J. High Res. Chromatogr. & Chromatogr. Comm.* 8 (1985) 186–188.

[43]   L.R. Snyder, M.A. Quarry, J.L. Glajch, Solvent-strength selectivity in reversed-phase HPLC, *Chromatographia* 24 (1987) 33–44.

[44]   J.R. Torres Lapasió, M.C. García Álvarez-Coque, Levels in the interpretive optimisation of selectivity in high-performance liquid chromatography: A magical mystery tour, *J. Chromatogr. A* 1120 (2006) 308–321.

[45]   R. Cela, E.Y. Ordoñez, J.B. Quintana, R. Rodil, Chemometric-assisted method development in RPLC, *J. Chromatogr. A* 1287 (2013) 2–22.

[46]   M.C. García Álvarez-Coque, J.R. Torres Lapasió, J.J. Baeza Baeza, Models and objective functions for the optimisation of selectivity in reversed-phase liquid chromatography, *Anal. Chim. Acta* 579 (2006) 125–145.

[47]   I. Molnar, Computerized design of separation strategies by reversed-phase liquid chromatography: Development of DryLab software, *J. Chromatogr. A* 965 (2002) 175–194.

[48]   S.V. Galushko, A.A. Kamenchuk, G.L. Pit, Calculation of retention in reversed-phase liquid chromatography. IV. ChromDream software for the selection of initial conditions and for simulating chromatographic behaviour, *J. Chromatogr. A* 660 (1994) 47–59.

[49]   S. Heinisch, E. Lesellier, C. Podevin, J.L. Rocca, A. Tchapla, Computerized optimization of RP-HPLC separation with nonaqueous or partially aqueous mobile phases, *Chromatographia* 44 (1997) 529–537.

[50] R. Cela, E. Leira, O. Cabaleiro, M. Lores, PREOPT-W: Off-line optimization of binary gradient separations in HPLC by simulation. IV. Phase 3, *Comp. Chem.* 20 (1996) 315–330.

[51] J.R. Torres Lapasió, M.C. García Álvarez-Coque, J.J. Baeza Baeza, Global treatment of chromatographic data with MICHROM, *Anal. Chim. Acta* 348 (1997) 187–196.

[52] V.V. Berry, Universal liquid chromatography methods: V. "Pulsing" with weak eluent to resolve peaks, *J. Chromatogr.* 321 (1985) 33–43.

[53] J.H. Wahl, C.G. Enke, V.L. McGuffin, Solvent modulation in liquid chromatography: Experimental verification and comparison with conventional premixed mobile phases, *Anal. Chem.* 63 (1991) 1118–1126.

[54] C. Ortiz Bolsico, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Approaches to find complementary separation conditions for resolving complex mixtures by high-performance liquid chromatography, *J. Chromatogr. A* 1229 (2012) 180–189.

[55] S. Louw, A.S. Pereira, F. Lynen, M. Hanna-Brown, P. Sandra, Serial coupling of reversed-phase and hydrophilic interaction liquid chromatography to broaden the elution window for the analysis of pharmaceutical compounds, *J. Chromatogr. A* 1208 (2008) 90–94.

[56] K.R. Chalcraft, B.E. McCarry, Tandem LC columns for the simultaneous retention of polar and nonpolar molecules in comprehensive metabolomics analysis, *J. Sep. Sci.* 36 (2013) 3478–3485.

[57] P. Jandera, Comprehensive two-dimensional liquid chromatography: Practical impacts of theoretical considerations, *Cent. Eur. J. Chem.* 10 (2012) 844–875.

[58] V.J. Barwick, Strategies for solvent selection: A literature review, *Trends Anal. Chem.* 16 (1997) 293–309.

[59] A. Berthod, K. Faure, Solvents in chromatography and electrophoresis, in *Analytical Separation Science Series*, Vol. 1 (edited by J.L. Anderson, A. Stalcup, A. Berthod, V. Pino), Wiley, New York, 2015, pp. 135–158.

[60] S. Heinisch, J.L. Rocca, Sense and nonsense of high-temperature liquid chromatography, *J. Chromatogr. A* 1216 (2009) 642–658.

[61] T.M. Pawlowski, C.F. Poole, Solvation characteristics of pressurized hot water and its use in chromatography, *Anal. Commun.* 36 (1999) 71–75.

[62] R.M. Smith, Superheated water chromatography: A green technology for the future, *J. Chromatogr. A* 1184 (2007) 441–455.

Part 1

INCREASING THE MODELLING CAPABILITY IN LIQUID CHROMATOGRAPHY

CHAPTER 2

MODELLING RETENTION AND PEAK SHAPE OF SMALL POLAR SOLUTES ANALYSED BY NANO-HPLC USING METHACRYLATE-BASED MONOLITHIC COLUMNS

2.1. Abstract

The development of methacrylate-based monolithic columns was studied for the separation of pharmaceutical hydrophilic compounds in nano-liquid chromatography. The selected polymerisation mixture consisted of 7.5% hexyl methacrylate, 4.5% methacrylic acid and 18.0% ethylene dimethacrylate (*w*/*w*), in a binary porogenic solvent (35:35 *w*/*w* 1-propanol/1,4-butanediol). The polymer synthesised with this mixture has a good permeability, not excessive back-pressure, and reasonable retention times for polar and non-polar solutes. Monolithic columns (12 cm total capillary length, 100 μm i.d.), prepared with this mixture, were able to produce hydrogen bonding and electrostatic interactions, giving rise to promising separations. To evaluate the chromatographic system, alkylbenzenes (neutral and hydrophobic compounds) and sulphonamides (hydrophilic drugs) were assayed. To optimise the chromatographic mobile phase in isocratic elution and characterise the retention mechanism for a mixture of eight sulphonamides, the performance of several mathematic models was checked in the description of retention. The behaviour of the monolithic capillary column was compared, in terms of selectivity and peak profile, to that obtained with a C18 column (9 cm × 4.6 mm i.d., 5 μm particle size) using a conventional HPLC equipment. The results revealed substantial differences in the interactions established, for sulphonamides, between the monolithic and C18 columns.

2.2. Introduction

The seminal work by Svec and Fréchet in 1992 started the era of organic polymer monoliths [1]. In the two last decades, the use of polymer-based monoliths as stationary phases has gained a large acceptance as an alternative to conventional silica-packed columns in liquid chromatography, due to their advantages in terms of low back pressure, good permeability, as well as wider pH range compatibility [2–4]. The fast and easy preparation, together with the possibility to adjust the morphological properties and functionality to obtain the desired chromatographic aspects, has extended their use in different chromatographic separation techniques, such as conventional liquid chromatography [5,6], capillary and nano-liquid chromatography [7,8], and capillary electrochromatography [9,10], among others.

Numerous investigations have been addressed by different research groups in order to study and control the separation properties of the monolithic polymers [11–14]. During the monolithic polymerisation, the functional monomer composition has been established as one of the most important parameters to control the selectivity and the separation mechanism of the resulting monolithic stationary phase [15,16]. The polymerisation time and initiation mode have also been determined as key factors to establish the desired morphology [17]. The possibility of adjusting the chromatographic properties of the monolithic polymers by varying the composition of the polymerisation mixture allows adapting the stationary phase to the samples be analysed. As a result, monolithic polymers have been synthesised to separate successfully large biomolecules, such as proteins and DNA [18,19], and/or small non-polar molecules [11,14,20]. In addition, the use of monolithic polymers as solid-phase extraction sorbents for sample pre-treatment has extended significantly in bioanalysis and environmental applications [21,22].

However, the separation of small polar molecules in reversed-phase mode is more limited. Several strategies have been developed to enhance the separation of this type of compounds, based on the increase of hydrophilic interactions between the stationary phase and the analytes. The addition to the polymerisation mixture of polar functional monomers, such as hydroxyethyl methacrylate (HEMA) [23], methacrylic acid (MAA) [24–27], or the combination with divinylbenzene (DVB) monomer to enhance π-π interactions [23,25,28], have resulted in satisfactory separations of different sets of small aromatic polar solutes.

Interpretive strategies are frequently used in conventional high-performance liquid chromatography (HPLC) to find the optimal separation conditions for all, or at least, selected compounds in the sample [29–31]. These strategies are based on the accurate description of the chromatographic behaviour, using mathematical models [29]. These models also offer information on the interactions of the solutes with the stationary phase. The first step in this type of studies consists in gathering information about the chromatographic behaviour of the compounds in the sample, focusing mainly on the retention and covering wide regions of the involved factor(s). The predictive capability of the fitted model for each solute will depend on the quality of the information provided according to an experimental design, which may contain isocratic, gradient, or mixed experimental data. The models allow the prediction of the retention time and other peak properties for particular solutes and under different conditions, in either isocratic or gradient modes. However, the validity of this strategy to characterise and optimise the separation in polymer monolithic columns in capillary/nano-HPLC has still to be proved. In this regard, previous work by Jandera et al. dealing with mathematical modelling of retention behaviour of small compounds in monolithic supports should be mentioned [32,33].

In this work, several polymeric monolithic columns containing different amounts of MAA were prepared according to the results presented by Lin et al. [27], and tested. From these, a polymeric monolith composed of hexyl methacrylate (HMA), MAA, and ethylene dimethacrylate (EDMA) (poly(HMA-*co*-MAA-*co*-EDMA)) was selected owing to its good permeability, not excessive back pressure, and reasonable retention times for non-polar and polar solutes. To study the column performance, several alkylbenzenes and sulphonamides were selected as probe compounds, which were eluted in isocratic mode according to particular experimental designs. The retention data were fitted to several mathematical models used in conventional HPLC (involving different retention mechanisms), in order to characterise the retention behaviour and study the interaction between solutes and column. The results show a regular behaviour for the analysed compounds, which is reproducibly modelled to be further used in the optimisation of a mixture of the analytes.

The behaviour of the capillary monolithic columns was also compared, in terms of selectivity and peak shape, to that obtained with a C18 column, using conventional HPLC equipment. The obtained information revealed substantial differences in the interactions established between sulphonamides with the monolithic and C18 column.

2.3. Theory

2.3.1. Retention modelling

The organic solvent content in the mobile phase is the experimental factor most frequently optimised in liquid chromatography to get appropriate elution strength, selectivity and analysis time. A wide variety of retention models have been proposed in the literature to describe the chromatographic behaviour [29,34–37]. These models allow the prediction of the retention factor (k) as a function of the volumetric fraction of the modifier, φ. In this work, we have considered some of the models most frequently used in RPLC, which are described below.

(i) Logarithmic-linear model

This is the simplest model, extensively used in RPLC [38]:

$$\log k = c_0 + c_1 \varphi = \log k_{\text{w}} - S \varphi \tag{2.1}$$

where

$$k = \frac{t_{\text{R}} - t_0}{t_0 - t_{\text{ext}}} \tag{2.2}$$

t_{R} being the retention time of the compound of interest, t_{ext} the extra-column time, and t_0 the time for an unretained compound (the dead time). Very often, t_{ext} is neglected in the calculation of k. The intercept of the fitted straight-line, $\log k_{\text{w}}$, refers to a mobile phase composed of pure water. The sensitivity of retention to changes in the organic modifier content (the slope S) is a measurement of the elution strength of the mobile phase.

(ii) Bosch-Rosés' model

The model proposed by Bosch et al. [39] describes the retention as a linear dependence, where the polarity contributions of the solute, stationary phase and mobile phase are separated:

$$\log k = (\log k)_0 + p_S (P_M^N - P_S^N) \tag{2.3}$$

p_S and P_M^N are polarity descriptors for solute and mobile phase, respectively, and $(\log k)_0$ and P_S^N quantify the hydrophobicity of the stationary phase; P_M^N is related to the volumetric fraction of organic solvent in the mobile phase. For acetonitrile-water mobile phases (used in this work):

$$P_M^N = 1.00 - \frac{2.068\varphi}{1+1.341\varphi} \tag{2.4}$$

Equation (2.3) can be simplified by grouping the parameters related to the hydrophobicity of the stationary phase ($(\log k)_0$ and $p_S P_S^N$), giving rise to a simpler two-parameter model:

$$\log k = q + p_S P_M^N \tag{2.5}$$

(iii) Logarithmic-quadratic model

Equation (2.1) gives an accurate description of the retention in RPLC only for moderate ranges of organic modifier. For larger ranges, a more complex model is needed [40]:

$$\log k = c_0 + c_1 \varphi + c_2 \varphi^2 = \log k_w - S \varphi + T \varphi^2 \tag{2.6}$$

(iv) Quadratic model with $P_{\mathrm{M}}^{\mathrm{N}}$ transformation

The accuracy of the Bosch-Rosés' model (Equation (2.5)) is also increased by adding a quadratic term:

$$\log k = q + S_1 P_{\mathrm{M}}^{\mathrm{N}} + S_2 (P_{\mathrm{M}}^{\mathrm{N}})^2 \tag{2.7}$$

(v) Neue-Kuss' model

Neue and Kuss proposed a model that shows excellent performance in wide domains of organic modifier [41]:

$$k = k_{\mathrm{w}} (1 + c\,\varphi)^2\, e^{-\frac{B\varphi}{1+c\varphi}} \tag{2.8}$$

where *c* is a curvature parameter, and *B* measures the elution strength.

(vi) Jandera's model

Although initially proposed to describe the retention in normal-phase liquid chromatography (NPLC), the so-called ABM model (named after the *a*, *b* and *m* parameters) has extended to RPLC, offering accurate results [33,42]. Instead of using a logarithmic value for the retention factor, this model is expressed as the inverse of the retention factor:

$$\frac{1}{k} = (a + b\,\varphi)^m \tag{2.9}$$

(vii) Partition and adsorption mixed model

This model differentiates the contributions of adsorption and partition in the chromatographic retention, being useful to evaluate new materials [37]:

$$\log k = c_0 - \frac{c_1\,\varphi}{1+c_2\,\varphi} + c_3\,\varphi + \log\left(1 + c_4\,e^{\frac{c_1\,\varphi}{1+c_2\,\varphi}}\right) \tag{2.10}$$

Parameter c_4 tends to zero when the retention process is mainly driven by adsorption. When this contribution is negligible, this parameter tends to infinity and partition is considered as the most important retention mechanism. The presence of five adjustable parameters (c_0, c_1, c_2, c_3, c_4) forces the user to implement designs including six or more experimental points, in order to fit the retention model with enough levels of freedom. To reduce the required experimental work, Equation (2.10) can be simplified considering the c_3 term is ≈ 0:

$$\log k = c_0 - \frac{c_1\,\varphi}{1+c_2\,\varphi} + \log\left(1 + c_4\,e^{\frac{c_1\,\varphi}{1+c_2\,\varphi}}\right) \tag{2.11}$$

(viii) Adsorption model

In the retention models above, adsorption plays a secondary role on the retention mechanism, being considered as a displacement process. The following model also considers the adsorption mechanism as a process taking place during the separation along the column [43]:

$$\log k = c_0 - r \log(1 + c_2\,\varphi) + c_3\,\varphi \tag{2.12}$$

2.3.2. Powell's method

The Powell's method [44] was designed to find the local minimum of arbitrary functions. It is based on the numerical construction of a set of conjugate (i.e., non-interfering) searching directions, that is, a system of axes pointing to the maximal mathematical gradient of the function to be minimised.

The algorithm starts by a convenient initial estimate of the solute parameters to be fitted, **p**, which is successively refined along iterations. At the start of the process, each initial searching direction corresponds to a specific parameter in the model and it can be expressed as a unitary vector, **u**. In each step, the unidimensional minimum along each direction $\mathbf{u}_i$ is established by finding the scalar λ that minimises the objective function, $F(\mathbf{p})$ (F can be, for example, the sum of squared residuals in a least squares problem), by doing $\mathbf{p}_i = \mathbf{p}_0 + \lambda\,\mathbf{u}$. After a number of iterations, N (which is related to the number of parameters), the direction which led to the best improvement along its corresponding unidimensional search is replaced by $\mathbf{p}_N - \mathbf{p}_0$, and normalised. This completes a cycle. For accelerating the process, after carrying out a certain number of cycles typically matching the number of model parameters, a new set of orthogonal directions is generated, in such a way that one of them is aligned to the pattern discovered along the former iterations, whereas orthogonality grants more efficient exploration capability.

The Powell's method can be classified as a direct search minimisation algorithm, and it is particularly efficient in multidimensional optimisation problems. It has interesting advantages with regard to other algorithms, such as its stability, safety, and the fact of not requiring any knowledge about the function derivatives to evaluate the mathematical gradient. It can be applied whenever derivatives are unknown or too complex to be calculated. It can be a good choice even in problems involving non-differentiable functions. Its efficiency depends critically on the selected unidimensional minimisation algorithm. A Fibonnacci search, the Golden Ratio or the Brent algorithm has been suggested for this purpose [45].

2.4. Experimental

2.4.1. Materials and reagents

In this work, two groups of probe compounds with different characteristics were considered (Table 2.1). The first set of compounds consisted of six alkylbenzenes: toluene (1A), ethylbenzene (2A), propylbenzene (3A), butylbenzene (4A), pentylbenzene (5A), and hexylbenzene (6A), from Riedel de Haën (Seelze, Germany). The second set included eight sulphonamides: sulphathiazole (1S), sulphaguanidine (2S), sulphisoxazole (3S), sulphapyridine (4S), sulphamethazine (5S), sulphadiazine (6S), sulphamethoxazole (7S), and sulphamonomethoxine (8S), from Sigma (Roedermark, Germany).

Stock solutions of the probe compounds containing 1000 μg/mL were prepared in acetonitrile (Scharlau, Barcelona, Spain), and stored at 4 ℃. Working standard solutions were obtained by weekly dilution of the stocks in nano-pure water (obtained with a purification system of Adrona B30 Trace, Burladingen, Germany), and the corresponding amount of acetonitrile to get the mobile phase composition. Duplicate injections were carried out.

For the preparation of the monolithic columns, the following reagents were used: 2,2-azobis[2-methylproprionitrile] (AIBN) from Fluka (Buchs, Switzerland), 1-propanol, 1,4-butanediol from Scharlau, lauryl methacrylate (LMA), hexyl methacrylate, methacrylic acid, and ethylene dimethacrylate, from Aldrich (Milwaukee, WI, USA). Other reagents were: acetone, NaOH, 37% HCl from Panreac (Barcelona, Spain), 3-[trimethoxysilyl]propyl methacrylate (γ-MPS), ethanol, methanol from Scharlau, and nitrogen from Carburos Metálicos (Valencia, Spain). All reagents were of analytical grade or better.

Table 2.1. Identities, structures, dissociation constants (pK_a) and/or octanol-water partition coefficients ($\log P_{o/w}$) for the sets of alkylbenzenes and sulphonamides studied in this work.

Compound	Structure	pK_a	$\log P_{o/w}$[a]
Toluene		–	2.73
Ethylbenzene		–	3.15
Propylbenzene		–	3.69
Butylbenzene		–	4.38
Pentylbenzene		–	4.80
Hexylbenzene		–	5.34
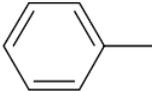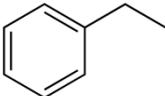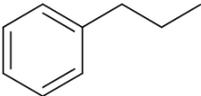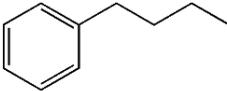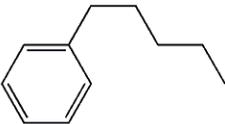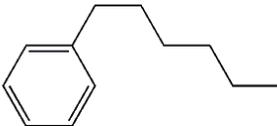
Table 2.1 (continued).

Compound	Structure	pK_a[b]	log $P_{o/w}$[b]
Sulphathiazole		2.1, 7.1	−0.04
Sulphaguanidine		2.8, 12.1	−1.07
Sulphisoxazole		1.8, 5.0	0.81
Sulphapyridine		2.4, 8.2	0.03
Sulphamethazine		2.4, 7.4	0.27
Sulphadiazine		2.0, 6.4	−0.06
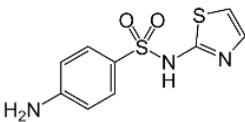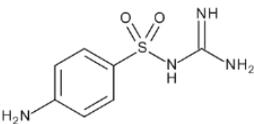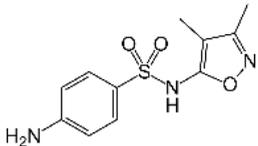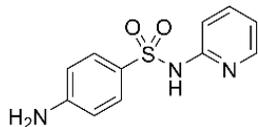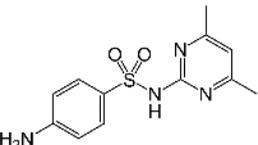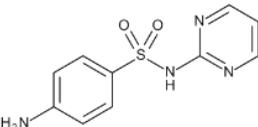
Table 2.1 (continued).

Compound	Structure	pK_a[b]	log $P_{o/w}$[b]
Sulphamethoxazole		1.7, 5.6	0.85
Sulphamonomethoxine		NA	0.74
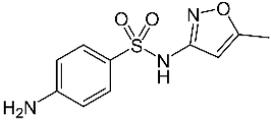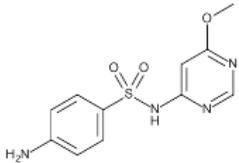
[a] From https://www.drugbank.ca/. [b] From Ref. [46]. NA: not available

2.4.2. Preparation of the monolithic columns

Prior to the monolith polymerisation, the inner surface of the fused silica capillaries was treated to anchor the monolithic stationary phase [15]. With this purpose, the capillaries were sequentially washed with acetone, water, NaOH, HCl, and γ-MPS, and dried overnight under nitrogen steam, to remove all the impurities and activate the inner capillary wall. The polymerisation solution consisted of 7.5% HMA, 4.5% MAA, 18.0% EDMA, and 35% 1-propanol and 35% 1,4-butanediol (all *w/w*) as porogenic solvents. AIBN at 1% (*w/w*, relative to the total amount of monomers) was used as initiator of the polymerisation reaction. The resulting solution was then sonicated during 15 min, and transferred with a syringe pump to the silanised capillary. The extremes of the filled capillary were sealed with gas chromatography septa, and heated for 2 h in an oven at 60 ºC. After the polymerisation, the synthesised monolithic

capillary column was washed with methanol to remove the remaining non-reacted chemicals.

Table 2.2 gives the column-to-column reproducibility, associated to the preparation of poly(HMA-*co*-MAA-*co*-EDMA) capillary monolithic columns, expressed as retention factor. The table gives the results found when the same polymerisation mixture was used to fill the capillaries immediately after its preparation (k_1), and after 15 (k_2) and 26 days (k_3).

2.4.3. Apparatus, columns and experimental designs

The chromatographic analyses with the monolithic nano-columns were carried out with a Thermo Scientific Ultimate 3000 UHPLC instrument, equipped with a nano-pump, autosampler (model Ultimate 3000TPL RS) with 2 mL vials thermostated at 4 ºC, and a variable wavelength detector (model Ultimate 3400 RS) with a 3 nL z-shaped capillary detection cell. The analyses were performed with a flow rate of 1.0 µL/min using 12 cm fused-silica capillaries (100 µm i.d.) (Polymicro Technologies, Phoenix, AZ, USA) for accommodating monolithic stationary phases. The injection volume was 50 nL, and the detection wavelength was set at 214 and 254 nm for alkylbenzenes and sulphonamides, respectively. All injections were performed at room temperature conditions (25 ºC). The dead time was determined in all assayed mobile phases by injection of uracil (U) from Sigma-Aldrich (Steinheim, Germany). The mean value of dead time using the poly(HMA-*co*-MAA-*co*-EDMA) capillary monolithic column was 0.987±0.005 min for alkylbenzenes, and 1.07±0.07 min for sulphonamides. The instrumental extra-column contributions were determined experimentally by removing the column from the system and injecting hexylbenzene and sulphamonomethoxine, using as

eluents 48% and 17% acetonitrile in water (*v/v*), respectively (extra-column time was 0.383 min at both mobile phase compositions).

Table 2.2. Reproducibility, expressed as the variability in retention factor, associated to the preparation of poly(HMA-*co*-MAA-*co*-EDMA) capillary monolithic columns, using the same polymerization batch, and 45% acetonitrile for alkylbenzenes or 10% acetonitrile for sulphonamides.

Compounds	k_1	k_2	k_3
Toluene	3.96	3.90	4.06
Ethylbenzene	5.53	5.48	5.65
Propylbenzene	7.97	7.94	8.13
Butylbenzene	11.63	11.64	11.87
Pentylbenzene	16.68	16.75	17.02
Hexylbenzene	23.91	24.09	24.40
Sulphathiazole	0.82	0.83	0.91
Sulphaguanidine	7.14	7.39	7.93
Sulphisoxazole	8.10	8.35	8.93
Sulphapyridine	9.42	9.59	10.53
Sulphamethazine	14.44	14.56	15.96
Sulphadiazine	35.39	35.48	38.55
Sulphamethoxazole	45.48	45.30	48.73
Sulphamonomethoxine	51.82	51.62	55.21

The work with the conventional column (C18, 9 cm total length, 4.6 mm i.d., and 5 μm particle size, from ACE, Aberdeen, Scotland, United Kingdom) was made using an Agilent (Waldbronn, Germany) instrument, equipped with a quaternary pump (Model 1260 Infinity) run at 1 mL/min, an autosampler (Model 1200) with 2 mL vials, a multiple-variable wavelength UV-visible detector (Model 1200), and a temperature controller (Model 1100) fixed at 25 °C. The injection volume was 20 μL. The dead time was determined by injection of KBr from Acros Organics (Fair Lawn, NJ, USA) (mean value for sulphonamides was 1.04±0.05 min). The extra-column contribution was evaluated similarly to the monolithic column using sulphamonomethoxine (0.14 min).

Chromatographic elution was carried out in the isocratic mode. For the monolithic columns, the mobile phases were prepared with HPLC-grade acetonitrile and nano-pure water containing 0.1% (v/v) acetic acid from Scharlau, and filtered with 0.22 μm Nylon membrane. The retention behaviour of alkylbenzenes and sulphonamides separated with the poly(HMA-*co*-MAA-*co*-EDMA) column using nano-HPLC was studied using experimental designs, taking into account the solutes polarity range. The acetonitrile percentages in water (v/v) were 40, 43, 48, 54 and 60% for alkylbenzenes, and 10, 13, 15, 17, 21 and 25% for sulphonamides. The experimental design for sulphonamides eluted from the ACE C18 column, using conventional HPLC, consisted of 10, 13, 15, 17 and 20% (v/v) acetonitrile in water. The pH was buffered at 3.5 with 0.01 M anhydrous dihydrogen phosphate from Sigma (Roedermark, Germany).

2.4.4. Software

The acquisition of signals for the experiments with the monolithic columns was made with a Chromeleon workstation (Thermo Scientific, version 7.2 SR4). An OpenLAB CDS LC ChemStation (Agilent B.04.03) was used to control the conventional HPLC equipment.

Retention times and peak half-widths were measured with the MICHROM software [47]. For the mathematical treatment, the data were processed with Matlab 2017b (The MathWorks Inc., Natick, MA, USA).

2.5. Results and discussion

2.5.1. Use of lauryl- and hexyl-methacrylate-based monolithic columns to analyse alkylbenzenes and sulphonamides

A series of LMA-based columns were first adapted from previous work [48]. The chromatograms obtained with a hydrophobic lauryl methacrylate column for a mixture of alkylbenzenes and sulphonamides are shown in Figures 2.1a and b, respectively (the column length and diameter of these initial columns were 15 cm and 320 μm i.d., respectively, and were operated with a capillary system). Columns containing only non-polar monomers, such as LMA, resolve alkylbenzenes to the baseline, while sulphonamides appeared totally overlapped, with only one broad peak observed. This should be interpreted as due to the absence of enough interactions of these polar compounds with the column to differentiate each compound.
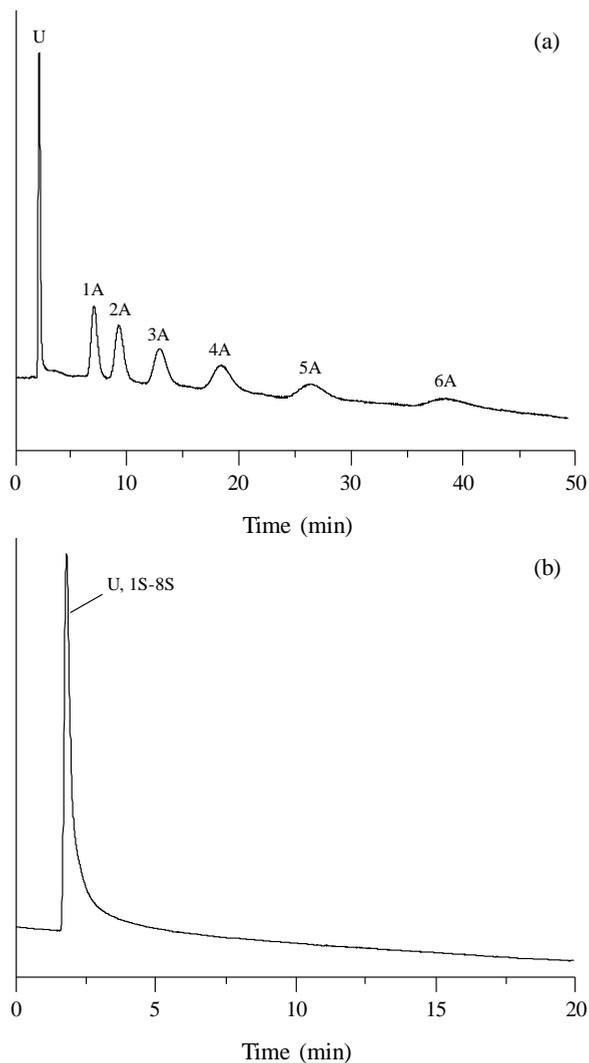
Figure 2.1. Experimental chromatograms showing the separation performance of a mixture of: (a) six alkylbenzenes and (b) eight sulphonamides, eluted using a 15 cm poly(LMA-*co*-EDMA) capillary monolithic column (320 μm i.d.). Composition of the polymerisation mixture (*w/w*): 20% LMA, 30% EDMA, 25% 1,4-butanediol, and 25% 1-propanol. Mobile phase composition (acetonitrile, *v/v*) was: (a) 50% and (b) 20%. Solute identities are given in Section 2.4.1.

In order to separate the sulphonamides, an LMA column was initially used, to which a hydrophilic monomer (MAA) was added, according to the work reported by Lin et al. [27]. As expected, the alkylbenzenes were again resolved to the baseline (Figure 2.2a). In contrast with the poly(LMA-*co*-EDMA) column, sulphonamides interacted with the methacrylate groups of the poly(LMA-*co*-MAA-*co*-EDMA) stationary phase, resulting in the separation of these compounds (Figure 2.2b). However, it was not possible to resolve the most retained compounds.

To achieve a more complete separation of the sulphonamides, still keeping an adequate resolution of the set of alkylbenzenes, the non-polar monomer LMA (with 12 carbons) was replaced with another of lesser hydrophobicity: HMA (6 carbons). Optimal chromatograms using a nano-HPLC system, obtained with the poly(HMA-*co*-MAA-*co*-EDMA) column for alkylbenzenes and sulphonamides are depicted in Figures 2.3a and c, respectively. The aqueous mobile phases contained 40% and 10% acetonitrile (*v/v*), respectively. Note that the elution order is the same for the alkylbenzenes analysed with the poly(LMA-*co*-MAA-*co*-EDMA) and poly(HMA-*co*-MAA-*co*-EDMA) columns, but important changes in relative retention are observed for the sulphonamides, especially for sulphathiazole (1S) and sulphaguanidine (2S).

A detailed study is presented next, which evaluates the retention behaviour and peak profiles obtained with the poly(HMA-*co*-MAA-*co*-EDMA) column, for both sets of compounds.
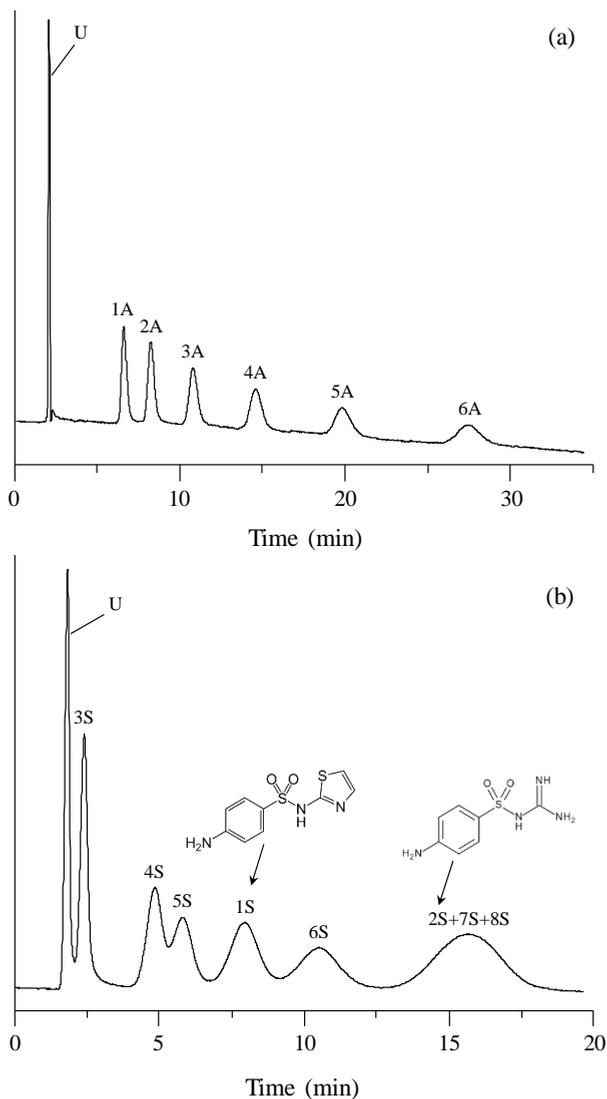
Figure 2.2. Experimental chromatograms showing the separation performance obtained for mixtures of six alkylbenzenes (a), and eight sulphonamides (b), separated with a 15 cm capillary monolithic column (320 µm i.d.), containing 12.5% LMA, 7.5% MAA and 30% EDMA (*w/w*). Mobile phase composition (acetonitrile, *v/v*) was: (a) 50% and (b) 20%. Solute identities are given in Section 2.4.1.
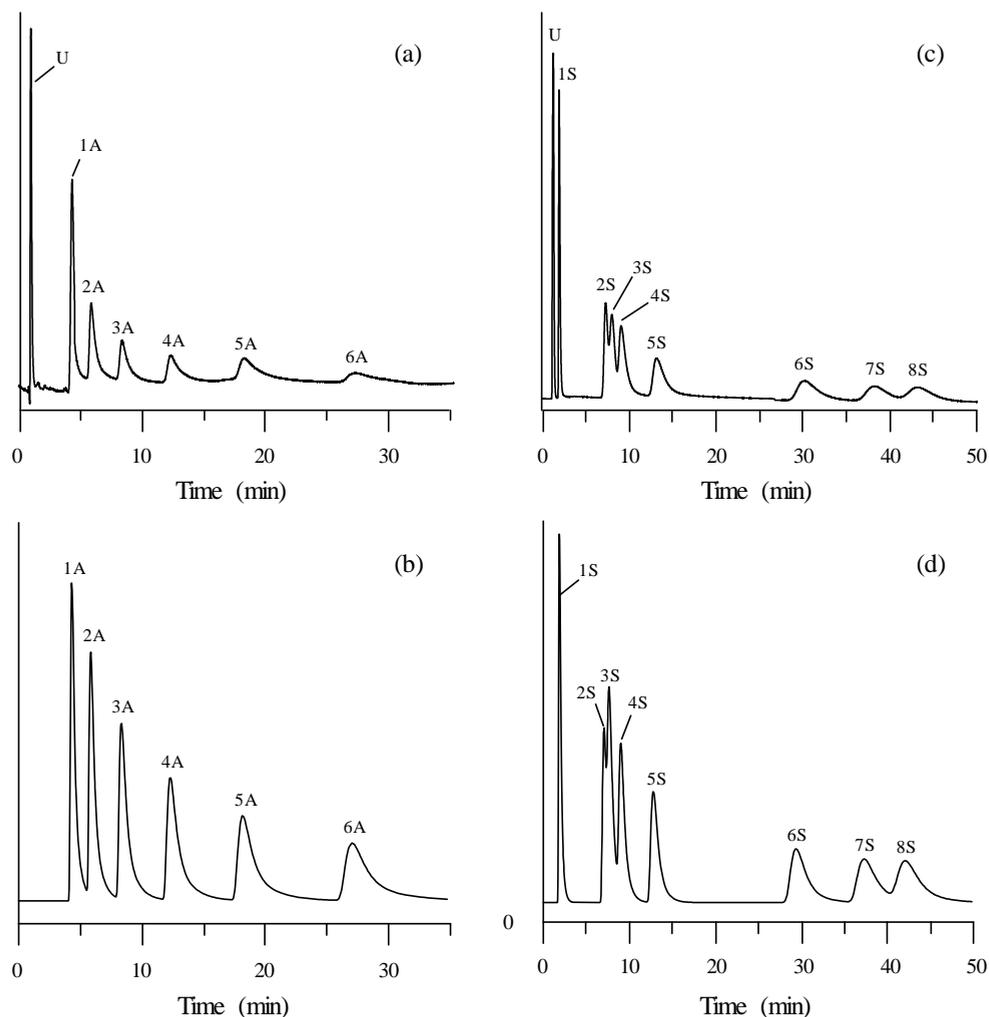
Figure 2.3. Experimental (a,c) and predicted (b,d) chromatograms for alkylbenzenes (a,b), and sulphonamides (c,d), eluted using a 12 cm poly(HMA-*co*-MAA-*co*-EDMA) capillary monolithic column (100 μm i.d.). Composition of the polymerisation mixture (*w/w*): 7.5% HMA, 4.5% MAA, 18% EDMA, 35% 1,4-butanediol, and 35% 1-propanol. Mobile phase composition (acetonitrile, *v/v*) was: (a,b) 40%, and (c,d) 10%. Solute identities are given in Section 2.4.1.

2.5.2. *Retention behaviour of the probe compounds at varying mobile phase composition*

Figure 2.4 shows the retention behaviour for each set of probe compounds, for all mobile phases in the experimental designs (see Section 2.4.3), using the poly(HMA-*co*-MAA-*co*-EDMA) capillary monolithic column. As can be seen, alkylbenzenes (Figure 2.4a) show very regular behaviour, with an almost linear trend. For the phases of low elution strength, the retention of the assayed solutes is sufficiently differentiated to resolve the sample of alkybenzenes in a reasonable analysis time. In spite of the decrease in the hydrophobicity of the column, due to the presence of moderately polar MAA groups in the stationary phase, hydrophobic interactions continue to play an important role in the retention, succeeding in the separation of non-polar samples. The elution order for the group of alkylbenzenes is correlated to the hydrophobicity of each solute (see Table 2.1).

For sulphonamides (Figure 2.4b), the retention behaviour is not as regular as observed with alkylbenzenes: the compounds are distributed into three groups according to their chromatographic behaviour (solute 1S, solutes 2S to 5S, and solutes 6S to 8S). Sulphathiazole (1S) was eluted very close to the dead time marker. Within each group, there are solutes giving rise to close retention, seemingly due to the similar interactions with the stationary phase. This makes complete separation of the probe compounds a challenge. This is especially the case of sulphamethoxazole (7S) and sulphamonomethoxine (8S), which co-elute in most assayed experimental conditions. At high organic modifier contents, peak reversal of both analytes occurs.
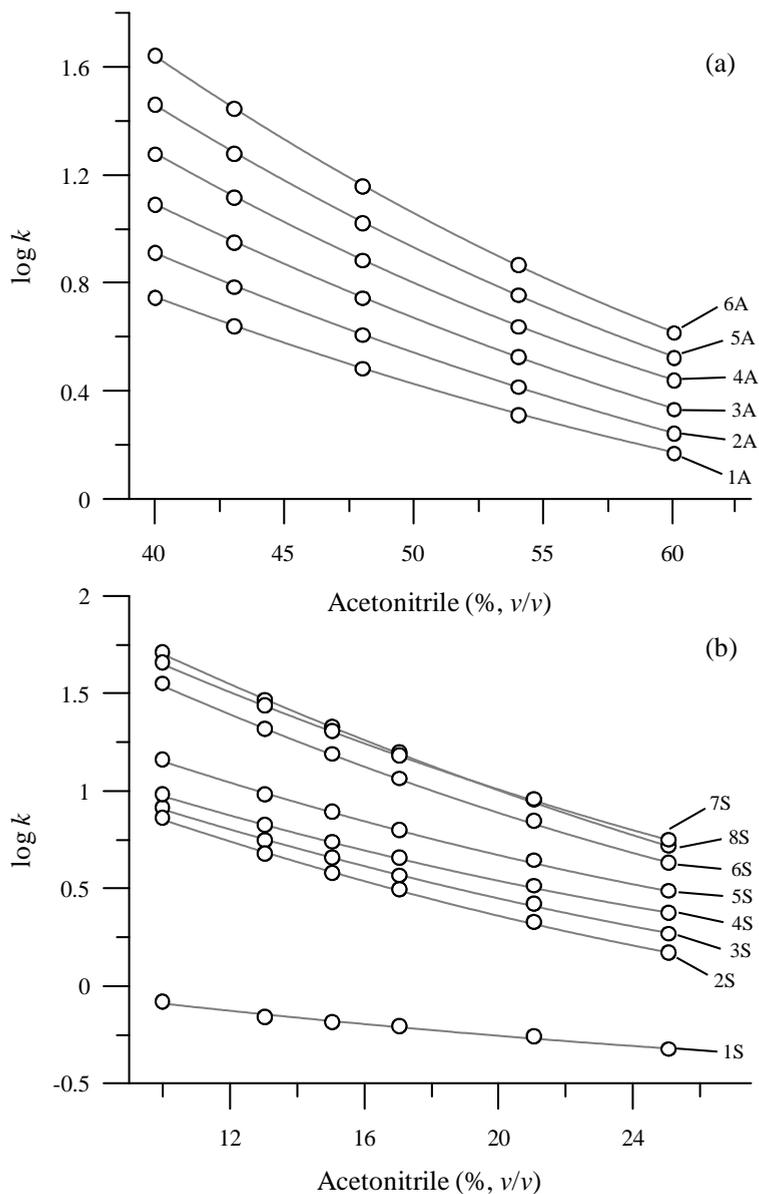
Figure 2.4. Retention behaviour for the poly(HMA-*co*-MAA-*co*-EDMA) capillary monolithic column, expressed as log *k* versus acetonitrile content in the mobile phase for the set of: (a) alkylbenzenes, and (b) sulphonamides. Solute identities are given in Section 2.4.1.

The elution order of sulphonamides is not marked by their hydrophobicity, as can be checked from the log $P_{o/w}$ values in Table 2.1, unlike alkylbenzenes. This can also be observed in Figures 2.5 and 2.6, where the retention times of alkylbenzenes and sulphonamides are plotted versus log $P_{o/w}$ for the assayed columns. This suggests that the retention for sulphonamides should be explained, besides the hydrophobic interactions with the stationary phase, by hydrophilic interactions, due to the presence of ionisable MAA groups in the stationary phase. We will return to this point later.

2.5.3. Selection of the retention model

In this section, the predictive capability of the different retention models described in Section 2.3.1 are commented for the poly(HMA-*co*-MAA-*co*-EDMA) monolithic column used in nano-HPLC. The study was performed for the two sets of probe compounds (alkylbenzenes and sulphonamides). In each case, the retention factors for the probe compounds, eluted with all mobile phases in the experimental designs, were processed. Given that some of the models considered in the present study are non-linear, the retention factors were adjusted by the iterative method of Powell [49] (see description in Section 2.3.2). To evaluate the modelling quality, the following statistics were calculated [50,51]:
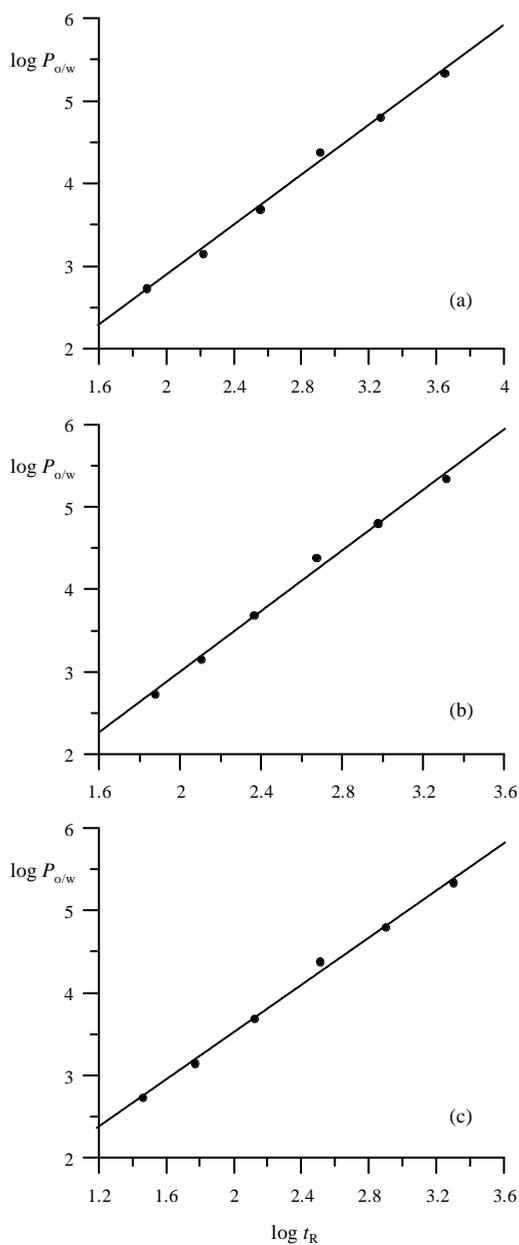
Figure 2.5. Correlation between retention and hydrophobicity for alkylbenzenes, analysed with: (a) poly(LMA-*co*-EDMA), (b) poly(LMA-*co*-MAA-*co*-EDMA), and (c) poly(HMA-*co*-MAA-*co*-EDMA).
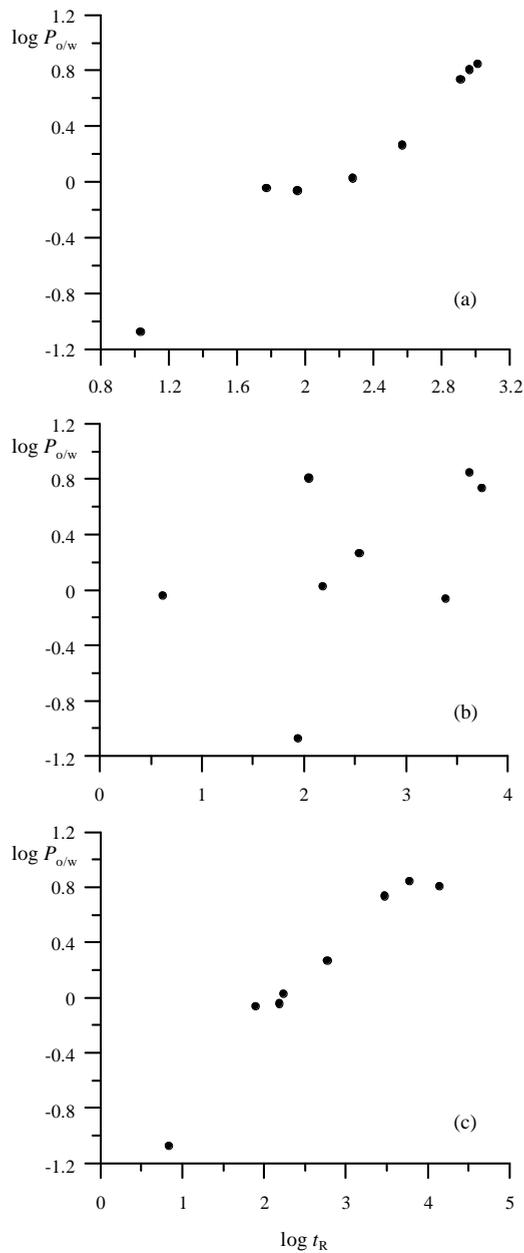
Figure 2.6. Correlation between retention and hydrophobicity for sulphonamides, analysed with: (a) poly(LMA-*co*-MAA-*co*-EDMA), (b) poly(HMA-*co*-MAA-*co*-EDMA), and (c) microparticulate C18.

(i) Adjusted correlation coefficient

$$R_{\text{adj}} = \sqrt{1 - (1 - R^2) \times \frac{ne - 1}{ne - np - 1}} \qquad (2.13)$$

where:

$$R = \sqrt{1 - \frac{\sum\limits_{i=1}^{ne} (\hat{k}_i - k_{\exp,i})^2}{\sum\limits_{i=1}^{ne} (k_{\exp,i} - \bar{k}_{\exp})^2}} \qquad (2.14)$$

ne corresponding to the number of mobile phases in the experimental design, *np* is the number of model parameters, $\hat{k}_i$ and $k_{\exp,i}$ are the predicted and experimental retention factors for each mobile phase *i*, respectively, and $\bar{k}_{\exp}$ is the mean experimental retention factor.

(ii) Mean relative error

$$RE = \frac{\sum\limits_{i=1}^{ne} \left| \hat{k}_i - k_{\exp,i} \right|}{\sum\limits_{i=1}^{ne} k_{\exp,i}} \times 100 \qquad (2.15)$$

(iii) Snedecor's F

$$F = \frac{\sum\limits_{i=1}^{ne} (\hat{k}_i - \hat{k}_{\text{mean}})^2}{\dfrac{\sum\limits_{i=1}^{ne} (k_{\exp,i} - \hat{k}_i)^2}{ne - np}} \qquad (2.16)$$

$\hat{k}_{\text{mean}}$ being the mean predicted retention factor.

Tables 2.3 and 2.4 show the parameters of all the models adjusted in this work, for alkylbenzenes and sulphonamides, analysed with the poly(HMA-*co*-MAA-*co*-EDMA) capillary column, respectively. Owing to the diversity in the transformation of the response (k) in the studied retention models, we have adopted the criterion of performing all fittings as $k = F(p_1, p_2,...)$, so that the scattering k_{pred} *vs.* k_{exp} is uniform for all models.

Tables 2.5 and 2.6 show the values for the different statistics, corresponding to the fitting of the retention data for the sets of alkylbenzenes and sulphonamides, respectively. As can be seen, all the models studied provided very acceptable predictions for both sets of compounds and, although there are differences in performance for each set of compounds, these are not substantial. The models fitted for sulphonamides usually yielded higher relative errors compared to alkylbenzenes.

Table 2.3. Regression parameters for the retention models evaluated, considering the set of six alkylbenzenes.

Solute	Parameters	Retention models							
		Linear[a]	Bosch-Rosés[b]	Quadratic[c]	Quadratic-P_M^{N}[d]	Neue-Kuss[e]	Jandera[f]	Mix[g]	Adsorption[h]
Toluene	p_1	1.96	−1.07	1.96	−0.72	18.85	0.00	3.49	19.35
	p_2	−3.04	3.94	−3.04	2.18	2.38	1.49	12.54	−3.72
	p_3			64.76	5259	50.62	3.30	−0.15	311.5
	p_4							0.01	0.82
Ethylbenzene	p_1	2.32	−1.19	2.32	−0.78	40.38	0.00	7.75	30.67
	p_2	−3.54	4.55	−3.54	2.51	2.72	1.44	33.95	−4.52
	p_3			57.67	5260	64.07	3.81	−2.56	1582
	p_4							2.46	1.50
Propylbenzene	p_1	2.71	−1.31	2.71	−0.84	229.43	0.00	5.91	20.06
	p_2	−4.07	5.19	−4.07	2.85	2.58	1.40	21.89	−5.23
	p_3			66.77	5262	65.97	4.34	−2.36	78.37
	p_4							0.01	1.62
Butylbenzene	p_1	3.13	−1.44	3.13	−0.89	321.95	0.00	5.59	44.87
	p_2	−4.64	5.87	−4.64	3.22	3.45	1.37	19.52	−6.52
	p_3			−1.39	5332	100.5	4.90	−0.28	1924
	p_4							0.00	3.50

Table 2.3 (continued).

Solute	Parameters	Retention models							
		Linear[a]	Bosch-Rosés[b]	Quadratic[c]	Quadratic-P_M^N [d]	Neue-Kuss[e]	Jandera[f]	Mix[g]	Adsorption[h]
	p_1	3.54	−1.57	3.54	−0.96	1014	0.00	7.70	14.37
Pentylbenzene	p_2	−5.21	6.55	−5.21	3.58	3.94	1.35	26.92	−21.77
	p_3			101.4	5283	128.2	5.46	−2.82	3.03
	p_4							3.50	15.64
	p_1	3.95	−1.70	3.95	−1.01	7700	0.00	7.00	18.89
Hexylbenzene	p_2	−5.79	7.22	−5.79	3.93	3.75	1.33	22.45	−12.67
	p_3			84020	5298	128.8	6.01	−1.71	8.40
	p_4							0.00	8.85

[a] Logarithmic-linear model (Equation (2.1)): $p_1 = \log k_w$, $p_2 = S$; [b] Bosch-Rosés's model (Equation (2.5)): $p_1 = q$, $p_2 = p_S$; [c] Logarithmic-quadratic model (Equation (2.6)): $p_1 = \log k_w$, $p_2 = S$, $p_3 = T$; [d] Logarithmic-quadratic model with P_M^N transformation (Equation (2.7)): $p_1 = q$, $p_2 = S_1$, $p_3 = S_2$; [e] Neue-Kuss's model (Equation (2.8)): $p_1 = k_w$, $p_2 = c$, $p_3 = B$; [f] Jandera's model (Equation (2.9)): $p_1 = a$, $p_2 = b$, $p_3 = m$; [g] Mix model (Equation (2.11)): $p_1 = c_0$, $p_2 = c_1$, $p_3 = c_2$, $p_4 = c_4$; [h] Adsorption model (Equation (2.12)): $p_1 = c_0$, $p_2 = r$, $p_3 = c_2$, $p_4 = c_3$.

Table 2.4. Regression parameters for the retention models evaluated, considering the set of eight sulphonamides.

Solute	Parameters	Linear[a]	Bosch-Rosés[b]	Quadratic[c]	Quadratic-P_M^N[d]	Neue-Kuss[e]	Jandera[f]	Mix[g]	Adsorption[h]
Sulphathiazole	p_1	5.99	−1.01	0.06	−0.77	0.00	0.19	1.18	3.07
	p_2	−1.55	1.12	−1.55	0.46	3.60	11.70	50.19	−0.47
	p_3			5246	7.06	40.83	0.63	−1.23	9000
	p_4							0.30	−0.62
Sulphaguanidine	p_1	1.34	−1.97	1.34	−1.20	0.00	0.18	3.07	3.62
	p_2	−4.96	3.45	−1.55	1.38	4.47	2.64	50.29	−146.7
	p_3			8855	5272	69.42	2.44	−3.95	0.62
	p_4							0.15	71.40
Sulphisoxazole	p_1	1.36	−1.71	1.36	−1.00	0.00	0.20	2.20	8.23
	p_2	−4.58	3.20	−4.58	1.29	4.46	2.29	38.42	−0.99
	p_3			8199	5271	67.14	2.47	−3.58	3384
	p_4							0.35	−3.72
Sulphapyridine	p_1	1.40	−1.50	1.40	−0.83	0.00	0.15	1.99	3.70
	p_2	−4.33	3.03	−4.33	1.22	4.48	2.08	33.83	−170.3
	p_3			−6546	5264	65.81	2.17	−3.24	0.53
	p_4							0.44	73.51

Modelling methacrylate-based monolithic columns

Table 2.4 (continued).

Solute	Parameters	Retention models							
		Linear[a]	Bosch-Rosés[b]	Quadratic[c]	Quadratic-P_M^{N}[d]	Neue-Kuss[e]	Jandera[f]	Mix[g]	Adsorption[h]
Sulphamethazine	p_1	1.63	−1.59	1.63	−0.85	0.00	0.14	2.80	10.10
	p_2	−4.82	3.36	−4.82	1.35	4.62	1.93	42.29	−1.10
	p_3			8581	5291	70.42	2.40	−3.74	6425
	p_4							0.25	−3.45
Sulphadiazine	p_1	2.22	−2.24	2.22	−1.19	0.00	0.18	2.95	5.70
	p_2	−6.78	4.63	−6.78	1.84	4.45	1.93	32.86	−174.9
	p_3			7879	5312	79.19	3.59	−5.03	0.63
	p_4							0.34	85.70
Sulphamethoxazole	p_1	2.31	−2.03	2.31	−1.01	0.02	0.27	2.27	5.79
	p_2	−6.59	4.50	−6.59	1.79	3.82	1.66	12.64	−74.12
	p_3			5256	5315	70.03	4.59	−3.74	0.89
	p_4							0.44	43.15
Sulphamonomethoxine	p_1	2.43	−2.33	2.43	−1.20	0.01	0.23	3.08	6.13
	p_2	−7.26	4.93	−7.26	1.96	4.03	1.80	32.72	−122.1
	p_3			8994	5323	76.00	4.45	−5.60	0.74
	p_4							0.38	65.30

[a] Logarithmic-linear model (Equation (2.1)): $p_1 = \log k_w$, $p_2 = S$; [b] Bosch-Rosés's model (Equation (2.5)): $p_1 = q$, $p_2 = ps$; [c] Logarithmic-quadratic model (Equation (2.6)): $p_1 = (\log k)_0$, $p_2 = S$, $p_3 = T$; [d] Logarithmic-quadratic model with P_M^{N} transformation (Equation (2.7)): $p_1 = q$, $p_2 = S_1$, $p_3 = S_2$; [e] Neue-Kuss's model (Equation (2.8)): $p_1 = k_w$, $p_2 = c$, $p_3 = B$; [f] Jandera's model (Equation (2.9)): $p_1 = a$, $p_2 = b$, $p_3 = m$; [g] Mix model (Equation (2.11)): $p_1 = c_0$, $p_2 = c_1$, $p_3 = c_2$, $p_4 = c_4$; [h] Adsorption model (Equation (2.12)): $p_1 = c_0$, $p_2 = r$, $p_3 = c_2$, $p_4 = c_3$.

Table 2.5. Fitting quality statistics obtained for each retention model evaluated with the individual retention data of the alkylbenzenes. Experimental designs are described in Section 2.4.3.

Solute	Statistics[a]	Retention models							
		Linear	Bosch-Rosés	Quadratic	Quadratic- P_M^N	Neue-Kuss	Jandera	Mixed[b]	Adsorption[b]
Toluene	R_{adj}	0.99898	0.99990	0.99944	0.99996	0.99996	0.99996	0.99999	0.99999
	RE	2.28	0.71	2.28	0.26	0.18	0.29	0.20	0.22
	F	949	10082	633	49271	89028	30722	40628	30420
Ethylbenzene	R_{adj}	0.99870	0.99982	0.99928	0.99996	0.99996	0.99996	0.99999	0.99999
	RE	2.85	1.04	2.85	0.44	0.44	0.51	0.14	0.40
	F	883	6235	588	23552	23343	13041	102749	14250
Propylbenzene	R_{adj}	0.99876	0.99988	0.99924	0.99996	0.99996	0.99996	0.99999	0.99999
	RE	2.94	0.96	2.94	0.10	0.08	0.38	0.11	0.04
	F	1055	10020	704	389246	874772	41746	207573	2245279
Butylbenzene	R_{adj}	0.99828	0.99970	0.99888	0.99996	0.99996	0.99992	0.99999	0.99999
	RE	3.57	1.46	3.57	0.40	0.16	0.85	0.13	0.18
	F	862	4754	574	34657	194975	8322	143439	82657
Pentylbenzene	R_{adj}	0.99808	0.99962	0.99872	0.99996	0.99996	0.99988	0.99999	0.99999
	RE	3.94	1.77	3.94	0.64	0.26	1.13	0.08	0.38
	F	872	4124	582	17682	114060	6531	485171	29252
Hexylbenzene	R_{adj}	0.99836	0.99968	0.99884	0.99996	0.99996	0.99988	0.99999	0.99999
	RE	3.84	1.67	3.84	0.53	0.07	1.02	0.06	0.06
	F	1063	5529	708	37093	1390971	9728	936695	1140006

[a] n (number of experiments) = 5; R_{adj} = adjusted correlation coefficient (Equation (2.13)); RE = mean relative error (Equation (2.15)); F = Snedecor's F (Equation (2.16)). [b] For the mixed and adsorption models, the values of R_{adj} (see Equation (2.13)) could not be calculated due to the insufficient number of experiments. Instead, the values of R (Pearson correlation coefficient, Equation (2.14)) are shown in both cases.

Table 2.6. Fitting quality statistics obtained for each retention model evaluated with the individual retention data of the sulphonamides. Experimental designs are described in Section 2.4.3.

Solute	Statistics[a]	Retention models							
		Linear	Bosch-Rosés	Quadratic	Quadratic-P_M^N	Neue-Kuss	Jandera	Mixed	Adsorption
Sulphathiazole	R_{adj}	0.99782	0.99857	0.99910	0.99950	0.99850	0.99960	0.99990	0.99980
	RE	2.30	2.21	2.30	1.86	3.05	1.65	1.18	1.65
	F	163	248	122	220	76	276	339	193
Sulphaguanidine	R_{adj}	0.99676	0.99910	0.99805	0.99980	0.99980	0.99992	0.99995	0.99990
	RE	4.05	2.15	4.05	1.28	1.21	0.69	0.26	1.05
	F	536	1889	402	3880	3891	10570	63211	3526
Sulphisoxazole	R_{adj}	0.99720	0.99928	0.99837	0.99985	0.99982	0.99992	0.99995	0.99995
	RE	3.86	1.86	3.86	0.97	1.08	0.85	0.53	0.76
	F	576	2222	432	4586	4011	7788	8474	6666
Sulphapyridine	R_{adj}	0.99695	0.99920	0.99830	0.99985	0.99985	0.99995	0.99995	0.99995
	RE	3.87	1.94	3.87	0.97	0.90	0.54	0.18	0.82
	F	501	1871	375	4431	3991	16532	60724	4654
Sulphamethazine	R_{adj}	0.99680	0.99910	0.99810	0.99980	0.99980	0.99990	0.99995	0.99995
	RE	4.11	2.06	4.11	1.03	1.17	0.83	0.38	0.67
	F	528	1837	396	3579	3872	8141	21028	7741
Sulphadiazine	R_{adj}	0.99752	0.99943	0.99822	0.99990	0.99990	0.99995	0.99995	0.99995
	RE	3.99	1.84	3.99	0.82	0.93	0.60	0.20	0.92
	F	883	3818	662	11117	12581	29173	114330	8664
Sulphamethoxazole	R_{adj}	0.99850	0.99987	0.99895	0.99997	0.99997	0.99997	0.99995	0.99995
	RE	3.15	0.94	3.15	0.29	0.52	0.23	0.18	0.36
	F	1434	15149	1075	77568	38113	188757	189387	51060
Sulphamonomethoxine	R_{adj}	0.99823	0.99972	0.99870	0.99995	0.99992	0.99997	0.99995	0.99995
	RE	3.54	1.41	3.54	0.67	0.88	0.57	0.09	0.80
	F	1289	7739	966	26220	15037	37019	934653	12450

[a] n (number of experiments) = 6; R_{adj} = adjusted correlation coefficient (Equation (2.13)); RE = mean relative error (Equation (2.15)); F = Snedecor's F (Equation (2.16)).

The relative errors in predictions (*RE*) for conventional HPLC columns can reach values between 0.5 and 1.5%, typically obtaining relative errors between 1 and 3% [52,53]. The prediction errors obtained with the monolithic polymer column studied in this work were similar, even reaching smaller values (see also Figure 2.7). Thus, for both alkylbenzenes and sulphonamides, the logarithmic-linear (Equation (2.1)), Bosch-Rosés (Equation (2.5)), and logarithmic-quadratic (Equation (2.6)) models offered the lowest quality of fit, with relative errors between 1 and 4%. On the other hand, more complex models, such as the mixed model (Equation (2.11)) and the adsorption model (Equation (2.12)), do not offer a significant improvement in predictive performance, compared to other simpler models recognised as excellent, such as the Neue-Kuss equation (Equation (2.8)), which offers relative errors between 0.3 and 1.5%. In addition, these models require a larger number of experimental points to grant enough degrees of freedom, since they have four parameters. The Jandera's model (Equation (2.9)) stands out for its good performance, considering that it was initially proposed for NPLC, with relative errors between 0.2 and 1.2%.
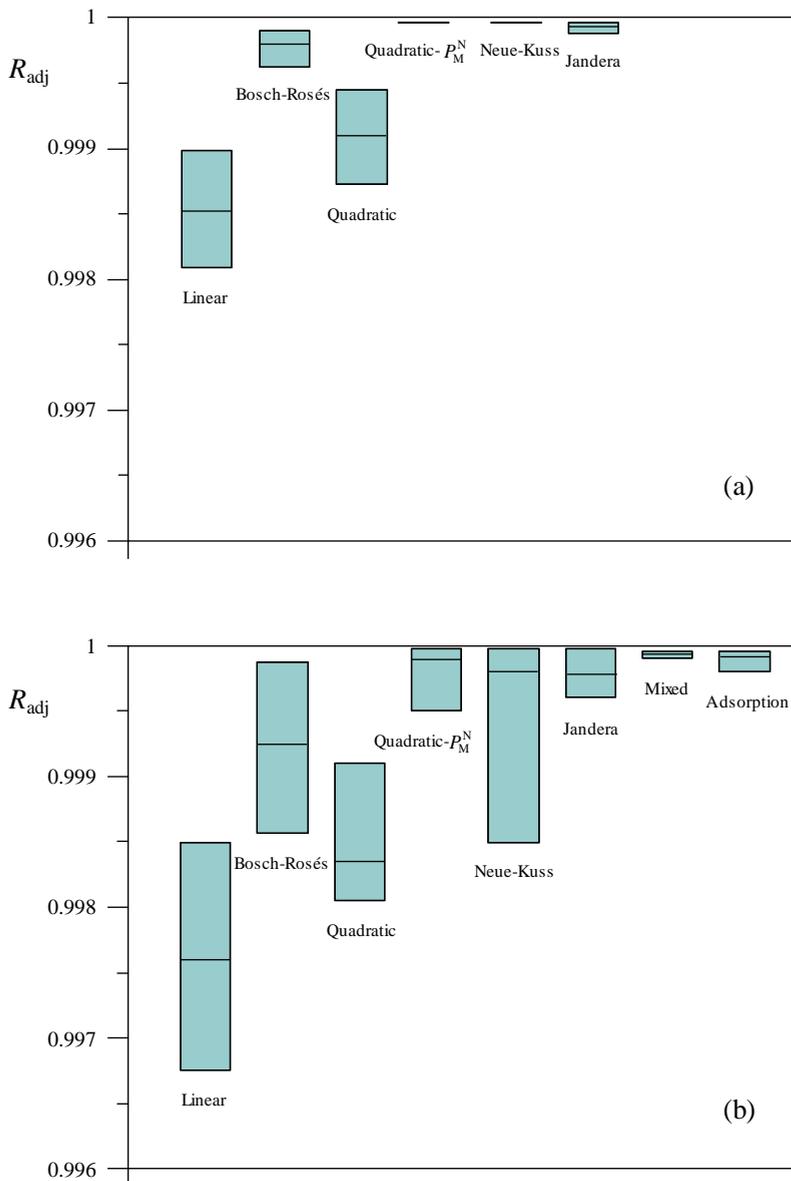
Figure 2.7. Boxes showing the individual values of the adjusted correlation coefficient (R_{adj}, Equation (2.13)) for the set of: (a) alkylbenzenes, and (b) sulphonamides, during the evaluation of the predictive quality of each assayed retention model. Median values are represented within each box.

Among all the models studied, the quadratic model with $P_{\mathrm{M}}^{\mathrm{N}}$ transformation (Equation (2.7)) was selected to describe the retention behaviour of the probe compounds analysed with the monolithic column, due to its simplicity and good predictive capability (errors between 0.3% and 1.3%). However, any of the models explored in this work can be used in the fittings, since all of them present a predictive performance of the same order as for conventional HPLC.

Figure 2.8 shows the correlation between the retention factor predicted using the selected model (Equation (2.7)), and those obtained experimentally for alkylbenzenes (Figure 2.8a, $n = 30$, $R^2 = 0.99992$, $RE = 0.86$ and $F = 1.88 \times 10^5$), and sulphonamides (Figure 2.8b, $n = 48$, $R^2 = 0.99979$, $RE = 0.54$ and $F = 1.12 \times 10^6$). As can be seen, the correlation obtained for both sets of solutes is nearly perfect, suggesting that the predictions of the chromatographic behaviour of polar and non-polar solutes separated with polymeric monolithic columns will be reliable.

2.5.4. Interactions of the probe compounds with the stationary phase

2.5.4.1. Selectivity

The retention behaviour allows drawing some conclusions about the separation mechanisms. As commented above, the c_4 coefficient in the mixed model (Equation (2.11), p_4 with the coding followed in Tables 2.3 and 2.4) indicates the relative magnitude of the partition and adsorption processes, so that if this coefficient is close to zero, the mechanism is based primarily on adsorption, and if it tends to infinity, the process is exclusively partitioning. As can be seen, this coefficient has small values, very close to zero for both alkylbenzenes and sulphonamides, so that the separation mechanism should be attributable to adsorption processes for both sets of solutes.
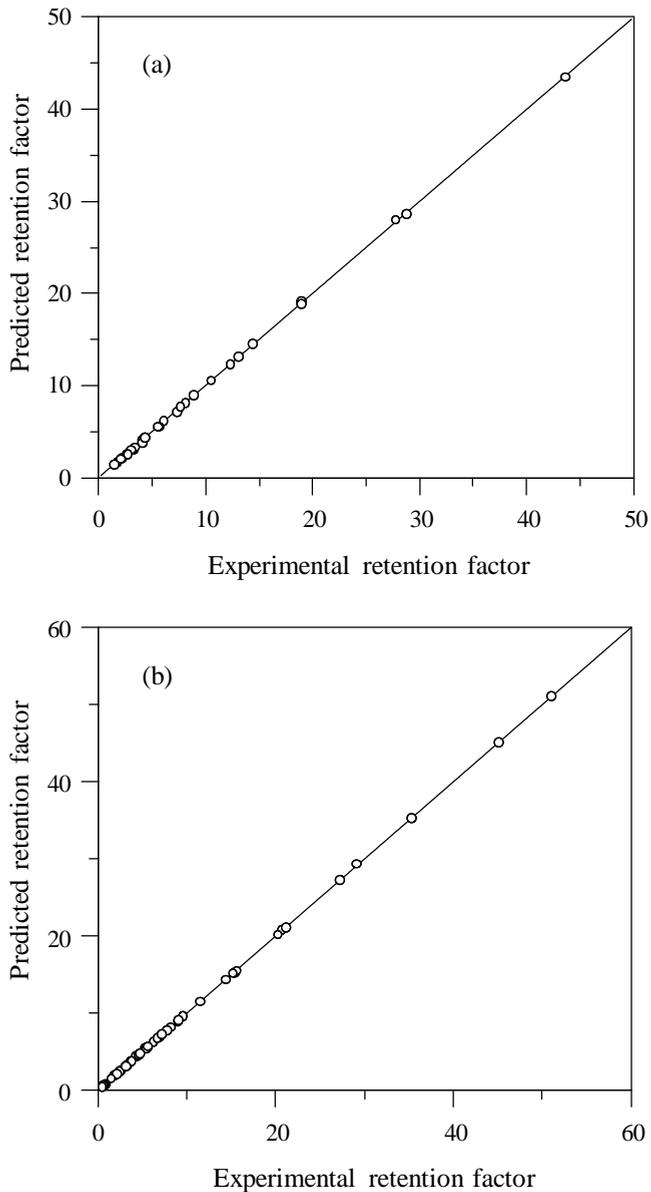
Figure 2.8. Correlation between predicted and experimental retention factor for: (a) alkylbenzenes, and (b) sulphonamides, obtained using the quadratic model with $P_{\mathrm{M}}^{\mathrm{N}}$ transformation (Equation (2.7)). The compounds were analysed with the poly(HMA-*co*-MAA-*co*-EDMA) capillary monolithic column.

On the other hand, structural similarities among the compounds can be investigated by examining the correlation between the parameters in the logarithmic-linear model (Equation (2.1)): S and log k_w [54–56]. The correlations are obtained for series of compounds belonging to particular families, such as alkylbenzenes, sulphonamides, amino acids and steroids. Solute retention in RPLC increases with the molecular size and hydrophobicity; therefore, S (which is positive) should be larger for later eluting solutes in this chromatographic mode. Considering polar solutes, structurally related compounds with similar dipolarity/polarizability and hydrogen bonding energy, will yield a linear relationship between S and log k_w. The linear regression coefficients of the correlations are, consequently, a measurement of the similarity of the interactions among the solutes in a set: the more similar, the larger the regression coefficient of the correlations.

In this work, instead of correlating the parameters in Equation (2.1), we preferred making the process by building correlations between the parameters in Equation (2.7), since we checked that the use of P_M^N, instead of φ, gives rise to better correlations between the regression parameters. Also, secondary interactions are isolated in the quadratic term:

$$S_1 = a + b\,q \tag{2.17}$$

$$S_2 = c + d\,q \tag{2.18}$$
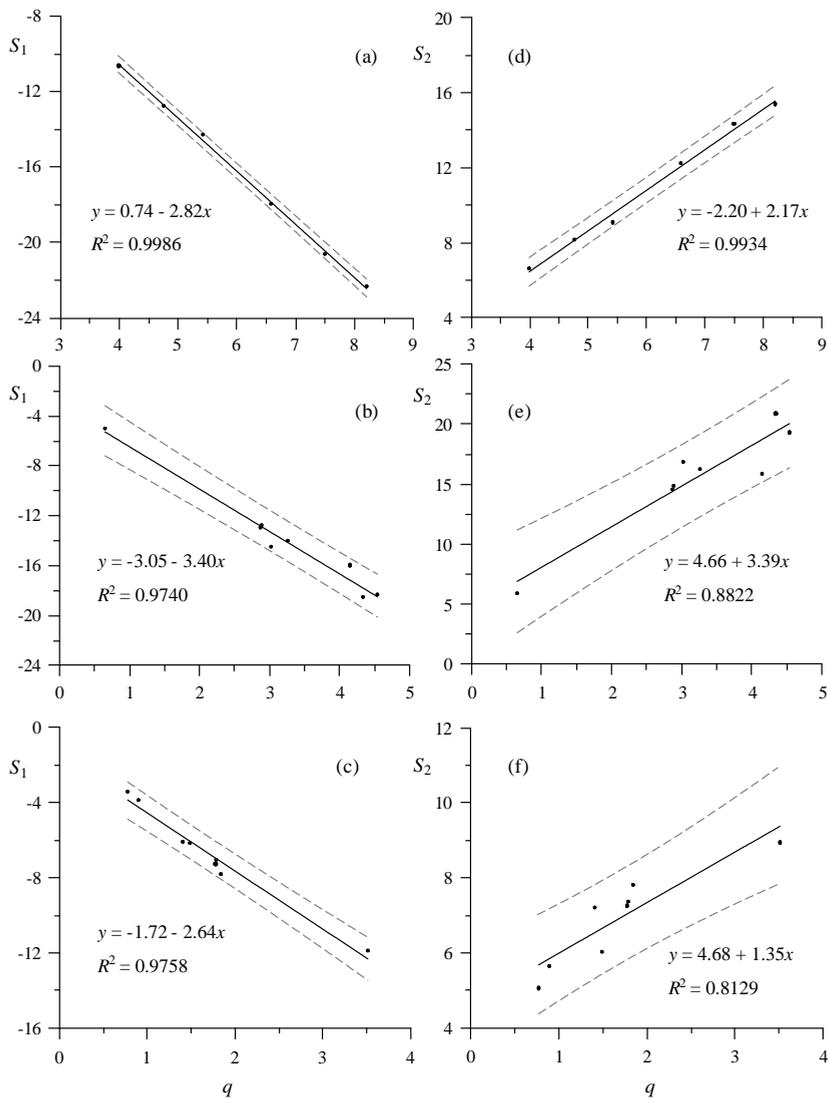
Figure 2.9. Correlation between the parameters of the logarithmic-quadratic model with P_M^N transformation (Equation (2.7)), for the mixture of alkylbenzenes (a,d), and sulphonamides (b,c,e,f), eluted from a 12 cm poly(HMA-*co*-MAA-*co*-EDMA) capillary monolithic column (a,b,d,e), and a conventional 9 cm C18 column (c,f). The 95% confidence intervals and regression straight-line are given.

Figure 2.9 compares the established correlations for S_1 and S_2 versus q, for the alkylbenzenes (Figures 2.9a and d) and sulphonamides (Figures 2.9b and e), analysed with the poly(HMA-*co*-MAA-*co*-EDMA) capillary monolithic column, and sulphonamides with the conventional C18 column (Figures 2.9c and f). We should remind that the chromatographic data for the six alkylbenzenes and eight sulphonamides, analysed with the poly(HMA-*co*-MAA-*co*-EDMA) column, were obtained with five and six isocratic mobile phases, respectively. The sulphonamides analysed with the C18 column were eluted with five mobile phases. In Figure 2.9, each dot corresponds to one compound and comes from a regression where several mobile phases were involved.

It can be observed that the data used to evaluate S_1 (Figures 2.9a to c) are, in general, less scattered than for S_2 (Figures 2.9d to f). The scattering observed in the plots denotes the variability in the molecular structure and its translation in terms of retention. Thus, the correlations achieved for alkylbenzenes have better quality than those for sulphonamides, using the poly(HMA-*co*-MAA-*co*-EDMA) monolithic column, and these are similar to those for sulphonamides with the C18 column. This indicates larger variability in the interactions of sulphonamides with the stationary phases.
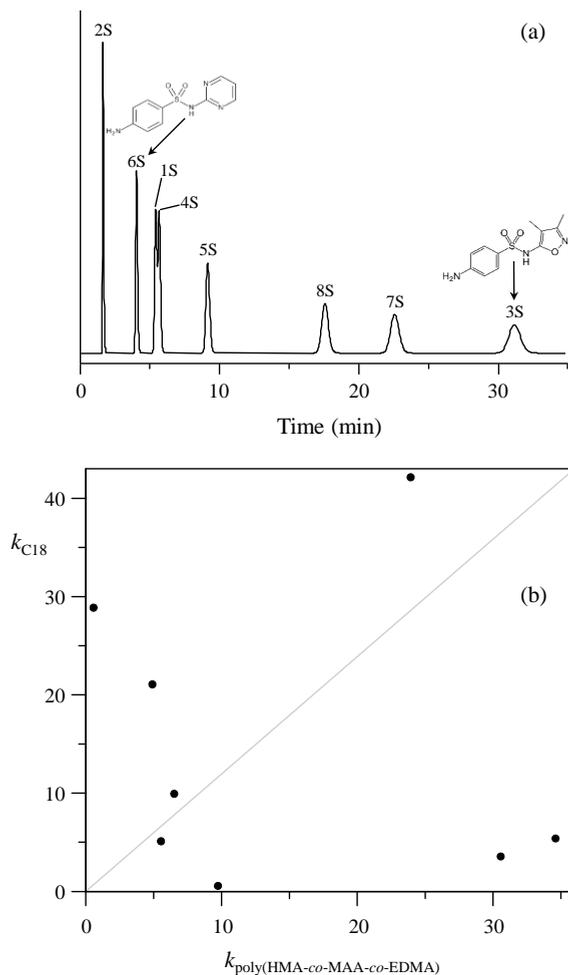
Figure 2.10. (a) Chromatogram of the mixture of eight sulphonamides, analysed with a conventional 9 cm C18 column, using 10% acetonitrile (*v/v*). The molecular structure for sulphadiazine (6S) and sulphisoxazole (3S), which experienced important changes in relative retention time with regard to the monolithic column (Figure 2.3c) are drawn. Other solute identities are given in Section 2.4.1. (b) Comparison of selectivity between the 9 cm C18 column and 12 cm poly(HMA-*co*-MAA-*co*-EDMA) capillary monolithic column, both eluted with 10% acetonitrile.

Here it is interesting to examine in more detail the interactions of sulphonamides with the monolithic and C18 columns. A chromatogram obtained for the set of sulphonamides with the C18 column, using the HPLC equipment and a flow rate of 1 mL/min at 25 °C is shown in Figure 2.10a. This chromatogram should be compared with the chromatogram in Figure 2.3c obtained for the same sulphonamides with the poly(HMA-*co*-MAA-*co*-EDMA) monolithic column. Attention should be drawn to the elution order for the eight sulphonamides in both columns, which indicates very different selectivity. To appraise better the change in selectivity, the retention factors with both columns were correlated. The results obtained for 10% acetonitrile are shown (Figure 2.10b), but similar selectivity mismatch was observed at other mobile phase compositions. The high scattering of the data indicates the presence of highly different interactions in both columns. As commented before by observing Figures. 2.5 and 2.6, the high scattering indicates that the hydrophobic interactions have a minor weight in explaining the retention behaviour of sulphonamides with the monolithic column.

2.5.4.2. Peak profiles

The profile of chromatographic peaks can be also described in a simple way using quadratic or linear models, based on the representation of the left (*A*) and right (*B*) peak half-widths versus the retention time, which are conveniently measured at 10% peak height [57,58]. These plots provide information about the changes in the values of the peak half-widths (or widths) and asymmetry as the solutes are eluted from the column. For convenience, the data can be fitted to the following equations:

$$A = m_A t_R + A_0 \tag{2.19}$$

$$B = m_B t_R + B_0 \tag{2.20}$$

m_A and m_B being the slopes of the linear correlations for the left and right peak half-widths, respectively, and A_0 and B_0 the extra-column contribution to the peak broadening. These parameters are obtained from the fitting of the half-widths for one or more compounds eluted at different retention times, using one or more mobile phase compositions. The sum of m_A and m_B represents the broadening rate of chromatographic peaks inside the column, and its ratio (m_B/m_A) indicates the peak asymmetry at high retention times. As we will comment below, these plots also give information about the interaction kinetics of solutes: peak broadening does not only happen at longer retention times, but also due to slower interaction kinetics.

Figure 2.11 depicts the half-width plots for the sets of alkylbenzenes and sulphonamides, eluted with acetonitrile, using the monolithic column, and for sulphonamides with the C18 column. For alkylbenzenes, analysed with the monolithic column (Figure 2.11a), rather good linear correlation was obtained when all available values of *A* and *B* from the whole set of compounds eluted with all mobile phases in the experimental design were plotted.
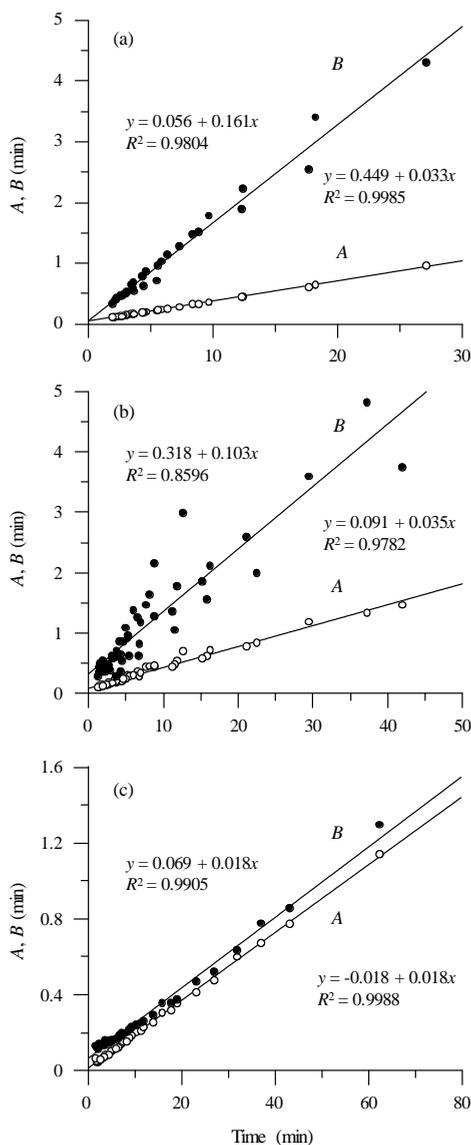
Figure 2.11. Global isocratic half-width plots for the set of: (a) alkylbenzenes, and (b) sulphonamides, analysed with the poly(HMA-*co*-MAA-*co*-EDMA) monolithic column; (c) sulphonamides analysed with the C18 column. All data in the experimental designs (see Section 2.4.3) were taken for the plots. The fitted straight-lines are overlaid. Left (*A*,○), and right (*B*,●) peak half-widths.
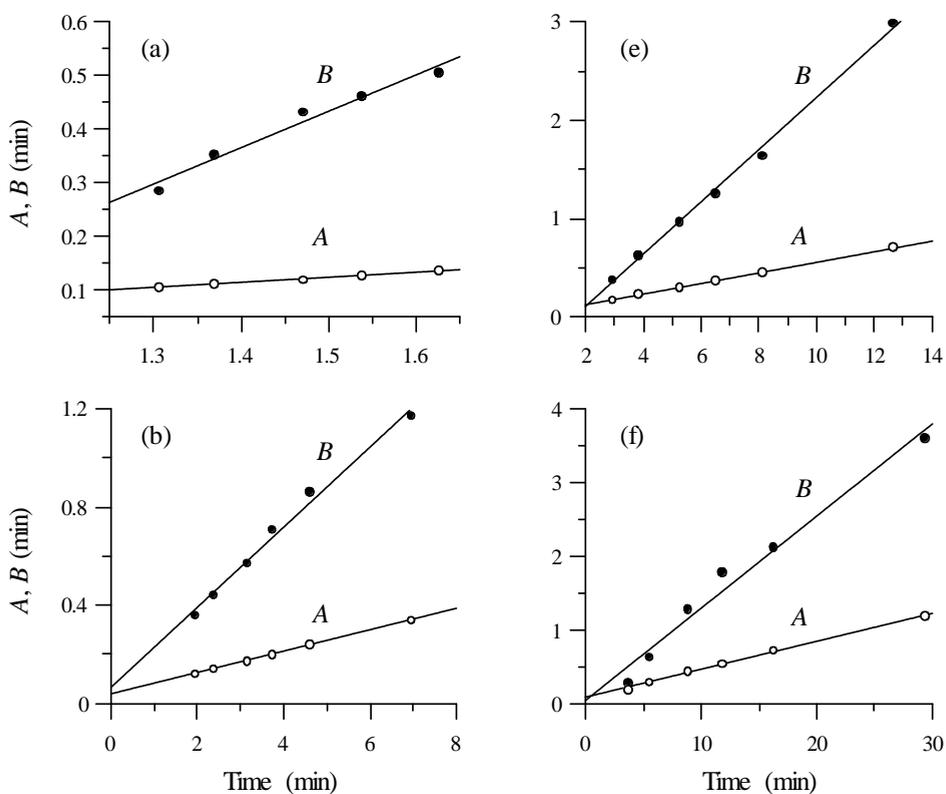
Figure 2.12. Individual isocratic half-width plots corresponding to the elution of each sulphonamide from the poly(HMA-*co*-MAA-*co*-EDMA) capillary monolithic column. Probe compounds (ordered according to the elution order): (a) sulphathiazole, (b) sulphaguanidine, (c) sulphisoxazole, (d) sulphapyridine, (e) sulphamethazine, (f) sulphadiazine, (g) sulphamethoxazole, and (h) sulphamonomethoxine. The fitted straight-lines are overlaid. Left (*A*,○), and right (*B*,●) half-widths. The slopes and determination coefficients for the fittings are given in Table 2.7.
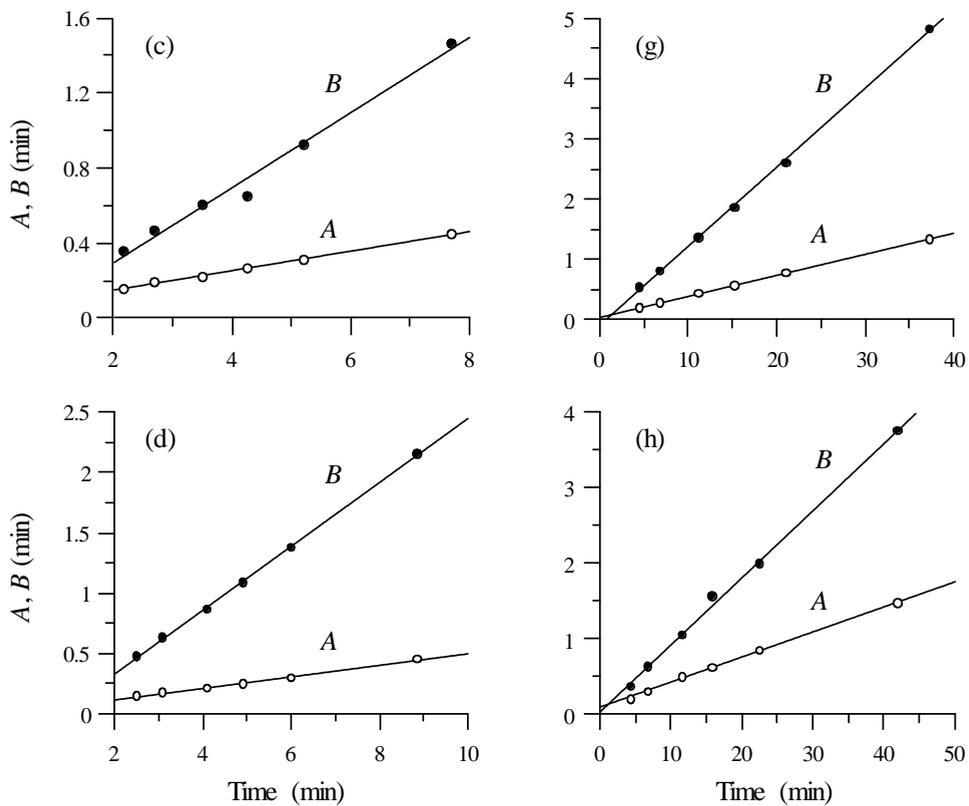
Figure 2.12 (continued).

For sulphonamides (Figure 2.11b), the B values showed high dispersion, making fitting of the data to a global model that considers all compounds and mobile phases unfeasible. This indicates the existence of different behaviours in the interaction kinetics for each sulphonamide with the monolithic column, due to the participation of different hydrophilic and hydrophobic forces in the retention of each compound, unlike alkylbenzenes that present a rather uniform behaviour, since the interactions are mainly hydrophobic. The different interaction behaviour for each sulphonamide with the monolithic column was confirmed by plotting the data for each sulphonamide, eluted with the whole range of assayed mobile phase compositions, and fitting them to individual models (Figure 2.12). This resulted in an enhanced fitting of the data, with regard to the global model (Figure 2.11b). In the case of alkylbenzenes, the individual models did not represent a significant improvement over the global model (see Figure 2.13). The half-width plots for the C18 column are depicted in Figure 2.11c for comparison purposes. As observed, the peaks are nearly Gaussian.
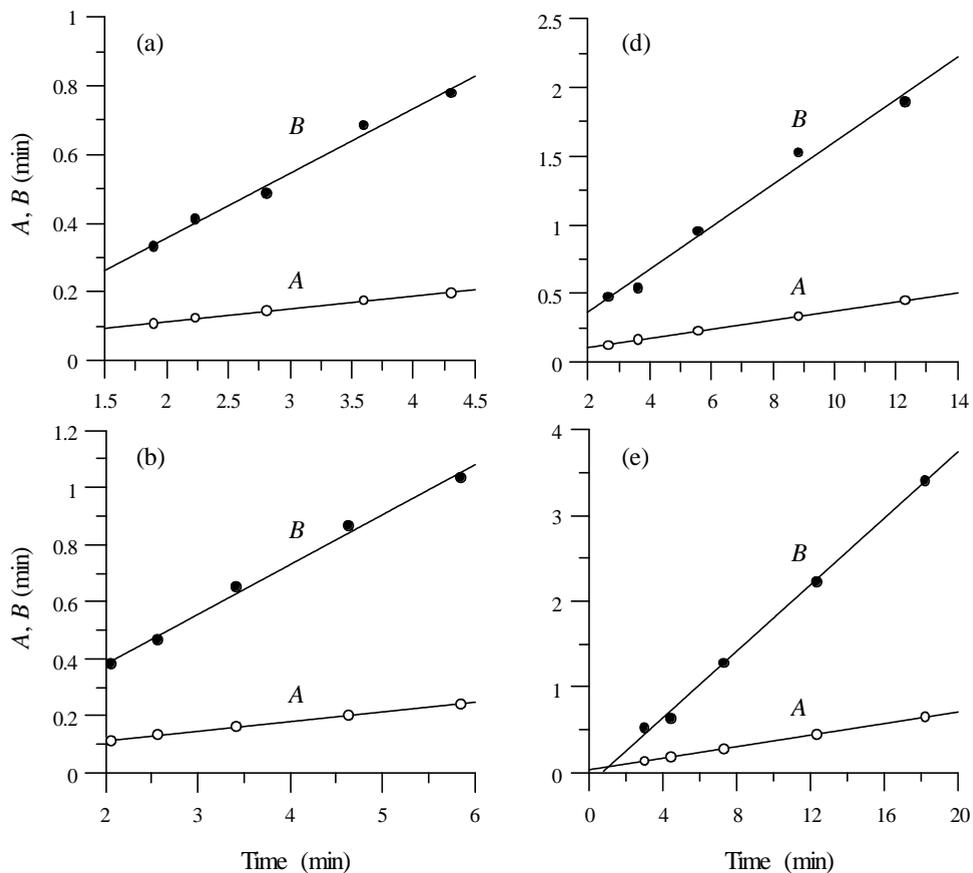
Figure 2.13. Individual isocratic half-width plots corresponding to the elution of each alkylbenzene with the poly(HMA-*co*-MAA-*co*-EDMA) capillary monolithic column: (a) toluene, (b) ethylbenzene, (c) propylbenzene, (d) butylbenzene, (e) pentylbenzene, and (f) hexylbenzene. The corresponding fitted straight-lines are overlaid. Left (A,○), and right (B,●) peak half-widths.
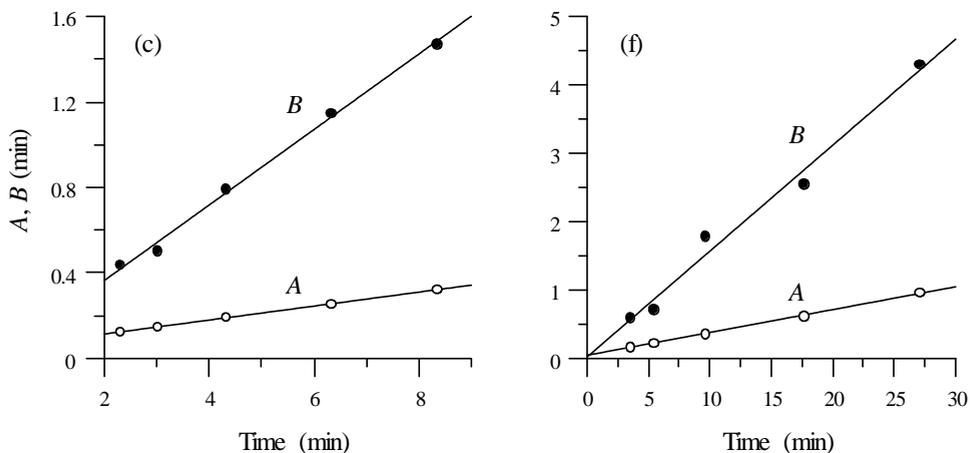
Figure 2.13 (continued).

The slopes of the linear segments for the left (m_A) and right (m_B) half-widths for the individual plots, and their sum ($m_A + m_B$) and ratio (m_B/m_A) for the assayed mobile phases are given in Table 2.7. It can be observed that the slope of the straight-line representing the right half-width (B) has significantly higher values than the left half-width (A), for both sets of compounds (Figures 2.11a and b, Figure 2.12, and Figure 2.13), with m_B/m_A ratios in the 4.57–5.31 range for alkylbenzenes, and 2.68–5.49 range for sulphonamides. This indicates that the solutes will present remarkably peak tailing in all the experimental conditions (see also Figures 2.3a and c).

Table 2.7. Slope values of the linear correlation for the left (m_A) and right (m_B) experimental half-widths for the set of alkylbenzenes and sulphonamides analysed with the monolithic column.

Solutes	m_A	R_A^2	m_B	R_B^2	$m_A + m_B$	m_B/m_A
Toluene	0.037	0.9972	0.189	0.9889	0.225	5.12
Ethylbenzene	0.033	0.9975	0.175	0.9936	0.208	5.26
Propylbenzene	0.033	0.9997	0.176	0.9960	0.209	5.31
Butylbenzene	0.033	0.9994	0.155	0.9859	0.189	4.65
Pentylbenzene	0.034	0.9998	0.194	0.9975	0.228	5.75
Hexylbenzene	0.034	0.9983	0.155	0.9844	0.189	4.57
Mean	0.034 ±0.002		0.174 ±0.016		0.208 ±0.017	5.11 ±0.44
Sulphathiazole	0.099	0.9919	0.681	0.9726	0.780	6.88
Sulphaguanidine	0.043	0.9996	0.163	0.9898	0.207	3.75
Sulphisoxazole	0.053	0.9983	0.200	0.9853	0.252	3.80
Sulphapyridine	0.048	0.9931	0.265	0.9994	0.313	5.49
Sulphamethazine	0.054	0.9995	0.266	0.9962	0.320	4.88
Sulphadiazine	0.038	0.9960	0.125	0.9767	0.164	3.27
Sulphamethoxazole	0.035	0.9999	0.131	0.9990	0.165	3.79
Sulphamonomethoxine	0.033	0.9988	0.089	0.9977	0.122	2.68
Mean	0.050 ±0.021		0.240 ±0.189		0.250 ±0.210	4.318 ±1.358

Note that the asymmetry is, in general, higher for alkylbenzenes (see Table 2.7), but the peaks of sulphonamides are also highly asymmetrical. This could be explained by the existence of microcavities in the connections between the equipment and the column, which causes that molecules in those chambers elute over time gradually (i.e., producing diffusion) [59]. In previous work, we observed the formation of asymmetrical peaks with conventional HPLC, using a coupled column system, where the columns were introduced in PEEK holders, which were screwed to maintain the columns sufficiently attached [60]. The deformed peaks were explained owing to the existence of column voids related to an insufficiently tight connection because of imperfect screwing of the PEEK holders, which created a small mixing chamber. In fact, a careless connection yielded extremely deformed peaks. In conventional systems, the formation of small cavities between the connections is not a great disadvantage, since the working scale is bigger, but in nano-HPLC, this can induce remarkable deformations of the peaks when working with such small volumes. We should here finally mention that peak tailing of prepared monolithic capillary columns might also originate from structural heterogeneity of polymer monolith and swelling of the stationary phase in mobile phases with various concentrations of acetonitrile.

In spite of the asymmetry achieved, it should be noted that the half-width models obtained in the study allow predicting the peak profiles at any acetonitrile composition, with values very similar to the experimental ones (compare Figures 2.3a and c with the predicted chromatograms in Figures 2.3b and d). The simulation of the peaks in the predicted chromatograms was carried out considering the prediction of peak profiles according to developments described elsewhere [56]. The similarity of the experimental and predicted

chromatograms is remarkable. Note that the simulations were carried out with normalised peaks.

2.6. Conclusions

In this work, the possibility of separating hydrophilic compounds with organic monolithic columns in miniaturised separation systems (in this case, in nano-HPLC) is studied. Three monolithic columns of diverse chemistry were prepared: (i) a column containing lauryl methacrylate (LMA), which confers a dominant hydrophobic character, (ii) a column of intermediate polarity formed with a mixture of hydrophobic (LMA) and ionisable (MAA) monomers, and (iii) a column with a more polar monomer (HMA) combined with MAA. All columns were characterised with alkylbenzenes (neutral and hydrophobic solutes) as probe compounds, which showed the columns had good permeability and good resolution for these compounds. However, the LMA column was not able to separate mixtures of hydrophilic compounds, such as sulphonamides. The use of a mixed column was found essential for the separation. The poly(LMA-*co*-MAA-*co*-EDMA) column was not able to completely resolve the mixture of eight sulphonamides due to its low efficiency and insufficient differentiation capability of the three more hydrophobic compounds. More favourable separation was achieved with the poly(HMA-*co*-MAA-*co*-EDMA) column.

The study of the poly(HMA-co-MAA-co-EDMA) column using retention models showed that it is comparable in terms of modelling performance to a conventional microparticulate alkyl-bonded column, with similar prediction errors in retention, although with reduced efficiency. The eight retention models tested showed good predictive capabilities, with the quadratic-logarithmic model with P_M^N transformation, the Jandera's model, and the mixed

and adsorption models providing the best performance. The mixed model indicated that the poly(HMA-*co*-MAA-*co*-EDMA) column operates according to an adsorption mechanism.

The use of several plots, based on the retention of the solutes and peak profiles, showed differences in the interactions of the solutes, among the different assayed columns. The comparison of the results obtained with the columns based on LMA and HMA, on the one hand, and a C18 column, on the other, indicated significant changes in selectivity, with remarkable reversals in the elution order.

The construction of plots of peak half-widths versus the retention time revealed large scattering when the data of all sulphonamides with all assayed mobile phases were represented for the poly(HMA-*co*-MAA-*co*-EDMA) column, but individual plots for each compound showed good performance. Meanwhile, the plots for alkylbenzenes altogether were similar to the individual representations for particular compounds. This behaviour reveals that each sulphonamide has a particular interaction with the monolithic column.

2.7. References

[1] F. Svec, J.M.J. Fréchet, Continuous rods of macroporous polymer as high-performance liquid chromatography separation media, *Anal. Chem.* 64 (1992) 820–822.

[2] F. Svec, T.B. Tennikova, Z. Deyl, *Monolithic Materials: Preparation Properties and Applications*, Journal of Chromatography Library, Elsevier, Amsterdam, 2003.

[3] J. Ou, Z. Liu, H. Wang, H. Lin, J. Dong, H. Zou, Recent development of hybrid organic-silica monolithic columns in CEC and capillary LC, *Electrophoresis* 36 (2015) 62–75.

[4] F. Svec, Y. Lv, Advances and recent trends in the field of monolithic columns for chromatography, *Anal. Chem.* 87 (2015) 250–273.

[5] M. Jonnada, Z. El Rassi, Robust naphthyl methacrylate monolithic column for high performance liquid chromatography of a wide range of solutes, *J. Chromatogr. A* 1409 (2015) 166–172.

[6] H. Zhang, L. Bai, Z. Wei, S. Liu, H. Liu, H. Yan, Fabrication of an ionic liquid-based macroporous polymer monolithic column via atom transfer radical polymerization for the separation of small molecules, *Talanta* 149 (2016) 62–68.

[7] R. Yu, W. Hu, G. Lin, Q. Xiao, J. Zheng, Z. Lin, One-pot synthesis of polymer monolithic column by combination of free radical polymerization and azide-alkyne cycloaddition "click" reaction and its application in capillary liquid chromatography, *RSC Adv.* 5 (2015) 9828–9836.

[8] H. Martínez Pérez-Cejuela, E.J. Carrasco Correa, A. Shahat, E.F. Simó Alfonso, J.M. Herrero Martínez, Incorporation of metal-organic frame-work amino-modified MIL-101 into glycidyl methacrylate monoliths for nano LC separation, *J. Sep. Sci.* 42 (2019) 834–842.

[9] E.J. Carrasco Correa, G. Ramis Ramos, J.M. Herrero Martínez, Hybrid methacrylate monolithic columns containing magnetic nanoparticles for capillary electrochromatography, *J. Chromatogr. A* 13 (2015) 77–84.

[10] E.J. Carrasco Correa, A. Martínez Vilata, J.M. Herrero Martínez, J.B. Parra, F. Maya, V. Cerdà, C.P. Cabello, G.T. Palomino, F. Svec, Incorporation of zeolitic imidazolate framework (ZIF-8)-derived nanoporous carbons in methacrylate polymeric monoliths for capillary electrochromatography, *Talanta* 164 (2017) 348–354.

[11] I. Nischang, O. Brüggemann, On the separation of small molecules by means of nano-liquid chromatography with methacrylate-based macroporous polymer monoliths, *J. Chromatogr. A* 1217 (2010) 5389–5397.

[12] J. Urban, F. Svec, J.M.J. Fréchet, Hypercrosslinking: New approach to porous polymer monolithic capillary columns with large surface area for the highly efficient separation of small molecules, *J. Chromatogr. A* 1217 (2010) 8212–8221.

[13] K. Liu, H.D. Tolley, M.L. Lee, Highly crosslinked polymeric monoliths for reversed-phase capillary liquid chromatography of small molecules, *J. Chromatogr. A* 1227 (2012) 96–104.

[14] S.L. Lin, Y.R. Wu, T.Y. Lin, M.R. Fuh, Preparation and evaluation of 1,6-hexanediol ethoxylate diacrylate-based alkyl methacrylate monolithic capillary column for separating small molecules, *J. Chromatogr. A* 1298 (2013) 35–43.

[15] E.C. Peters, M. Petro, F. Svec, J.M.J. Fréchet, Molded rigid polymer monoliths as separation media for capillary electrochromatography: 1. Fine control of porous properties and surface chemistry, *Anal. Chem.* 70 (1998) 2288–2295.

[16] E.C. Peters, M. Petro, F. Svec, J.M.J. Fréchet, Molded rigid polymer monoliths as separation media for capillary electrochromatography: 2. Effect of chromatographic conditions on the separation, *Anal. Chem.* 70 (1998) 2296–2302.

[17] F. Svec, Porous polymer monoliths: Amazingly wide variety of techniques enabling their preparation, *J. Chromatogr. A* 1217 (2010) 902–924.

[18] R.D. Arrua, M. Talebi, T.J. Causon, E.F. Hilder, Review of recent advances in the preparation of organic polymer monoliths for liquid chromatography of large molecules, *Anal. Chim. Acta* 738 (2012) 1–12.

[19] Y. Liang, L. Zhang, Y. Zhang, Recent advances in monolithic columns for protein and peptide separation by capillary liquid chromatography, *Anal. Bioanal. Chem*. 405 (2013) 2095–2106.

[20] A. Weller, E.J. Carrasco Correa, C. Belenguer Sapiña, A.R. Mauri Aucejo, P. Amorós, J.M. Herrero Martínez, Organo-silica hybrid capillary monolithic column with mesoporous silica particles for separation of small aromatic molecules, *Microchim. Acta* 184 (2017) 3799–3808.

[21] A. Namera, T. Saito, Advances in monolithic materials for sample preparation in drug and pharmaceutical analysis, *Trends Anal. Chem.* 45 (2013) 182–196.

[22] T. Nema, E.C.Y. Chan, P.C. Ho, Applications of monolithic materials for sample preparation, *J. Pharm. Biomed. Anal.* 87 (2014) 130–141.

[23] K.N. Smirnov, I.A. Dyantchkov, M.V. Telnov, A.V. Pirogov, O.A. Shpigun, Effect of monomer mixture composition on structure and chromatographic properties of poly(divinylbenzene-*co*-ethylvinyl-benzene-*co*-2-hydroxyethyl methacrylate) monolithic rod columns for separation of small molecules, *J. Chromatogr. A* 1218 (2011) 5010–5019.

[24] X. Wang, H. Lü, X. Lin, Z. Xie, Electrochromatographic characterization of methacrylate-based monolith with mixed mode of hydrophilic and weak electrostatic interactions by pressurized capillary electro-chromatography, *J. Chromatogr. A* 1190 (2008) 365–371.

[25] A. Svobodová, T. Knzek, J. Sirc, P. Salek, E. Tesarová, P. Coufal, K. Stulík, Monolithic columns based on a poly(styrene-divinylbenzene-methacrylic acid) copolymer for capillary liquid chromatography of small organic molecules, *J. Chromatogr. A* 1218 (2011) 1544–1547.

[26] M.L. Chen, L.M. Li, B.F. Yuan, Q. Ma, Y.Q. Feng, Preparation and characterization of methacrylate-based monolith for capillary hydrophilic interaction chromatography, *J. Chromatogr. A* 1230 (2012) 54–60.

[27] S.L. Lin, Y.R. Wu, T.Y. Lin, M.R. Fuh, Preparation and evaluation of poly(alkyl methacrylate-*co*-methacrylic acid-*co*-ethylene dimethacrylate) monolithic columns for separating polar small molecules by capillary liquid chromatography, *Anal. Chim. Acta* 871 (2015) 57–65.

[28] Y.J. Cheng, S.H. Huang, B. Singco, H.Y. Huang, Analyses of sulphonamide antibiotics in meat samples by on-line concentration capillary electrochromatography-mass spectrometry, *J. Chromatogr. A* 1218 (2011) 7640–7647.

[29] M.C. García Alvarez-Coque, J.R. Torres Lapasió, J.J. Baeza Baeza, Models and objective functions for the optimisation of selectivity in reversed-phase liquid chromatography, *Anal. Chim. Acta* 579 (2006) 125–145.

[30] R. Cela, E.Y. Ordoñez, J.B. Quintana, R. Rodil, Chemometric-assisted method development in RPLC, *J. Chromatogr. A* 1287 (2013) 2–22.

[31] J.R. Torres Lapasió, M.C. García Álvarez-Coque, Levels in the interpretive optimisation of selectivity in high-performance liquid chromatography: A magical mystery tour, *J. Chromatogr. A* 1120 (2006) 308–321.

[32] P. Jandera, M. Staňková, T. Hájek, New zwitterionic polymethacrylate monolithic columns for one- and two-dimensional microliquid chromatography, *J. Sep. Sci.* 36 (2013) 2430–2440.

[33] P. Jandera, T. Hájek, Z. Šromová, Mobile phase effects in reversed-phase and hydrophilic interaction liquid chromatography revisited, *J. Chromatogr. A* 1543 (2018) 48–57.

[34] A. Pappa-Louisi, P. Nikitas, Non-linear least-squares fitting with Microsoft Excel Solver and related routines in HPLC modelling of retention: I. Considerations of the problems of the method, *Chromatographia* 53 (2000) 477–486.

[35] A. Pappa-Louisi, P. Nikitas, Non-linear least-squares fitting with Excel Solver and related routines in HPLC modelling of retention: II. Considerations of selection of optimal fit, *Chromatographia* 53 (2000) 487–493.

[36] A. Pappa-Louisi, P. Nikitas, P. Balkatzopoulou, C. Malliakas, Two- and three-parameter equations for representation of retention data in reversed-phase liquid chromatography, *J. Chromatogr. A* 1033 (2004) 29–41.

[37] P. Nikitas, A. Pappa-Louisi, Retention models for isocratic and gradient elution in reversed-phase liquid chromatography, *J. Chromatogr. A* 1216 (2009) 1737–1755.

[38] L.R. Snyder, J.J. Kirkland, J.W. Dolan, *Introduction to Modern Liquid Chromatography*, 2nd ed., John Wiley & Sons, New York, 2011.

[39] E. Bosch, P. Bou, M. Rosés, Linear description of solute retention in reversed-phase liquid chromatography by a new mobile phase polarity parameter, *Anal. Chim. Acta* 299 (1994) 219–229.

[40] P.J. Schoenmakers, H.A.H. Billiet, L. de Galan, Description of solute retention over the full range of mobile phase compositions in reversed-phase liquid chromatography, *J. Chromatogr. A* 282 (1983) 107–121.

[41] U.D. Neue, H.J. Kuss, Improved reversed-phase gradient retention modeling, *J. Chromatogr. A* 1217 (2010) 3794–3803.

[42] P. Jandera, M. Janderová, J. Churáček, Gradient elution in liquid chromatography: VIII. Selection of the optimal composition of the mobile phase in liquid chromatography under isocratic conditions, *J. Chromatogr.A* 148 (1978) 79–97.

[43] P. Nikitas, A. Pappa-Louisi, P. Agrafiotou, Effect of the organic modifier concentration on the retention in reversed-phase liquid chromatography: II. Tests using various simplified models, *J. Chromatogr. A* 946 (2002) 33–45.

[44] M.J.D. Powell, An efficient method for finding the minimum of a function of several variables without calculating derivatives, *Computer J.* 7 (1964) 155–162.

[45] W.H. Press, B.P. Flannery, S.A. Teukolsky, *Numerical Recipes in FORTRAN 77: The Art of Scientific Computing*, Cambridge University Press, New York, USA, 1992.

[46] E. Peris García, M.T. Úbeda Torres, M.J. Ruiz Ángel, M.C. García Álvarez-Coque, Effect of sodium dodecyl sulphate and Brij-35 on the analysis of sulphonamides in physiological samples using direct injection and acetonitrile gradients, *Anal. Methods* 8 (2016) 3941–3952.

[47] J.R. Torres Lapasió, *MICHROM Software*, Marcel Dekker, New York, 2000.

[48] V. Bernabé Zafón, M. Beneito Cambra, E.F. Simó Alfonso, J.M. Herrero Martínez, Comparison on photo-initiators for the preparation of methacrylate monolithic columns for capillary electrochromatography, *J. Chromatogr. A* 1217 (2010) 3231–3237.

[49] M.J.D. Powell, *Unconstrained minimization algorithms without computation of derivatives*, Technical Report TP483, AERE Harwell Report, 1972.

[50] I.E. Frank, R. Todeschini, *The Data Analysis Handbook: Data Handling in Science and Technology*, Vol. 14, Elsevier Science, Amsterdam, 1994.

[51] S.N. Deming, Y. Michotte, D.L. Massart, L. Kaufman, B.G.M. Vandeginste, *Handbook of Chemometrics and Qualimetrics*, Part B, Elsevier Science, Amsterdam, 1998.

[52] J.R. Torres Lapasió, M. Rosés, E. Bosch, M.C. García Alvarez-Coque, Interpretive optimisation strategy applied to the isocratic separation of phenols by reversed-phase liquid chromatography with acetonitrile-water and methanol-water mobile phases, *J. Chromatogr. A* 886 (2000) 31–46.

[53] T. Álvarez Segura, C. Camacho Molinero, J.R. Torres Lapasió, M.C. García Alvarez-Coque, Analysis of amino acids using serially coupled columns, *J. Sep. Sci.* 40 (2017) 2741–2751.

[54] H.B. Xiao, X.M. Liang, R.C. Lu, Classification of structurally related compounds from Astragalus extract by correlation of the log k_w and S, *Chromatographia* 51 (2000) 212–220.

[55] C.F. Poole, *The Essence of Chromatography*, Elsevier, Amsterdam, 2003, pp. 303–304.

[56] J.A. Navarro Huerta, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Estimation of peak capacity based on peak simulation, *J. Chromatogr. A* 1574 (2018) 101–113.

[57] M.J. Ruiz Ángel, S. Carda Broch, M.C. García Álvarez-Coque, Peak half-width plots to study the effect of organic solvents on the peak performance of basic drugs in micellar liquid chromatography, *J. Chromatogr. A* 1217 (2010) 1786–1798.

[58] J.J. Baeza Baeza, M.J. Ruiz Ángel, S. Carda Broch, M.C. García Álvarez-Coque, Half-width plots, a simple tool to predict peak shape, reveal column kinetics and characterise chromatographic columns in liquid chromatography: State of the art and new results, *J. Chromatogr. A* 1314 (2013) 142–153.

[59] J.P. Grinias, B. Bunner, M. Gilar, J.W. Jorgenson, Measurement and modeling of extra-column effects due to injection and connections in capillary liquid chromatography, *Chromatography* 2 (2015) 669–690.

[60] C. Ortiz Bolsico, J.R. Torres Lapasió, M.J. Ruiz Ángel, M.C. García Álvarez-Coque, Comparison of two serially-coupled column systems and optimization software in isocratic liquid chromatography for resolving complex mixtures, *J. Chromatogr. A* 1281 (2013) 94–105.

CHAPTER 3

BENEFITS OF SOLVENT CONCENTRATION PULSES IN RETENTION TIME MODELLING OF LIQUID CHROMATOGRAPHY

3.1. Abstract

The advantages and disadvantages of the use of isocratic experimental designs including transient increments of organic solvent (i.e., pulses) in the mobile phase(s) of lowest elution strength are explored with modelling purposes. For retained solutes, this type of mixed design offers similar or better predictive capability than gradient designs, shorter measurement time than pure isocratic designs, and retention model parameters that agree with those derived from pure isocratic experiments, with similar uncertainties. The predicted retention times are comparable to those offered by models adjusted from pure isocratic designs, and the solvent waste is appreciably smaller. Under a practical standpoint, mixed designs including pulse(s) can be easily constructed by replacing the slowest isocratic runs with runs containing a pulse of short duration at an intermediate time. This allows the elution of the fastest solutes with appreciable retention in the initial sector of the elution program, previous to the pulse, and the elution of the slow solutes after the pulse, also in acceptable times. The fitting of the retention data obtained with pulses is simpler compared to gradient elution, and involves solving the integral equation of gradient elution, simplified by the presence of isocratic sectors. Experiments involving pulses reveal the existence of discrepancies in the predictions for solutes eluting in the nearby of the pulse, offered by the fundamental equation of gradient elution when this is solved using numerical integration. The correction of such discrepancies implies the inclusion of intra-column delays, in the arrival of changes in the concentration of organic modifier in the gradient to the instantaneous position of the solute, along the whole migration.

3.2. Introduction

Since the 70s, the use of liquid chromatography (LC) has been growing in analytical laboratories, due to its sensitivity, robustness, ease of use, and applicability to multiple problems in diverse fields (environmental, pharmaceutical, clinical and food analysis) [1,2]. Reversed-phase liquid chromatography (RPLC) is the most usual LC mode for non-volatile compounds, from small molecules to large biological macromolecules, in a wide range of polarities [3]. The most important step in the development of a chromatographic method is still the choice of column. At present, hundreds of columns are commercialised for RPLC, with very different performances. The choice of the organic modifier is instead particularly limited, being reduced almost exclusively to acetonitrile or methanol in mixtures with an aqueous buffer. Nevertheless, the chromatographic behaviour can be extensively modulated by varying the modifier concentration in the mobile phase. Since the initial conditions selected by the analyst rarely provide good enough resolution (except for very simple samples), an optimisation protocol must be applied to find out an appropriate isocratic mobile phase composition or gradient program [4,5].

Isocratic elution is suitable for samples containing a small group of analytes, within a small or moderate range of polarities. In this case, all solutes will be resolved in reasonable times with the proper mobile phase. In contrast, this elution mode is not recommended when solute polarities cover a wide range. Two situations are possible. The first would yield a chromatogram where the most retained solutes elute at appropriate times (at high modifier percentages), but early peaks will present poor resolution, or even be lost at the solvent front. The second situation is the opposite, giving rise to a chromatogram where the least retained solutes are well resolved (at low modifier percentages), but the

most retained solutes will elute at excessively long times, with broad peaks having sensitivity problems. Some highly retained solutes may be undetected (i.e., they will not be distinguished from the baseline), or even appear overlapped with the chromatogram of the next injected sample.

Therefore, it is not possible to improve both extremes of the chromatogram at the same time by using isocratic elution. This incompatibility is known as the "general problem of chromatographic elution" [2]. The usual solution is the application of a gradient of organic modifier, in which its concentration is gradually altered according to a program [6–9]. The main objective is to obtain adequate resolution for all sample components, by increasing the retention of the poorly retained solutes and reducing it for strongly retained ones. For this purpose, the elution strength of the mobile phase must initially be low and become stronger as the separation progresses (e.g., by increasing the percentage of organic modifier).

The interpretive optimisation of the resolution (i.e., based on models) allows finding the conditions that simultaneously separate all, or at least, the target compounds in the sample [4,5,10]. The first step in these optimisations consists in the collection of information about the chromatographic behaviour of the solutes in the sample, focusing mainly on retention and covering wide regions of the involved factors. With this aim, the data are collected under controlled elution conditions, according to a pre-established experimental design. For each solute, a mathematical model, adequately describing the chromatographic behaviour as a function of the experimental factors, is fitted. The models allow the prediction of retention times and other peak properties, for particular solutes and under different conditions, in isocratic or gradient modes [4].

In this work, the use of transient increases (pulses) in organic solvent concentration in LC is reported. The benefits and drawbacks of mixed designs

(including isocratic and/or gradient experiments with transient increases) are examined, with the main aim of obtaining more informative experimental designs. The pulses allow obtaining retention information from highly hydrophobic solutes, maintaining low modifier concentration during most of the elution. This type of elution program reduces the retention time of the most retained compounds to reasonable values. Moreover, it constitutes an interesting possibility in isocratic experimental designs to have access to measurements at low elution strength, which grants enriched information about the retention behaviour of slow compounds. In contrast to isocratic designs with pulses, low organic solvent concentrations have a marginal participation in designs constituted by several gradient ramps. Thus, the transient presence of a high concentration of modifier during a short time gives rise to measurable retention times, where the lowest concentrations of modifier in the pulse still have a significant weight.

3.3. Theory

3.3.1. Retention models

In RPLC, the experimental factor usually optimised is the organic modifier content in the mobile phase. It not only has a large impact on the elution strength and selectivity, but it can also be easily altered in wide ranges to modulate the retention of a large variety of compounds. The literature has provided a wide variety of models, useful to describe the retention behaviour [4,11–14], which allow the prediction of the retention factor (k), as a function of the volumetric fraction of modifier, φ. In this work, we have considered the models described below (the fitting parameters adopt particular values for each solute, column and modifier).

(i) *Logarithmic-linear model*

This model was proposed by Snyder et al. [2], and can be expressed as:

$$\ln k = c_0 + c_1 \varphi = \ln k_{\mathrm{w}} - S \varphi \qquad (3.1)$$

where

$$k = \frac{t_{\mathrm{R}} - t_0}{t_0 - t_{\mathrm{ext}}} \qquad (3.2)$$

being t_{ext} the extra-column time. Very often, t_{ext} is neglected in the calculation of k. In Equation (3.1), the intercept of the fitted straight-line, $\ln k_{\mathrm{w}}$, refers to the extrapolated value of a mobile phase composed of pure water. The slope S indicates the sensitivity of retention to changes in the organic modifier content, being a measurement of the elution strength of the mobile phase.

(ii) *Logarithmic-quadratic model*

Equation (3.1) accurately describes the retention in RPLC only in narrow or moderate organic modifier ranges. Large deviations from linearity are found at extreme high and low modifier concentrations. In this case, the logarithmic-quadratic model proposed by Schoenmakers [15] can be used instead, providing accurate fittings:

$$\ln k = c_0 + c_1 \varphi + c_2 \varphi^2 = \ln k_{\mathrm{w}} - S \varphi + T \varphi^2 \qquad (3.3)$$

(iii) *Neue-Kuss model*

The model proposed by Neue and Kuss [16] has demonstrated excellent performance in isocratic elution, in wide domains of organic modifier, being

one of the most accurate models for predicting the retention in RPLC, currently available:

$$k = k_w (1 + c\,\varphi)^2 e^{-\frac{B\,\varphi}{1 + c\,\varphi}} \tag{3.4}$$

where k_w is the retention factor extrapolated to a phase constituted only by water, like in Equation (3.1), c is a curvature parameter, and B a measurement of the elution strength.

3.3.2. Prediction of retention times in gradient elution

The retention times can be obtained by solving the so-called fundamental equation of gradient elution, expressed as [8,17–19]:

$$t_0 - t_{ext} = \int_0^{t_D} \frac{dt}{k_0} + \int_{t_D}^{t_g - t_0} \frac{dt}{k(\varphi(t))} \tag{3.5}$$

where t_0 is the dead time, k_0 the retention factor at the start of the gradient, t_g the retention time of a given solute in the assayed gradient conditions, and $k(\varphi(t))$ is an expression that describes the retention based on the gradient program. Tubing with appreciable volume between mixer and column implies the introduction of a certain delay in the arrival of the changes of composition programmed in the gradient, at the column inlet. The parameter t_D (dwell time) quantifies this delay. The real gradient profile must be obtained by adding t_D to the programmed time values.

The $k(\varphi(t))$ function implies two nested equations: the dependence of the modifier concentration with time (that is, the gradient program), and the retention factor as a function of the modifier (the retention model). The retention time can be calculated for any gradient, as long as the $k(\varphi(t))$ function

is known, but the solution can be obtained analytically only in limited cases. If the gradient only implies linear changes between ln k and φ (Equation (3.1)), and between the gradient program and time t, Equation (3.5) will have the following analytical solution:

$$t_g = t_0 + t_D + \frac{1}{S\,m} \ln \left[1 + S\,m\,(t_0\,k_w\,e^{-S\varphi_0} - t_D) \right] \tag{3.6}$$

In the development of Equation (3.6), for convenience, the gradient program has been shifted to compensate t_D:

$$\varphi = \varphi_0 + m\,(t - t_D) \tag{3.7}$$

where m is the gradient slope, φ_0 the initial gradient concentration, and S and k_w are the solute model parameters in Equation (3.1).

The analytical integration of the Neue-Kuss model (Equation (3.4)) gives rise to:

$$t_g = \frac{\ln H\,[1 + c(a - mt_0)] - B(a - mt_0)}{m\,(B - c\ln H)} \tag{3.8}$$

where

$$H = m\,B\ln k_w \left(t_0 - \frac{t_D}{k_w} \right) + e^{\frac{B(a + mt_D)}{1 + c(a + mt_D)}} \tag{3.9}$$

k_w, c and B are the solute model parameters, m is again the gradient slope, and a is the modifier concentration at the start of the ramp ($t = t_D$). Equations. (3.6) and (3.8)/(3.9) are valid for solutes eluting along a single gradient ramp.

The combination of the gradient function and the retention model often leads to expressions in Equation (3.5) lacking of simple analytical solution. In such cases, the retention time in gradient elution, t_g, can be obtained through

numerical integration, by dividing the integral in infinitesimal steps [19]. The last upper limit whose sum equals or exceeds t_0 gives the solution:

$$t_0 - t_{\text{ext}} = \int_0^{t_D} \frac{dt}{k_0} + \int_{t_D}^{t_g - t_0} \frac{dt}{k(\varphi(t))} = \frac{t_D}{k_0} + \int_{t_D}^{t_1} \frac{dt}{k(\varphi(t))}$$
$$+ \int_{t_1}^{t_2} \frac{dt}{k(\varphi(t))} + ... + \int_{t_{i-1}}^{t_i} \frac{dt}{k(\varphi(t))} + \int_{t_i}^{t_{i+1} - t_0} \frac{dt}{k(\varphi(t))}$$

(3.10)

If the steps in which the integral is divided are sufficiently small, it can be assumed that $k(t)$ will be constant in each infinitesimal step:

$$t_0 - t_{\text{ext}} \approx \frac{t_D}{k_0} + \frac{t_1 - t_D}{k(\varphi(t_1))} + \frac{t_2 - t_1}{k(\varphi(t_2))} + ... + \frac{t_i - t_{i-1}}{k(\varphi(t_i))} + \frac{t_{i+1} - t_i}{k(\varphi(t_{i+1}))}$$

(3.11)

where $k \approx k_i \approx k_{i+1}$ if the integration step $(t_{i+1} - t_i)$ is small enough.

Depending on the availability of primitive function in the integral and the accuracy level required for the solution, the analyst must make a decision about the use of analytical or numerical integration. For this study, we have chosen numerical integration, which is competitive whenever the required accuracy is not much smaller than 0.0001 min. For calculations such as gradient optimisations or fitting of gradient data for modelling, numerical integration is a valid option, in spite of the reduced speed.

3.3.3. Correction of the deviations in retention in gradient numerical integration associated to time delays

In gradient elution, the programmed changes in the eluent composition reach the solute location with an increasing delay along its migration, where several independent contributions can be distinguished. The first contribution is related to the dwell volume, associated to void spaces between mixer and column inlet.

If there were several columns in tandem, then the dead volume(s) of the column(s) inserted before the one where the solute is migrating will introduce a second delay. A third delay is associated to the time needed by the solvent front to reach the solute location from the column inlet (intra-column delay) [18–20].

In contrast to the other delays, which can be easily incorporated in the solution of the fundamental equation for gradient elution (Equations. (3.10) and (3.11)), the intra-column delay is difficult to implement when numerical integration is carried out. The reason is that it requires monitoring the actual solute position within the column along the gradient program, which implies evaluating the magnitude of the integral along the multiple successive infinitesimal steps. This can be corrected by modifying Equation (3.11) as follows:

$$
\begin{aligned}
t_0 - t_{\text{ext}} = \frac{t_{\text{D}}}{k_0} &+ \frac{t_1 - t_{\text{D}}}{k(\varphi(t_1 - \tau_1))} + \frac{t_2 - t_1}{k(\varphi(t_2 - \tau_2))} + ... \\
&+ \frac{t_i - t_{i-1}}{k(\varphi(t_i - \tau_i))} + \frac{t_{i+1} - t_i}{k(\varphi(t_{i+1} - \tau_{i+1}))}
\end{aligned}
\tag{3.12}
$$

where each of the terms τ_i are the cumulative sum, up to the term i:

$$
\tau_i = \frac{t_{\text{D}}}{k_0} + \sum_{j=1}^{i} \frac{t_{j+1} - t_j}{k(\varphi(t_j - \tau_j))}
\tag{3.13}
$$

The gradient retention time (t_{g}) will be then $t_i + t_0$, being t_i the time along the gradient in Equation (3.12) where the summation matches $t_0 - t_{\text{ext}}$. The intra-column correction is solute-dependent and requires being implemented in iterations in the numerical integration. The reason is that the calculated solute position along the gradient program happens earlier to what was calculated without considering the delay. This implies that a new delay should be

calculated with the corrected position. The successive corrections tend to zero, the process being quickly convergent.

In the case of analytical integration, the intra-column delay is intrinsically considered, whenever the gradient consists of a single linear ramp. In multi-linear gradients, however, the lower limits in the integral term associated to each linear segment must be corrected with the column fraction migrated at the start of each segment.

3.3.4. Fitting of the retention model and retention time predictions for non-isocratic experiments

If the training set includes exclusively isocratic data, the retention model can be straightforwardly fitted, using either linear or non-linear procedures, depending on the retention model (see Equations. (3.1), (3.3) and (3.4)). When the source data are gradient retention times, the calculation of the model parameters for a given solute is more complex. The simplest situation corresponds to retention models for which the fundamental equation (Equation (3.5)) has analytical solution, and t_g can be worked out as is the case of Equations. (3.1) and (3.4), using a linear gradient (Equations. (3.6) and (3.8)/(3.9)). In such cases, the solution includes the parameters defining each gradient in the training set (the slope and intercept, e.g., m and φ_0), the system dwell time, the dead and extra-column times, and the solute parameters according to the model (e.g., k_w and S in Equation (3.1)). In order to get the solute model parameters, non-linear fitting is required, where the agreement between predicted and experimental t_g values is monitored as the model parameters are adjusted by an algorithm. In this study, the tuning process of the model parameters was controlled by the Powell algorithm [21] (see Chapter 2 for method description). Highly accurate evaluations (until reaching the

machine internal precision) required typically 50 iterations, with experimental designs constituted of five elution programs.

3.4. Experimental

3.4.1. Reagents

For this study, the following 14 sulphonamides were considered: (1) sulphaguanidine, (2) sulphanilamide, (3) sulphadiazine, (4) sulphathiazole, (5) sulphapyridine, (6) sulphamerazine, (7) sulphamethazine, (8) sulpha-methizole, (9) sulphamonomethoxine, (10) sulphachloropyridazine, (11) sulphamethoxazole, (12) sulphisoxazole, (13) sulphadimethoxine, and (14) sulphaquinoxaline (Sigma, Roedermark, Germany). Stock solutions of these compounds containing 100 µg/mL were prepared with nanopure water (obtained with a purification system of Adrona B30 Trace, Burladingen, Germany), assisted with an ultrasonic bath (from Elmasonic, Singen, Germany). Adequate volumes of the stock solutions were mixed in order to get similar peak areas for all sulphonamides.

Chromatographic runs were carried out in both isocratic and gradient modes, using mobile phases prepared with HPLC-grade acetonitrile (Scharlau, Barcelona, Spain) and anhydrous sodium dihydrogen phosphate (Fluka, Germany), until reaching 0.01 M solutions with nanopure water. The pH was fixed at 3.0 by addition of 0.1 M HCl and NaOH (Scharlau). The training set consisted of five isocratic experiments at 10, 13, 16, 20 and 25% (v/v) acetonitrile. Duplicated injections were carried out.

All solutions were filtered through 0.45 µm Nylon membranes from Micron Separations (Westboro, MA, USA), before injection into the chromatographic system.

3.4.2. Apparatus and column

The analysis was performed with an HP1100 chromatograph (Agilent, Waldbronn, Germany), composed of the following modules: quaternary pump, autosampler equipped with 2 mL vials, thermostated column compartment, and UV-Vis detector set at 254 nm. The injection volume was 20 μL, and the mobile phase flow rate was kept constant at 1.0 mL/min. A Zorbax Eclipse XDB-C18 column (150×4.6 mm) with a particle size of 5 μm (Agilent) was used in the analyses. All injections were carried out under controlled temperature conditions at 25 ºC. The dead time was determined for different mobile phase compositions by injection of KBr (from Acros Organics, Fair Lawn, NJ, USA). The dwell time (1.16 min) was measured using an acetone gradient. The extra-column time was 0.12 min.

A pH-meter (model MicropH 2002, Crison, Barcelona) and a glass membrane electrode containing a Ag/AgCl reference electrode with 3.0 M KCl solution as salt bridge (model 8102, Orion, Barcelona) were used to measure the pH.

3.4.3. Software

For the acquisition of signals, an OpenLAB CDS LC workstation (Agilent, revision B.04.03) was used. The peak properties (retention time and half-widths) were measured with the MICHROM software [22]. Data treatment was carried out with home built-in functions written in Matlab 2016b (The MathWorks Inc., Natick, MA, USA).

3.5. Results and discussion

3.5.1. Selection of the retention model

To evaluate the modelling quality of Equations. (3.1), (3.3) and (3.4), the following statistics were calculated [23,24]:

(i) *Mean relative error*

$$RE = \frac{\sum\limits_{i=1}^{ne} \left| \hat{k}_i - k_{\exp,i} \right|}{\sum\limits_{i=1}^{ne} \hat{k}_{\exp,i}} \times 100 \qquad (3.14)$$

(ii) *Correlation coefficient*

$$R = \sqrt{1 - \frac{\sum\limits_{i=1}^{ne} (\hat{k}_i - k_{\exp,i})^2}{\sum\limits_{i=1}^{ne} (k_{\exp,i} - \bar{k}_{\exp,i})^2}} \qquad (3.15)$$

(iii) *Adjusted correlation coefficient*

$$R_{\mathrm{adj}} = \sqrt{1 - (1 - R^2) \times \frac{ne - 1}{ne - np - 1}} \qquad (3.16)$$

(iv) *Snedecor's F*

$$F = \frac{\sum\limits_{i=1}^{ne} (\hat{k}_i - \hat{k}_{\mathrm{mean}})^2}{\dfrac{\sum\limits_{i=1}^{ne} (k_{\exp,i} - \hat{k}_i)^2}{ne - np}} \qquad (3.17)$$

In the equations above, *ne* is the number of experiments (i.e. mobile phases) in the experimental design, *np* is the number of model parameters, $k_{\exp,i}$ and $\hat{k}_i$ are the experimental and predicted retention factors for experiment *i*, and $\hat{k}_{\mathrm{mean}}$ and $\bar{k}_{\exp}$, the predicted and mean experimental retention factors, respectively. It should be noted that despite being used frequently, the correlation coefficient is not appropriate if the models being compared involve a different number of parameters.

Table 3.1 shows the performance of the retention models. Among them, the Snyder model (Equation (3.1), with two parameters) offered the poorest predictions, with relative errors of about 4–5%. The Schoenmakers (Equation (3.3)) and Neue-Kuss (Equation (3.4)) models contain three parameters. Their performance was excellent, with prediction errors usually in the 0.3–0.5% range, with almost identical predictive capability. The Neue-Kuss model was finally selected due to its good behaviour in extrapolations, the low uncertainty in the estimation of the model parameters, and because it allows a relatively simple analytical solution of the fundamental equation (Equations. (3.8) and (3.9)). The simulation of chromatograms was carried out considering also the prediction of peak profiles, according to developments described elsewhere [25].

Table 3.1. Fitting statistics corresponding to three retention models for the set of 14 sulphonamides, considering an experimental design with five isocratic mobile phases (10, 13, 16, 20 and 25% acetonitrile *v/v*). For sulphadimethoxine and sulphaquinoxaline, whose retention at 10% acetonitrile exceeded 2.5 hours, the design included only the other four mobile phases.

Solute	Statistics	Retention model		
		Eq. (3.1)	Eq. (3.3)	Eq. (3.4)
Sulphaguanidine	RE	5.15	0.24	0.67
	R	0.99719	0.99999	0.99999
	R_{adj}	0.99437	0.99999	0.99994
	F Snedecor	78.2	16953.3	2345.1
Sulphanilamide	RE	3.96	0.45	0.14
	R	0.99821	0.99999	0.99999
	R_{adj}	0.99641	0.99998	1.00000
	F Snedecor	85.4	3976.7	38684.4
Sulphadiazine	RE	4.89	0.46	0.24
	R	0.99769	0.99999	0.99999
	R_{adj}	0.99538	0.99998	0.99999
	F Snedecor	204.0	13348.3	43689.3
Sulphathiazole	RE	5.87	0.46	0.22
	R	0.99745	0.99999	0.99999
	R_{adj}	0.99489	0.99998	0.99999
	F Snedecor	264.5	21293.8	89193.9
Sulphapyridine	RE	5.34	0.47	0.56
	R	0.99769	0.99999	0.99999
	R_{adj}	0.99538	0.99997	0.99996
	F Snedecor	266.9	17500.4	11159.9
Sulphamerazine	RE	5.15	0.61	0.67
	R	0.99776	0.99999	0.99998
	R_{adj}	0.99553	0.99996	0.99994
	F Snedecor	260.7	9923.0	7339.8
Sulphamethazine	RE	5.48	0.58	0.69
	R	0.99782	0.99999	0.99998
	R_{adj}	0.99564	0.99996	0.99994
	F Snedecor	318.3	14451.6	9226.3

Table 3.1 (continued).

Solute	Statistics	Retention model		
		Eq. (3.1)	Eq. (3.3)	Eq. (3.4)
Sulphamethizole	RE	5.43	0.38	0.31
	R	0.99808	0.99999	0.99999
	R_{adj}	0.99615	0.99998	0.99999
	F Snedecor	403.9	38320.1	57137.0
Sulphamonomethoxine	RE	5.07	0.58	0.68
	R	0.99843	0.99999	0.99999
	R_{adj}	0.99686	0.99997	0.99995
	F Snedecor	518.9	20065.7	13698.5
Sulphachloropyridazine	RE	4.66	0.37	0.30
	R	0.99852	0.99999	0.99999
	R_{adj}	0.99704	0.99998	0.99999
	F Snedecor	486.5	33863.1	54464.7
Sulphamethoxazole	RE	4.38	0.49	0.48
	R	0.99873	0.99999	0.99999
	R_{adj}	0.99746	0.99997	0.99997
	F Snedecor	576.8	21244.6	20922.2
Sulphisoxazole	RE	4.29	0.41	0.45
	R	0.99887	0.99999	0.99999
	R_{adj}	0.99774	0.99998	0.99998
	F Snedecor	709.1	39899.2	30554.2
Sulphadimetoxine	RE	4.00	0.92	0.64
	R	0.99927	0.99999	0.99999
	R_{adj}	0.99781	–	–
	F Snedecor	411.6	3800.4	7706.0
Sulphaquinoxaline	RE	3.51	0.64	0.42
	R	0.99946	0.99999	0.99999
	R_{adj}	0.99838	–	–
	F Snedecor	595.8	8723.1	19996.4

3.5.2. Effect of pulses of modifier on retention and efficiency

Dolan gave the following analogy to understand the elution mechanism in gradient LC [26]: "a solute sits at the head of a column until a strong enough solvent comes along to push it through the column leaving the other solutes behind, then it travels to the column outlet fairly quickly". This will help to understand the explanations below.

In gradient elution, the affinity of solutes towards the mobile phase is favoured as the elution strength increases, sometimes requiring a complex program to accommodate the requirements of solutes [27]. In the chromatographic practice, sometimes such complex gradients imply segments with small slope, followed by strong increases in the modifier content. In this case, solutes suffer a strong acceleration, which benefits the reduction of retention times for solutes already separated. Let us reckon what would happen if this high concentration of organic modifier is kept constant during a certain time, and afterwards, the concentration before the sudden increase is recovered (i.e., a transient increase in organic solvent is generated) (see for example the elution program in Figure 3.1b to d). In this particular case, solutes of high hydrophobicity, eluting after the transition, would pass from moving very fast to do it again slowly. The extent of the magnitude of the effect will depend on the solute hydrophobicity. For brevity, henceforth we will refer to the transient increases of organic solvent as "pulses".
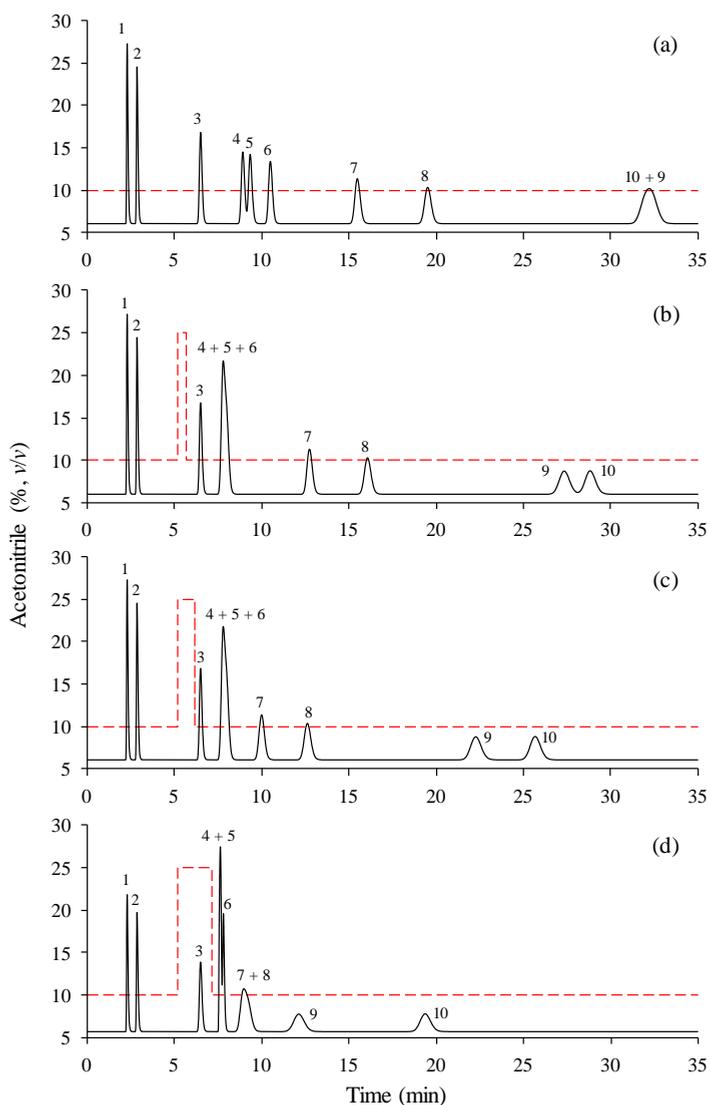
Figure 3.1. Effect of the application of a pulse of increasing duration in the separation of sulphonamides: isocratic elution with 10% acetonitrile (a), and pulses where the concentration of acetonitrile was increased to 25% with duration of 0.5 min (b), 1.0 min (c), and 2.0 min (d). All pulses were applied at 5.16 min (4.0 min + 1.16 min dwell time). See Section 3.4.1 for compound identity and other details.

The above effects would be undesirable with optimisation purposes, due to the increased peak overlapping, especially in (and immediately after) the pulse region, and because solutes suffer important drops in efficiency in the region right after the pulse, due to the reduction in the migration speed. However, as will be shown, the use of pulses is beneficial for modelling. The effect is similar to translating in block late eluting compounds to earlier times (i.e., giving the separation an impulse in between, using a type of stepped gradient with positive and negative sudden changes). Pulses are thus a special case of multi-step isocratic run.

The use of "pulse experiments" constitutes an interesting possibility to have access to chromatographic information for highly hydrophobic solutes at low modifier contents (i.e., the contents before and after the pulse), because the retention times will be significantly decreased. Note that, along a conventional gradient, the highest modifier concentrations that a solute experiences are the only ones contributing significantly to solute migration, while the effect of smaller concentrations becomes negligible except for very fast eluting solutes. In other words, transient increases or pulses provide an insight about the behaviour of solutes in the lowest part of the experimental design, which otherwise would give rise to excessive retention times for hydrophobic solutes, but are decisive for the separation of solutes. In addition, narrower ranges of organic modifier give rise to more uncertain parameters in modelling studies.

In order to explain the effect of the use of pulses of organic modifier on the elution of a mixture (a subset of 10 sulphonamides), two sequences of simulated chromatograms are first shown. In each run, the concentration of modifier was maintained at 10%, before and after the pulse where it was abruptly increased to 25%. In the first sequence (Figure 3.1), the pulse is applied at a time $t = 4$ min and the duration of the pulse is increased from 0.5 to

2 min, along the subsequent experiments. In the second sequence (Figure 3.2), a pulse of 1.0 min duration is shifted gradually from 2 to 4 min. It should be noted that the elution programs are plotted taking into account the delay due to the dwell volume (i.e., the actual location of the pulse at the column inlet is depicted). The elution in the absence of pulses is plotted for comparison purposes (Figures 3.1a and 3.2a). In all cases, the first two solutes (solutes 1 and 2), located before the pulses, are naturally not affected. The details relative to the simulation of chromatograms are given in Sections 3.3.1 to 3.3.3.

Figure 3.1 shows that solutes eluting after the pulse shortened their elution with regard to the absence of pulse. The reduction in retention is considerably larger for late eluting compounds (solutes 7 to 10), which were more significantly affected at increasing pulse duration (the longer the retention, the larger the reduction). Although the aim of this work is only modelling, it can be observed that pulses have also effects on selectivity. Thus, solutes 9 and 10, which co-eluted in the absence of pulse, are well separated when a pulse is applied. Also, the resolution of solutes 4 to 6 was changed. Note that the pulse affects the solutes after an extra delay produced by the dead time. Therefore, in the chromatograms, the effects of a pulse will be perceptible after a time $t_p + t_{dwell} + t_0$, where t_p is the time at the programmed pulse start.

When a pulse of fixed duration is shifted to longer times (Figure 3.2b to d), the most retained solutes are unaffected. Meanwhile, the four intermediate solutes experience important variations in their elution. When the elution of a solute is very close to the end of the pulse, its retention time is scarcely affected (see solute 3 in Figure 3.2c and d).
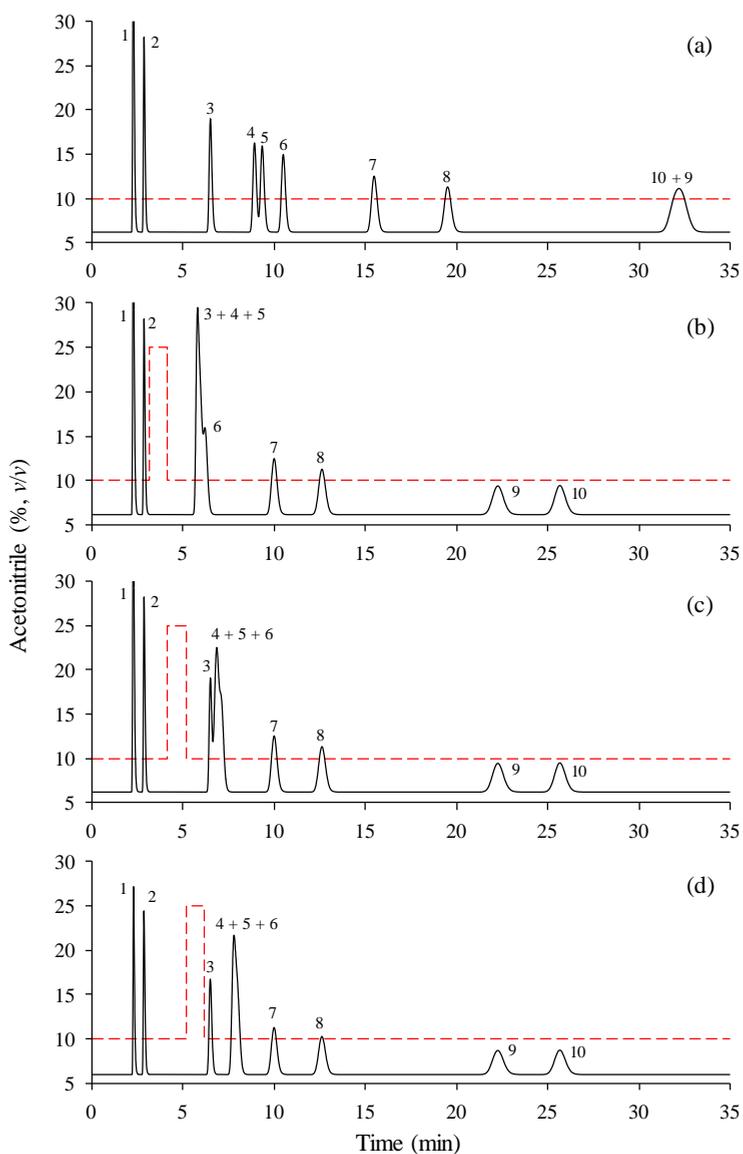
Figure. 3.2. Effect of the shift of a pulse of 1 min duration in the separation of sulphonamides, starting at 2 min (b), 3 min (c), and 4 min (d), to which the dwell time (1.16 min) was added. Isocratic elution with 10% acetonitrile is given for comparison purposes in (a). See Section 3.4.1 for compound identity and other details.

3.5.3. Modelling of retention times

3.5.3.1. Deviations of raw predictions with regard to experimental chromatograms

To inspect in more detail the effect of pulses of organic modifier in real chromatograms, four experiments were designed with the set of 10 sulphonamides (Figure 3.3). Pulses of 1 min of duration gradually shifting to longer times were run. In each run and after the pulse, the elution ended with a 10–25% acetonitrile gradient in the 10–12 min range (plus the dwell time). If the isocratic elution would have been kept after the pulse (without the applied gradient), the elution of solutes 8 to 10 would have reached around 25 min (Figure 3.2b). It should be noted that the least retained sulphonamide (sulphaguanidine) appears split in two peaks due to the degradation of the compound. Also, note that there is no consequence in the baseline associated to the application of both the pulse and the rapid gradient, owing to the same buffer concentration level (0.01 M phosphate buffer) used in the two solutions being mixed (10 and 25% acetonitrile) to generate pulses and gradients. In general, the column pressure fluctuation due to the transient change of organic solvent in the mobile phase varied between 119 and 136 bars (see Figure 3.4). As observed in the sequence of simulated chromatograms obtained for isocratic elution (Figure 3.2), the position where the pulse is applied does not affect the peaks eluting far enough behind the pulse. Only peaks eluting in the neighbourhood inside or after the pulse are affected, depending on the pulse duration.
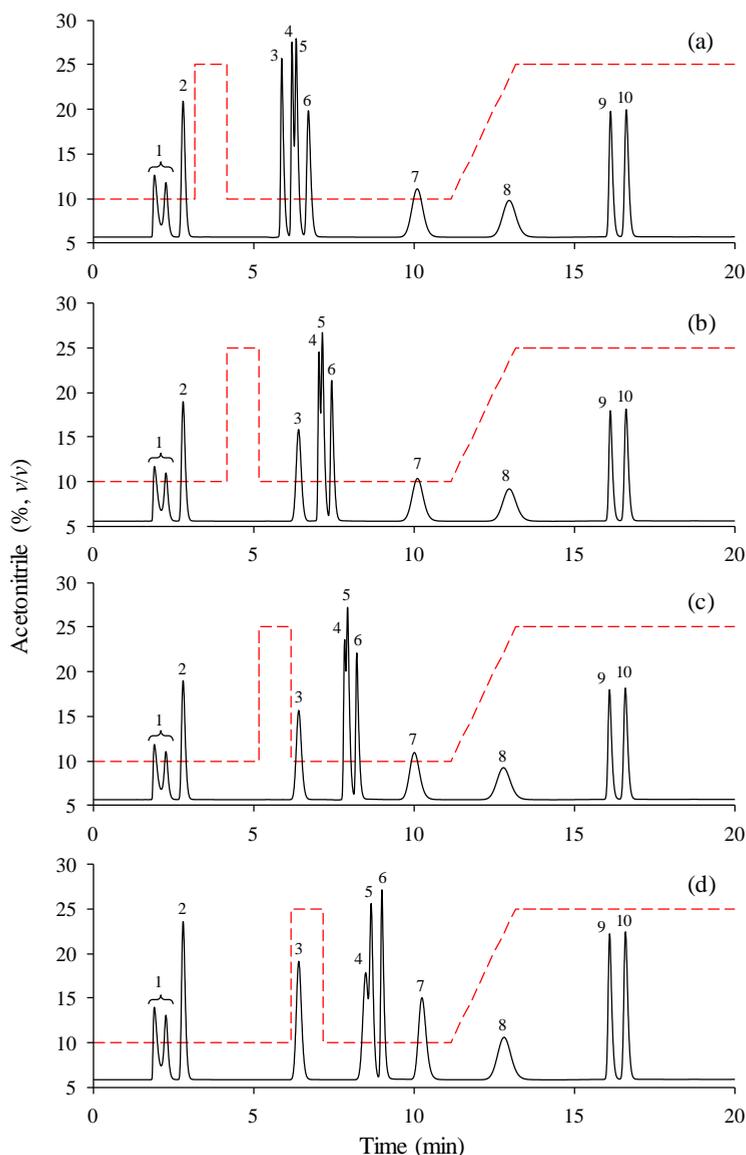
Figure 3.3. Experimental chromatograms showing the effect of the shift of a pulse of 1 min duration in the separation of sulphonamides, starting at 2 min (a), 3 min (b), 4 min (c), and 5 min (d), to which the dwell time (1.16 min) was added. A fast gradient was applied between 10 and 12 min. See Section 3.4.1 for compound identity.
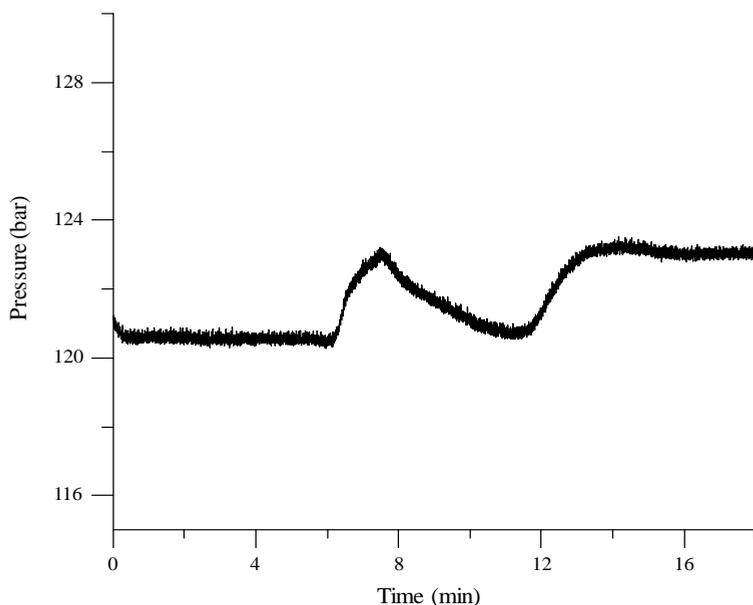
Figure 3.4. Pressure profile along the pulse experiment, corresponding to Figure 3.5c.

Figures 3.5a and b show predicted chromatograms obtained by applying a pulse of organic modifier at 5.16 min (4 min + dwell time). The experimental chromatogram obtained in the laboratory is depicted in Figure 3.5c. The predictions were carried out according to the Neue-Kuss model, using two experimental designs containing as training set: (i) the data from the five isocratic runs of the training set (Figure 3.5a), and (ii) a mixed set where the data obtained with the four pulse experiments in Figure 3.3 were processed altogether with the five isocratic runs (Figure 3.5b).
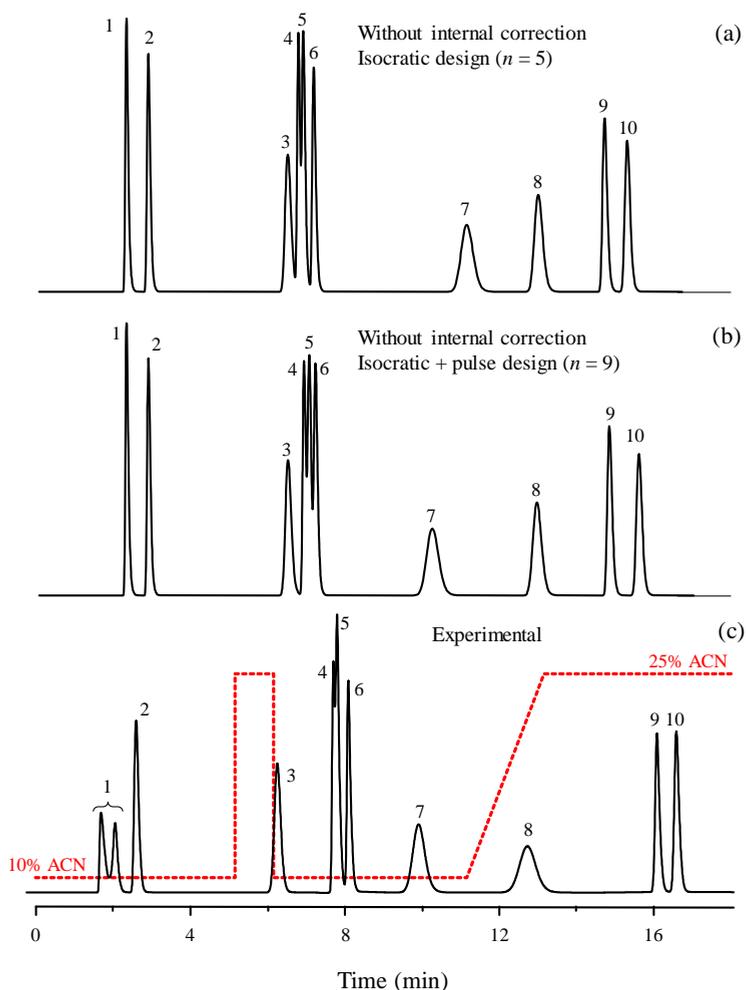
Figure 3.5. Predicted (a,b) and experimental (c) chromatograms for the elution program plotted in (c) for a mixture of sulphonamides. Predictions were made with training sets constituted by: (a) the five isocratic experiments indicated in Section 3.4.1, and (b) the five isocratic experiments and the four pulses shown in Figure 3.3. The elution program consisted of a pulse of 1 min duration starting at 4 min, followed by a fast gradient between 10 and 12 min (to which the dwell time, 1.16 min, was added). See Section 3.4.1 for compound identity and other details.

In principle, the prediction errors obtained with the Neue-Kuss model would guarantee very low error in the prediction of retention times (see Table 3.1). Taking this into account, it calls strongly the attention the deficient predictions of retention times in the presence of sudden changes in the concentration of organic modifier (Figure 3.5). It could also be thought that the training set of isocratic runs do not provide enough information for the predictions and hence the errors. However, the incorporation of information from chromatograms obtained applying pulses, to the experimental design, scarcely improved the situation (Figure 3.5b). This suggests that it is necessary to consider another type of correction in the numerical integration, to improve predictions of retention for solutes eluting close to the pulses.

3.5.3.2. *Correction of retention times in numerical integration due to the migration inside the column*

One of the objectives of this work is the improvement in the predictions under critical gradient conditions, such as gradients that include sudden changes in the modifier concentration, or situations where extreme compositions participate, such as pulses. It should be considered that, when a change in the composition of the mobile phase (i.e., a gradient change, or a transient increase) reaches the column inlet, its effects always take a small additional time to reach the solute neighbourhood, owing to the distance travelled by the solute from the inlet. Thus, as commented in Section 3.3.3, the instant composition experienced by the solute includes two delays: one of them is the dwell time (associated to the distance travelled by the mobile phase from the mixer to the column inlet), and the second one is an intra-column delay accounting the gradual discrepancy between the gradient program and the instant composition at the solute location. Naturally, the magnitude of such

intra-column delay increases along the solute migration, and the need for a correction becomes more mandatory (the maximal delay is t_0 and is found at the column outlet). The intra-column delay is considered in the analytical integration in gradients with a single ramp, because it is based on the application of the Barrow's rule, which states that the integral is calculated at two points, one of them where the intra-column delay is t_D (the lower limit of the integral), and the other at the outlet, where the intra-column delay is $t_g - t_0$ (the upper limit of the integral in Equation (3.5)).

Numerical integration requires correcting each infinitesimal term up to the column outlet. The intra-column delay up to a certain column location was incorporated in the prediction of retention times through Equations. (3.12) and (3.13). It should be noted that the true variable in $k(\varphi(t))$ is the organic solvent and only when the delay implies a variation in the solvent content, the correction is significant [20].

Two different scenarios are next considered: chromatograms obtained using a linear or multi-linear gradient (Figure 3.6), and by the application of a pulse (Figure 3.7). In each case, the predictions without taking into account the intra-column correction in the numerical integration (Figures 3.6a and d), and considering this correction (Figures 3.6b and e, and 3.7a and b), are compared with the experimental chromatograms (Figures 3.6c and f, and 3.7c).
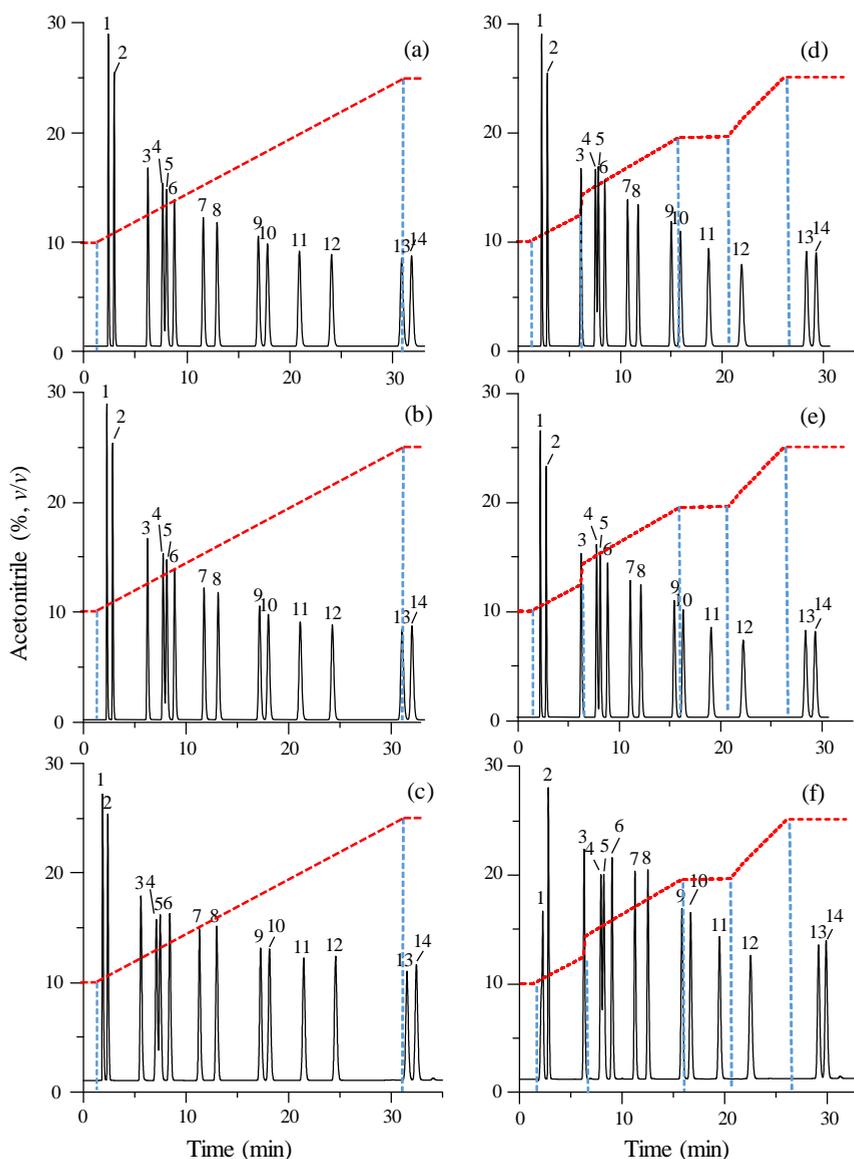
Figure 3.6. Predicted (a,b,d,e) and experimental (c,f) chromatograms for linear (a,b,c) and multi-linear (d,e,f) gradients, applied to the separation of sulphonamides. The predictions were made by numerical integration without (a,d) and with (b,e) intra-column corrections. The gradient programs are overlaid. See Section 3.4.1 for compound identity and other details.
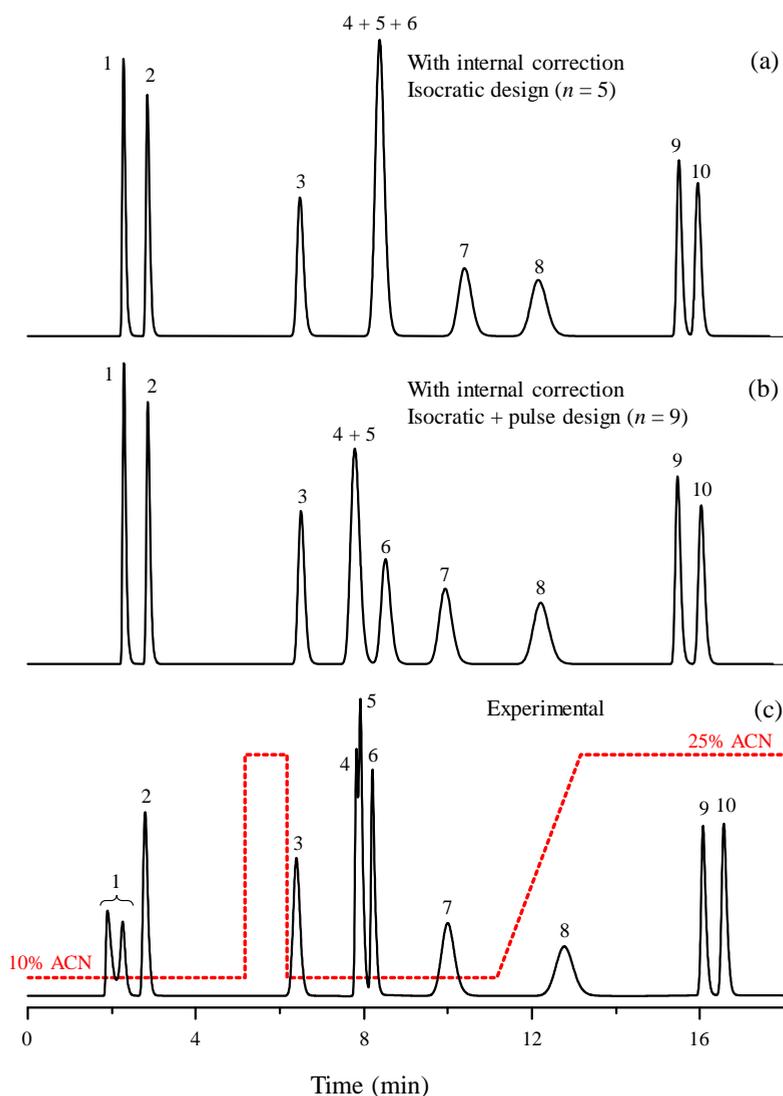
Figure 3.7. Predicted (a,b) and experimental (c) chromatograms in the separation of sulphonamides. Predictions were made by numerical integration with training sets constituted by: (a) the five isocratic experiments indicated in Section 3.4.1, and (b) the five isocratic experiments and the four pulses shown in Figure 3.3. In this case, in contrast to Figure 3.5, the intra-column corrections were applied. See Figure 3.5 for more details.

In the first scenario (Figure 3.6), where the composition changes are gradual and increasing, the effects of intra-column delays were not significant, since the difference in organic modifier at the column inlet does not differ appreciably from that in the solute neighbourhood. Therefore, the intra-column corrections were not translated in large differences in the prediction of retention times. This explains why this type of correction (i.e., intra-column delays) is not considered in numerical integration for gradients. In spite of this, if the chromatograms in Figure 3.6 are inspected in detail, the discrepancies and performance of the correction are perceptible (see for instance the peaks of solutes 11 and 12 in Figures 3.6d, e and f; the vertical dashed lines delimiting the transitions between the linear segments in the gradient help to appraise the differences in retention).

In the second scenario, in which a pulse is applied in a chromatogram ending with a linear gradient, the improvements in the predicted chromatograms when the intra-column corrections are applied (Figure 3.7) are appreciable with regard to the raw predictions (see Figure 3.5). This can be explained taking into account that, when sudden changes in organic modifier occur, solute speed inside the column may vary drastically between the two extreme compositions that affect the solute neighbourhood: the speed can be very slow if the solute is migrating at the lower pulse composition, and very fast when it migrates at the highest composition. The difference in speed in the proximity of the change in organic solvent at the end of the pulse can be critical for solutes of close polarity and lead to the observed important prediction errors if the intra-column correction is neglected.

Figure 3.7 shows the consequences of including the intra-column correction in the second scenario, both in the predictions obtained with the training set of five isocratic experiments, and with the extended design including the isocratic

experiments plus the four runs containing the pulses altogether (Figure 3.3a to d). Note that the intra-column correction does not require including in the design the information obtained with pulses to originate fairly good predictions for most peaks. The agreement between experimental and predicted chromatograms is very satisfactory, although not perfect. Problems such as those associated with deformations in the pulse profile, diffusion effects, or re-equilibration of the stationary phase may contribute to these small variations.

Figure 3.8 complements the information in Figure 3.7, depicting simulated chromatograms considering the intra-column delays, which should be compared with the experimental chromatograms in Figure 3.3. The predicted chromatograms in Figure 3.8 were obtained using the training set of five isocratic experiments and the intra-column corrections for the predictions. Here, we must indicate that peak profiles were predicted using the Jandera's approximation [28], which as observed, tends to overestimate the peak width close to the pulse (the prediction for the other peaks is satisfactory).

Finally, we must remark the importance of using a small step in the numerical integration for obtaining chromatograms including pulses. In this work, we have used an integration step of 0.0001 min. Figure 3.9 shows the consequences of using variable integration steps ranging from 10^{-5} to 0.1 min. A value of 0.01 min is usually a good choice for conventional linear and multi-linear gradients. In the presence of strong transitions, especially at decreasing concentrations, such step is insufficiently small, giving rise to occasional wrong predictions for peaks eluting critically close to the end of the applied pulse.
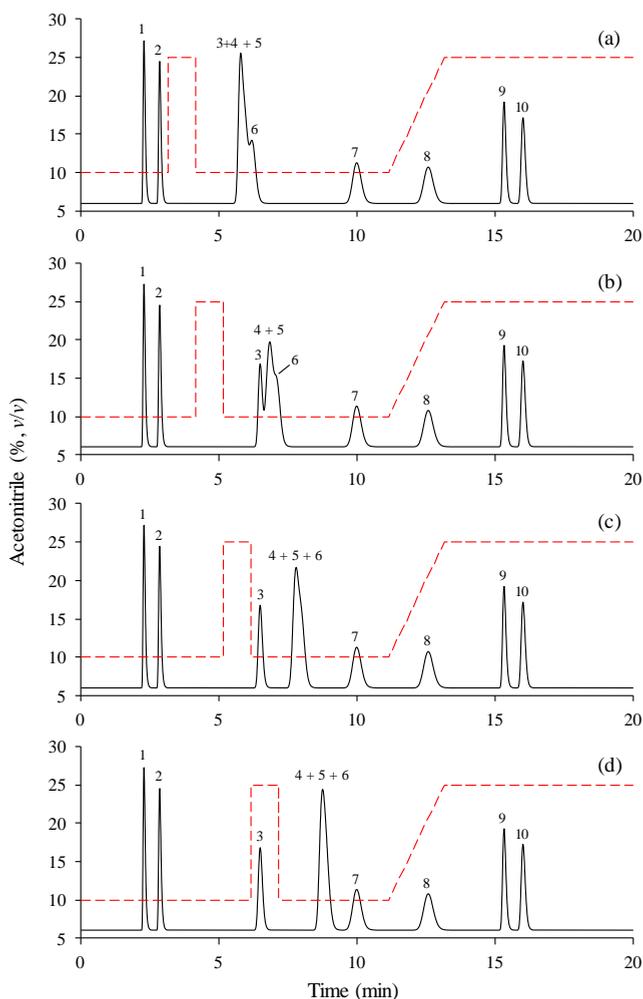
Figure 3.8. Predicted chromatograms showing the effect of the shift of a pulse of 1 min duration in the separation of sulphonamides, starting at 2 min (a), 3 min (b), 4 min (c), and 5 min (d), to which the dwell time (1.16 min) was added. A fast gradient was applied between 10 and 12 min. See Figure 3.1 for more details. The chromatograms were obtained by numerical integration, using the training set of five isocratic experiments and the intra-column corrections for the predictions. They should be compared with the experimental counterparts in Figure 3.3.
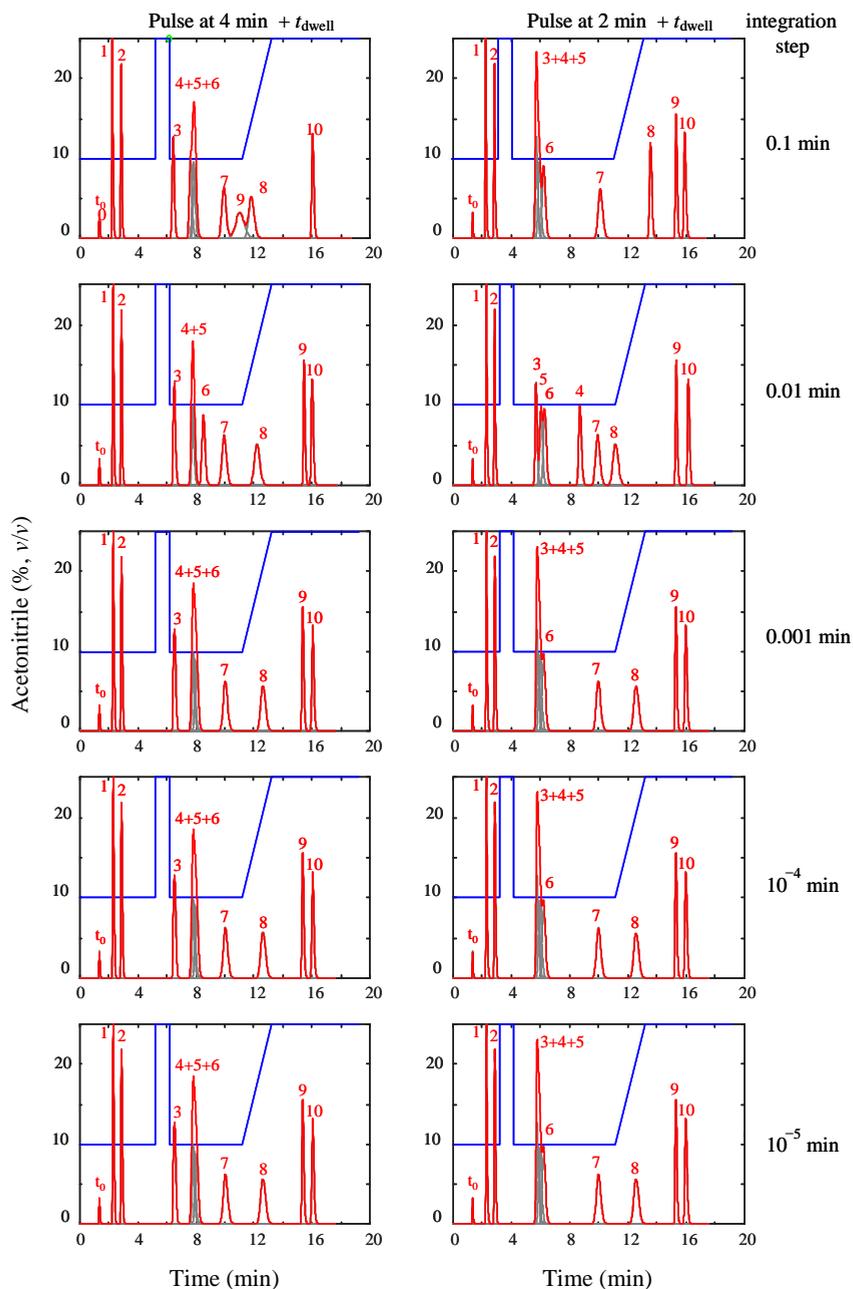
Figure 3.9. Effect of the integration step on the quality of the predictions.

3.5.4. Predictive performance of experimental designs involving pulses

As commented above, the use of pulses gives rise to a significant reduction of retention times for the most hydrophobic solutes, when these are eluted using pure isocratic conditions (Figures 3.1 and 3.2). In this section, the performance of mixed experimental designs constituted by isocratic runs and pulses is evaluated and compared with designs including conventional linear gradients, whose purpose is also the reduction of analysis times. Here it should be mentioned that several software packages commercially available implement training sets constituted of simple linear gradients for modelling purposes, usually according to the linear solvent strength model (Equation (3.1)). In this work, we have fitted the different training sets to the Neue-Kuss model (Equation (3.4)), which offers low error in wide ranges of organic solvent in the mobile phase, as commented in Section 3.5.1. The applied software was written in our laboratory, which allowed more flexibility in the design of experiments and in the calculations, with regard to commercial software, and a more fair comparison. Moreover, it is not possible to know all details in the construction of commercial software, which may lead to wrong conclusions.

For non-linear fitting, it is important to choose appropriate initial values for the model parameters, to avoid be trapped into a local optimum. The fittings started giving arbitrary positive values, identical for all solutes in a first run. The parameters found for those solutes where the regression was successful were then used, in a next run, as initial values for those solutes that failed. Usually, $k_w = 800$, $c = 4$, and $B = 50$ were adequate default values to fit the Neue-Kuss model (Equation (3.4)), for typical solutes and favourable designs. For the most hydrophobic solutes and less informative designs, larger k_w values (e.g., $k_w = 5000$) were more convenient.

3.5.4.1. Validation and experimental designs

The evaluation of the performance (benchmarking) involved several aspects:

(i)   Checking the quality of the parameters of the retention models found with experimental designs that include pulses, with regard to those obtained from purely isocratic designs, which are the richest in information.

(ii) Evaluate the predictive capability of models obtained from different experimental designs involving pulses or gradients. The evaluation implies both the prediction of the training data (those used in the fitting of the retention models), and external predictions for isocratic mobile phases of low elution strength. These compositions participate at the start of gradients, in a minor extent, and this is translated in more serious errors.

(iii) Evaluate if the designs that include pulses are competitive with regard to solvent consumption and analysis time, in comparison to pure isocratic and pure gradient designs.

To perform the study, we inspected the two slowest solutes in the set of 14 sulphonamides (sulphadimethoxine and sulphaquinoxaline, solutes 13 and 14, Tables 3.2 and 3.3), which are the most affected by the use of pulses. For solutes 7 to 12, the benefits are smaller. The measurement of retention times for solutes 13 and 14, analysed using isocratic elution, was problematic, since at the lowest concentration in the experimental design (10% acetonitrile) they amounted to around 2.5 and 3 hours, respectively.

Table 3.2. Predictive performance for the two most hydrophobic sulphonamides (see Section 3.5.4) and acetonitrile consumption for the experiments in the training sets.[a,b]

Design set	Experiment	Sulphadimethoxine		Sulphaquinoxaline		Waste
		$t_{R,target}$	$t_{R,found}$	$t_{R,target}$	$t_{R,found}$	Acetonitrile (mL)
5 isocratic	1	154.91		180.32		18.03
	2	76.17		88.82		11.55
	3	51.06		58.74		8.81
	4	22.83		24.70		4.94
	5	12.57		12.58		3.15
		Cumulative analysis time / 5 experiments (min)				365.16
		Total waste (mL)				46.48
		$t_{R,target}$	$t_{R,found}$	$t_{R,target}$	$t_{R,found}$	Acetonitrile (mL)
1 pulse + 4 isocratic	1	69.33	69.33	83.70	83.70	9.62
	2	76.17	76.17	88.82	88.84	11.55
	3	51.06	51.05	58.74	58.75	8.81
	4	22.83	22.83	24.70	24.70	4.94
	5	12.57	12.57	12.58	12.57	3.15
		Cumulative analysis time / 5 experiments (min)				268.54
		Total waste (mL)				38.07
		$t_{R,target}$	$t_{R,found}$	$t_{R,target}$	$t_{R,found}$	Acetonitrile (mL)
2 pulses + 3 isocratic	1	52.23	52.22	64.38	64.38	7.94
	2	56.44	56.45	67.72	67.72	9.48
	3	51.06	51.07	58.74	58.75	8.81
	4	22.83	22.84	24.70	24.72	4.94
	5	12.57	12.57	12.58	12.58	3.15
		Cumulative analysis time / 5 experiments (min)				228.13
		Total waste (mL)				34.33

Table 3.2 (continued).

Design set	Experiment	Sulphadimethoxine		Sulphaquinoxaline		Waste
		$t_{R,target}$	$t_{R,found}$	$t_{R,target}$	$t_{R,found}$	Acetonitrile (mL)
5 gradients, variable φ_{end}	1	43.81	43.81	46.05	46.05	7.18
	2	53.48	53.49	56.89	56.90	8.33
	3	74.10	74.12	82.15	82.17	10.79
	4	92.97	92.96	105.67	105.66	12.82
	5	35.33	35.31	36.68	36.67	6.11
				Cumulative analysis time / 5 experiments (min)		327.44
				Total waste (mL)		45.23
5 gradients, variable t_G	1	17.27	17.27	17.48	17.47	3.72
	2	21.57	21.58	22.00	22.00	4.29
	3	30.13	30.15	31.00	31.02	5.42
	4	37.65	37.66	39.23	39.25	6.41
	5	43.81	43.78	46.05	46.04	7.18
				Cumulative analysis time / 5 experiments (min)		55.77
				Total waste solute (mL)		27.02

[a] For the isocratic design, the acetonitrile concentration was 10, 13, 16, 20 and 25% for experiments 1 to 5, respectively. For the other designs, the experiments are indicated in Figure 3.10. [b] See Section 3.5.4 for the meaning of $t_{R,target}$ and $t_{R,found}$.

Table 3.3. Parameters of the Neue-Kuss retention model (Equation (3.4)) obtained from the designs in Figure 3.10, and predicted retention times for the two isocratic experiments with strongest retention for the two most hydrophobic sulphonamides.

Designs	Neue-Kuss parameters			Isocratic t_R (min)	
	k_w	c	B	10% acetonitrile	13% acetonitrile
		Sulphadimethoxine			
5 isocratic experiments	6617 ±179	4.52 ±0.06	68.2 ±0.79	154.9	76.2
1 pulses + 4 isocratic	6673 ±321	4.53 ±0.7	68.4 ±1.1	155.0	76.2
2 pulse + 3 isocratic	6590 ±637	4.51 ±0.13	68.1 ±2.1	154.8	76.2
5 gradients, variable φ_{end}	6106 ±816	4.37 ±0.25	66.1 ±3.5	154.4	76.2
5 gradients, variable t_G	4172 ±3210	3.88 ±1.08	58.3 ±16.2	147.4	75.0
		Sulphaquinoxaline			
5 isocratic experiments	4816 ±88.6	3.30 ±0.04	53.9 ±0.50	180.3	88.9
1 pulses + 4 isocratic	4817 ±159	3.30 ±0.04	53.9 ±0.67	180.4	88.8
2 pulse + 3 isocratic	4752 ±300	3.29 ±0.08	53.7 ±1.22	180.1	88.8
5 gradients, variable φ_{end}	4618 ±399	3.23 ±0.16	52.9 ±2.1	179.9	88.8
5 gradients, variable t_G	2930 ±2233	2.65 ±1.02	44.5 ±13.9	169.0	86.9

The study considered, on the one hand, the isocratic design measured in the laboratory, which included five mobile phases (10, 13, 16, 20 and 25% acetonitrile, Section 3.4.1), and on the other, the following four designs (Figure 3.10):

(i) The first design is the simplest (Figure 3.10a), since it only replaces the slowest experiment in the isocratic design with a pulse of 10 min duration, starting at 20 min including the dwell time. Along this first 20 min, the fastest solutes elute isocratically at 10% acetonitrile, which provides quality measurements for these solutes. The slowest solutes are also benefited because the pulse reduces significantly the elution at 10% acetonitrile.

(ii) The second design contains three isocratic experiments and two pulses of 12 and 15 min duration (Figure 3.10b) that, as in the previous case, begin at 20 min. The pulse duration was established so that the retention time of the slowest solute in both experiments was close to one hour. This implies a reduction to one third of the time that would be obtained if the elution were purely isocratic at 10% acetonitrile.

(iii) The third design includes five linear gradients, which reach the maximal target concentration of organic solvent, φ_{end} (25, 20, 15 and 13% v/v) in 60 min for four of the gradients, and 25% v/v in 40 min for the fifth gradient. The slowest gradient is practically isocratic, so this is one of the most favourable gradient designs that could be considered in information terms.

(iv) The fourth design consists of five gradients in the range 10–25% acetonitrile, with variable gradient time (t_G). This results in a larger reduction of the analysis time with regard to design (iii).
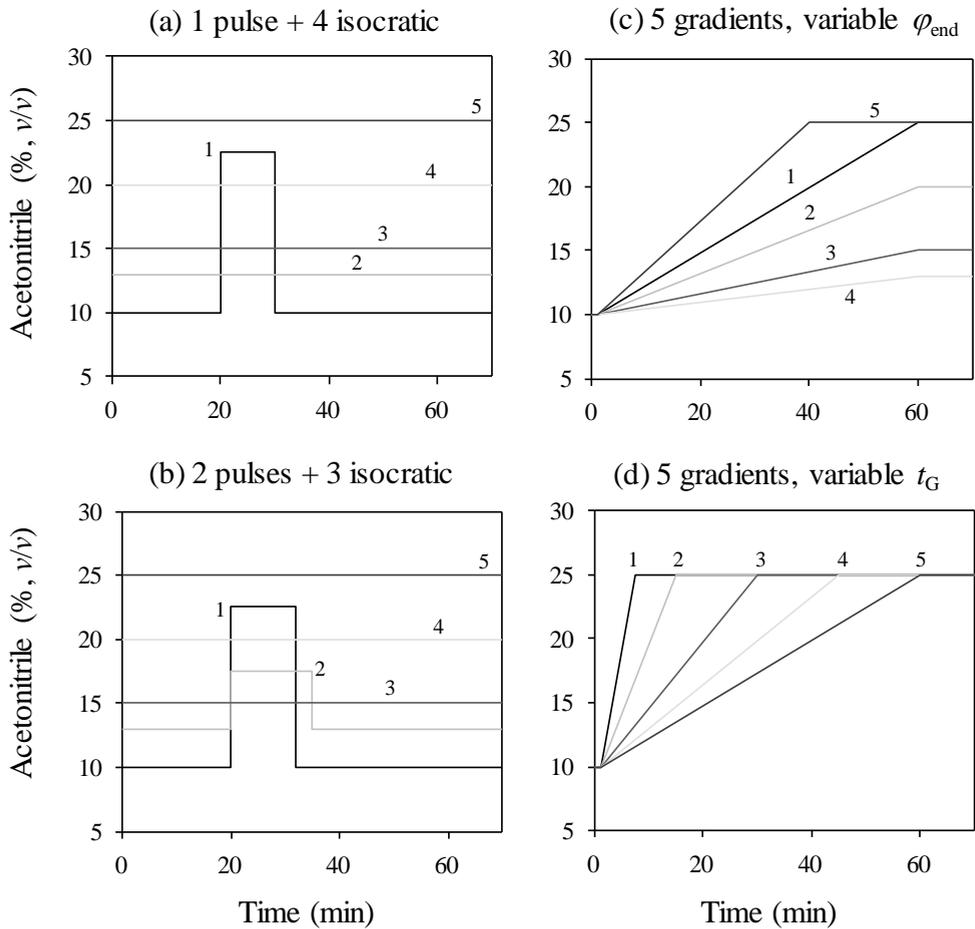
Figure 3.10. Experimental designs used to get the model parameters for solutes 13 and 14.

3.5.4.2. Methodology

With the aim of obtaining comparable results with the different designs in Figure 3.10, the following methodology was applied:

In a first step, the parameters k_w, c and B of the Neue-Kuss model (Equation (3.4)) were obtained by non-linear regression for each solute. In these fittings, the retention times measured with the five isocratic experiments were used as training set (see top of Table 3.2). As commented, the regression statistics are shown in Table 3.1. The model parameters for sulphadimethoxine and sulphaquinoxaline obtained from the isocratic data are indicated in Table 3.3. These parameters are considered as the most accurate and will be used to evaluate the results offered by the other designs.

The retention models obtained with the isocratic data were used to predict the retention times for the two sulphonamides, expected with the four designs in Figure 3.10 ($t_{R,target}$ in Table 3.2). To obtain these values, numerical integration was applied as explained in Sections 3.3.2 and 3.3.3, using an integration step of 10^{-5} min. The $t_{R,target}$ values were used for the subsequent calculations, as if they were experimental data, without any uncertainty. This allowed making a proper comparison of the performance of the different designs. The $t_{R,target}$ values for each design were next used as training data to recover the corresponding Neue-Kuss parameters, and predict the retention times of the two sulphonamides eluted according to the experiments of each design. These calculations were made following the methodology explained in Sections 3.3.2 to 3.3.4. The time consumption is given in Table 3.2 in terms of acquisition time in a single run (analysis time). Naturally, this time will depend on the way the solutes are injected (e.g., 4 solutes by run). The computation time for fitting was around 2 min by solute.

3.5.4.3. Comparison of performances

In Table 3.3, the Neue-Kuss parameters obtained by fitting the $t_{R,target}$ values for the four designs in Figure 3.10 are compared to the parameters evaluated from the experimental isocratic data. As can be observed, the designs with pulses provided model parameters practically coincident with those obtained from the isocratic experiments, being the design with a single pulse the most accurate. The designs containing gradients offered more discrepant parameters, particularly the design in Figure 3.10d, which includes the fastest gradients. Table 3.3 also provides the standard errors in the estimated parameters for each design. From these values, it can be concluded that not only the parameters offered by pulse experiments are more coincident than the gradient data with those obtained from pure isocratic data, but also the confidence intervals are narrower, denoting altogether results of better quality.

As can be observed in Table 3.2, the $t_{R,target}$ values were perfectly reproduced ($t_{R,found}$) by the respective model parameters fitted for each design, with discrepancies between both times ($t_{R,target}$ and $t_{R,found}$) generally around or below 0.01 min. This means that models fitted with the data of each design predict their own training data with excellent accuracy. However, this does not imply an equal predictive capability for out-of-domain experiments (especially for very slow eluents), or experiments corresponding to other designs. Table 3.3 indicates the expected retention time for the slowest eluents in the isocratic design (10 and 13% acetonitrile), using the model parameters indicated in the table. It can be observed that the designs with pulses and the gradient design in Figure 3.10c lead to extrapolated predictions with a quality practically equal to the isocratic experiments. The gradient design in Figure 3.10d provides the worst extrapolated predictions.

Table 3.2 also includes the analysis times associated to the measurement of the five experiments in each design, as well as the consumption of acetonitrile, using the retention time of sulphaquinoxaline. It can be observed that the designs with pulses imply intermediate analysis times and solvent waste, between those obtained with the two gradient designs studied. It should be taken into account, however, that the gradient design with lowest consumption and analysis time (the gradient design with variable t_G, design 3.10d) was also the one that yielded larger errors in the estimation of the parameters of the retention models, and in the predictions of retention at low concentration of organic solvent (Table 3.3). The reason of the poorer performance of design 3.10d is related to the scarce information in the overall retention that provides for low elution strength compositions, which are important for the separation. Design 3.10c (with variable φ_{end}) improves this insufficiency, but mixed pulse/isocratic designs improve it even more, especially design 3.10a.

It should be noted that the designs containing pulses involve considerable reductions in analysis time and a somewhat smaller solvent waste with respect to pure isocratic elution. Therefore, considering the obtained results globally, it is possible to conclude that the runs with pulses are very competitive with regard to gradient designs for obtaining high quality models.

3.6. Conclusions

Accurate retention modelling is crucial in the field of automatic method development of LC. For this purpose, elaborate designs of experiments are needed. Among the different possible choices in RPLC, isocratic experimental designs provide the richest information about the retention behaviour of solutes, but their use is hampered by the long retention times for the most hydrophobic

compounds at low elution strengths. Gradient experimental designs are able to reduce substantially the long acquisition times, but provide less rich information about the retention behaviour at compositions with low elution strength, which may be decisive for the separation. In this work, we have explored the advantages and disadvantages of the use of isocratic designs in which a transient increment in organic solvent (i.e., a pulse) is inserted in an intermediate position of the elution program at the isocratic experiment(s) with lowest elution strength. The application of pulses in retention time modelling is aimed to reduce the run time of slow solutes in collecting training data for model fitting.

From a practical standpoint, mixed designs with pulses are easily constructed by replacing the slowest isocratic runs by experiments with a pulse inserted at an intermediate time in the elution program. This allows the elution of the fastest solutes in the initial sector of the elution program, previous to the pulse, and the elution of the most retained solutes after the pulse in acceptable times. If the design is carried out running the experiments from higher to lower elution strength, the data fitting from the already measured isocratic runs allows selecting the most suitable duration and position of the pulse in the experiment(s) with less elution strength. Fitting of retention models using data obtained in experiments involving mixed designs with isocratic and pulse runs is simpler than using data from gradient experiments. It implies solving a simplified version of the fundamental equation of gradient elution, owing to the presence of isocratic sectors.

The study has revealed that mixed isocratic designs including pulse(s) offer better predictive capability in extrapolations than experimental designs of gradients, and a shorter measurement time than pure isocratic designs. Retention models fitted from these mixed designs practically match those

derived from pure isocratic experimental designs. The accuracy in prediction of retention times is also comparable, but the analysis time with the slowest eluents is considerably reduced for hydrophobic compounds. This type of mixed design is also competitive in terms of solvent consumption.

Experiments involving pulses showed, however, the existence of discrepancies in predictions close to the pulse when numerical integration was applied, which can be wrongly attributed to insufficient column re-equilibration. These deviations, negligible in conventional gradient experiments, are well evidenced in experiments involving strong alterations in the solvent content along the elution program (such as those that include pulses). The correction of such deviations implies the inclusion of delays in the arrival of gradient changes to the instant position of the solute along the whole migration. Once such intra-column corrections are incorporated in the numerical integration, accurate retention predictions are obtained in pulse experiments. Peak width was, however, overestimated for peaks close to the pulse when the Jandera's approximation was applied. Meanwhile, in the pulse runs, the baseline shows no perceptible alteration upon the introduction of sudden variations.

Numerical predictions ignoring corrections of the intra-column delay will be accurate as long as solutes do not elute close to the pulse. When the objective is developing an experimental design including pulses, and the pulse location can be set arbitrarily, the best choice is locating it in an empty intermediate region of the chromatogram (taking into account that its effects will take some time to disappear after the pulse ends: $t_{end} + t_0$). For solutes eluting too close to the pulses, moving the pulse location may be more practical to avoid introducing the intra-column corrections. For intermediate solutes (such as solutes 7 and 8

in this study), which are far enough beyond the pulse, the benefits are similar but smaller than those presented for slower solutes (such as solutes 9 and 10).

Finally, it is worth to comment that to shorten the time needed for slow solutes, there are other alternative approaches:

(i) Collection of the training data at higher isocratic concentration levels (e.g., 15, 20, 25 and 30% (*v/v*) versus 10, 13, 16, 20 and 25% (*v/v*) used in this work). For evaluating this possibility, we must take into account that a design should satisfy the requirements of information of solutes of very diverse polarity. Shifting the design to higher concentrations will be detrimental for the fastest solutes, since their elution is accelerated excessively making them to co-elute with the void volume. In addition, removing the data for 10 and 13% acetonitrile means that an eventual gradient would start with an unfavourable situation for the fastest solutes, which would elute in a too narrow time window.

(ii) Training data can be collected at higher flow rate (e.g., retention times can be reduced by a factor of two if the flow rate is doubled). This can be a valid strategy, whenever the heat diffusion effects are negligible, and the column allows larger pressure. Naturally, increasing the flow rate is not mutually exclusive with the use of pulse experiments: pulses can be applied at higher flow rate, as well.

(iii) Collection of training data on another column of same chemistry and different diameter and particle size (e.g., using a 50×2.1 mm column with particle size of 1.7 µm and flow rate of 0.2 mL/min, the corresponding retention time of each solute will be reduced by a factor of about 2.88 with regard to the original 150×4.6 mm column with particle size of 5 µm and flow rate of 1.0 mL/min). Again, this is another valid option. The

only problem is having a column in the laboratory with scaled-down geometry (and the appropriate pump).

3.7. References

[1] G. Guiochon, The limits of the separation power of unidimensional column liquid chromatography, *J. Chromatogr. A* 1126 (2006) 6–49.

[2] L.R. Snyder, J.J. Kirkland, J.W. Dolan, *Introduction to Modern Liquid Chromatography*, 2nd ed., John Wiley & Sons, New York, 2011.

[3] M.C. García Álvarez-Coque, J.J. Baeza Baeza, G. Ramis Ramos, Reversed phase liquid chromatography, in *Analytical Separation Science Series*, Vol. 1 (edited by J. Anderson, A. Berthod, V. Pino, A.M. Stalcup) Wiley, New York, 2015.

[4] M.C. García Álvarez-Coque, J.R. Torres Lapasió, J.J. Baeza Baeza, Models and objective functions for the optimisation of selectivity in reversed-phase liquid chromatography, *Anal. Chim. Acta* 579 (2006) 125–145.

[5] R. Cela, E.Y. Ordoñez, J.B. Quintana, R. Rodil, Chemometric-assisted method development in RPLC, *J. Chromatogr. A* 1287 (2013) 2–22.

[6] P. Jandera, Can the theory of gradient liquid chromatography be useful in solving practical problems?, *J. Chromatogr. A* 1126 (2006) 195–218.

[7] J.E. Haky, D.A. Teifer, Gradient elution, in *Encyclopedia of Chromatography*, Taylor and Francis, New York, 2006.

[8] L.R. Snyder, J.W. Dolan, *High-Performance Gradient Elution*, John Wiley & Sons, Hoboken, NJ, 2007.

[9] J.W. Dolan, L.R. Snyder, Gradient elution chromatography, in *Encyclopedia of Analytical Chemistry* (edited by R.A. Meyers), John Wiley & Sons, New York, 2012.

[10] J.R. Torres Lapasió, M.C. García Álvarez-Coque, Levels in the interpretive optimisation of selectivity in high-performance liquid chromatography: A magical mystery tour, *J. Chromatogr. A* 1120 (2006) 308–321.

[11] A. Pappa-Louisi, P. Nikitas, Non-linear least-squares fitting with Microsoft Excel Solver and related routines in HPLC modelling of retention: I. Considerations of the problems of the method, *Chromatographia* 53 (2000) 477–486.

[12] A. Pappa-Louisi, P. Nikitas, Non-linear least-squares fitting with Excel Solver and related routines in HPLC modelling of retention: II. Considerations of selection of optimal fit, *Chromatographia* 53 (2000) 487–493.

[13] A. Pappa-Louisi, P. Nikitas, P. Balkatzopoulou, C. Malliakas, Two- and three-parameter equations for representation of retention data in reversed-phase liquid chromatography, *J. Chromatogr. A* 1033 (2004) 29–41.

[14] P. Nikitas, A. Pappa-Louisi, Retention models for isocratic and gradient elution in reversed-phase liquid chromatography, *J. Chromatogr. A* 1216 (2009) 1737–1755.

[15] P.J. Schoenmakers, H.A.H. Billiet, L. de Galan, Description of solute retention over the full range of mobile phase compositions in reversed-phase liquid chromatography, *J. Chromatogr*. 282 (1983) 107–121.

[16] U.D. Neue, H.J. Kuss, Improved reversed-phase gradient retention modeling, *J. Chromatogr. A* 1217 (2010) 3794–3803.

[17] P. Nikitas, A. Pappa-Louisi, Expressions of the fundamental equation of gradient elution and a numerical solution of these equations under any gradient profile, *Anal. Chem.* 77 (2005) 5670–5677.

[18] J.J. Baeza Baeza, M.C. García Álvarez-Coque, Some insights on the description of gradient elution in reversed phase-liquid chromatography, *J. Sep. Sci.* 37 (2014) 2269–2277.

[19] C. Ortiz Bolsico, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Optimization of gradient elution with serially-coupled columns. Part I: Single linear gradients, *J. Chromatogr. A* 1350 (2014) 51–60.

[20] C. Ortiz Bolsico, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Optimisation of gradient elution with serially-coupled columns. Part II: Multi-linear gradients, *J. Chromatogr. A* 1373 (2014) 51–60.

[21] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B. Flannery, *Numerical Recipes: The Art of Scientific Computing*, 3rd ed., Cambridge University Press, New York, 2007.

[22] J.R. Torres Lapasió, *MICHROM Software*, Marcel Dekker, New York, 2000.

[23] I.E. Frank, R. Todeschini, *The Data Analysis Handbook: Data Handling in Science and Technology*, Vol. 14, Elsevier Science, Amsterdam, 1994.

[24] S.N. Deming, Y. Michotte, D.L. Massart, L. Kaufman, B.G.M. Vandeginste, *Handbook of Chemometrics and Qualimetrics*, Part B, Elsevier Science, Amsterdam, 1998.

[25] J.A. Navarro Huerta, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Estimation of peak capacity based on peak simulation, *J. Chromatogr. A* 1574 (2018) 101–113.

[26] J.W. Dolan, The hazards of adjusting gradients, *LCGC North Am.* 20 (2002) 940–946.

[27]   V. Concha Herrera, G. Vivó Truyols, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Limits of multi-linear gradient optimisation in reversed-phase liquid chromatography, *J. Chromatogr. A* 1063 (2005) 79–88.

[28]   P. Jandera, Predictive calculation methods for optimization of gradient elution using binary and ternary solvent gradients, *J. Chromatogr.* 485 (1989) 113–141.

CHAPTER 4

TESTING EXPERIMENTAL DESIGNS IN LIQUID CHROMATOGRAPHY: DEVELOPMENT AND VALIDATION OF A METHOD FOR THE COMPREHENSIVE INSPECTION OF EXPERIMENTAL DESIGNS

4.1. Abstract

The basis of interpretive optimisation in liquid chromatography is the prediction of resolution, from appropriate solute retention models. The reliability of the process depends critically on the quality of the experimental design. This work develops, validates and applies a general methodology aimed to evaluate the quality of any training experimental design. The methodology is based on the systematic evaluation of the uncertainties associated to the prediction of retention times in comprehensive scans of both isocratic and gradient experimental conditions. It is able to evaluate comprehensively experimental designs of arbitrary complexity. Five common training experimental designs were used to model the retention, according to the Linear Solvent Strength (LSS) and the Neue-Kuss (NK) equations, using a set of 14 sulphonamides of different polarity. The results are presented in terms of relative uncertainties in predictions, which provide significant and interpretable results. The magnitude of such uncertainties, together with the systematic, coherent and logical changes observed at decreasing solute polarity, give support to the results. The NK model offered smaller errors and unbiased predictions, whereas the LSS model gave rise to lack of fit. Isocratic training designs, which are widely accepted as the most informative, are confirmed as the best. As a general conclusion, gradients are predicted with intrinsically smaller uncertainties, independently of the training experimental design. In addition, gradients are more insensitive than isocratic predictions with regard to the type of training design used. Isocratic predictions deteriorate quickly at larger content of organic modifier in the mobile phase. This explains the better performance of gradient predictions, even with biased models.

4.2. Introduction

The need for providing quality and informative data, with a minimal experimental effort, is imperative in scientific areas where the acquisition of information is slow or laborious; such is the case of liquid chromatography (LC). Among the possible ways to obtain quality data, the most rational alternative is the use of design of experiments (DOE) [1,2]. Nowadays, DOE is, together with the Process Analytical Technology, the main tool in the Quality-by-Design paradigm [3]. In LC method development, DOE has been traditionally used to: (i) rank controlled experimental factors and set preliminary conditions (e.g., explore columns, solvents), (ii) fine-tune the most influent factors (method optimisation), and (iii) assess or forecast the robustness of selected optimal conditions [2]. This work is focused on the second item, more precisely on the assessment of the quality of experimental designs in the field of conventional reversed-phase liquid chromatography (RPLC), from which the best experimental conditions to fulfil resolution (i.e., method optimisation) will be inferred.

The basis of systematic LC method development is the enhancement of resolution, usually with the assistance of retention models, which relate the solute retention time with a series of experimental factors, such as the organic solvent content or the temperature. Retention models are built with the information obtained from standards, by fitting the data acquired from a small number of carefully planned experiments that follow a certain distribution: the experimental design [4–6]. The reliability of the expectancies of resolution, and therefore, the success of chromatographic optimisation depends critically on the quality of the solute retention models, and hence the importance of experimental designs, which determine the quality of the information available for the fittings.

Usually, method development starts with a "scouting gradient", which helps to establish the most suitable elution mode and solvent range for the sample. Then, retention data are acquired from a small number of isocratic or gradient runs (i.e., the training experimental design), which are set according to the distribution of polarity of the sample components. From this training design, proper retention models are established.

Isocratic experimental designs are maximally informative and their treatment is rather simple, but they suffer from the important drawback of needing a long time for data acquisition [7]. For this reason, in spite of being less informative, gradient experiments are considered more practical [8]. Gradient training designs are able to provide the information needed for the fittings, although finding an optimal gradient set for modelling is not straightforward. Moreover, the inclusion of the gradient complexity remains beyond the possibilities of the best DOE strategies (based on derivatives), because the calculation of gradient retention implies the resolution of an integral equation, which is only possible in some circumstances. Otherwise, only geometrical DOE strategies, based on properties such as orthogonality, rotatability, or uniformity [9], are applicable.

According to the aim, the designs used in LC can be divided, considering first the number of variables under study, in screening and optimisation designs [10]. With five or more variables, screening designs (fractional factorial [11] or Plackett-Burman [12]) are needed to rank and reduce the variables. Once the most important variables (four or less) have been found, central composite [13], Box-Behnken [14], or Taguchi [15] designs are applied, normally in combination with polynomial retention models. With specific models and the support of their mathematical properties, other more elaborated and comprehensive proposals are possible. D- or G-optimal designs belong to this category, and analyse the

properties of the so-called "design matrix" [16,17], which relates the distribution of the experiments with the retention model.

G-optimal experimental designs have been applied to ionisable solutes, in isocratic domains at constant temperature with variable pH and organic solvent content [18]. The strategy used in that work was based on the addition of an experimental point in the region of the design with the largest expected prediction error. This process was repeated sequentially. Crossed predictions from isocratic experimental data to gradient elution, and vice-versa, have also been studied under the perspective of the errors associated to the transference [19,20]. However, predictions were strongly constrained by calculation issues, and resulted unfeasible in many situations of highly practical interest. In addition, the slow calculation speed prevented the massive evaluation of partial derivatives of the retention model along the evaluation of the quality of the design. This happens whenever the combination of retention models and elution program does not yield an algebraic expression, when the fundamental equation of gradient elution is solved. Recently, an excellent predictive capability has been found in unconventional designs, where isocratic experiments are combined with runs involving temporary rises (pulses) in organic solvent [21]. It was found that these unusual designs, whose acquisition time was competitive with gradients, allowed good predictions for slow solutes at low solvent contents (see Chapter 3).

A methodology, able to rank by quality any set of experimental designs in both isocratic and gradient predictions, is still needed to discover the best training experimental design. This work proposes, develops, validates and applies a universal methodology for assessing the quality of training experimental designs, that can be applied to arbitrary configurations (involving multi-linear gradients, multi-isocratic runs, runs with pulses, etc.), in order to reveal the best. The method is not constrained by the availability of algebraic expressions for the

prediction of retention times in gradient elution. Using exact model parameters fitted from isocratic data, the predictive capability of the retention model is evaluated when the training experimental design is exhaustively applied to both isocratic and gradient predictions. A set of solutes of diverse polarity, well and poorly covered by the training designs, are analysed. The approach is applied to analyse the predictive performance and properties of five of the most usual training experimental designs in RPLC, using the Linear Solvent Strength (LSS) and Neue-Kuss retention models. The final aim is deriving conclusions about the design performance, finding eventually the optimal training designs, which allow the best fittings and subsequent predictions with the target retention model.

4.3. Theory

4.3.1. Prediction of retention times

For this study, two widely applied retention models have been considered. One of them was proposed by Snyder [8] and is the core of the LSS theory, extensively used in routine laboratories and optimisation software. It upholds a linear relationship between the logarithm of the retention factor (k) and the volume fraction of organic solvent in the mobile phase (φ):

$$\log k = \log \frac{t_R - t_0}{t_0 - t_{ext}} = \log k_w - S\,\varphi \qquad (4.1)$$

where t_R, t_0 and t_{ext} are the isocratic retention, dead and extra-column times, respectively. This model includes two adjustable parameters, namely the elution strength (S) and the logarithm of the retention factor in pure water ($\log k_w$). Equation (4.1) offers good predictions only in intermediate and narrow solvent concentration domains. It is applicable to gradient elution, but its integration is only possible for specific situations, such as when the gradient program consists

of a unique linear ramp. Composite solutions are, however, possible for multi-linear ramps. From now on, Equation (4.1) will be referred as the LSS model.

In 2010, Neue and Kuss [22] proposed a model with good performance in wide organic solvent domains:

$$k = k_{\mathrm{w}}(1+c\,\varphi)^2\, e^{-\frac{B\,\varphi}{1+c\,\varphi}} \tag{4.2}$$

where c is a curvature parameter, and B, a parameter related to the distribution of the solute between both phases, due to hydrophobic interactions. This parameter coincides with S in Equation (4.1) when the relationship is linear ($c = 0$). This model not only gives rise to excellent fittings under isocratic elution, but it also has antiderivative under linear gradients. It will be called the Neue-Kuss (NK) model henceforth.

Single linear gradients starting after a certain time delay t_{D} (i.e., dwell time) with slope m and intercept a, so that $\varphi = \varphi_0$ at $t = t_{\mathrm{D}}$, can be expressed as:

$$\varphi = a + mt = \varphi_0 + m(t - t_{\mathrm{D}}) \tag{4.3}$$

For such linear gradients, the primitive functions or antiderivatives ($I(t)$) of the LSS and NK models are, respectively:

$$I(t) = \frac{e^{S(a+mt)}}{k_{\mathrm{w}}\, S\, m} \tag{4.4}$$

$$I(t) = \frac{e^{\frac{B(a+mt)}{1+c(a+mt)}}}{k_{\mathrm{w}}\, m\, B} \tag{4.5}$$

The retention time in gradient elution is obtained by working out the upper limit of the fundamental equation [23]:

$$t_0 - t_{\text{ext}} = \int_0^{t_D} \frac{dt}{k(\varphi_0)} + \int_{t_D}^{t_g - t_0} \frac{dt}{k(\varphi(t))} = \frac{t_D}{k_0} + I(t_g - t_0) - I(t_D) \tag{4.6}$$

For the LSS model, and after operating Equation (4.6), the following expression is obtained:

$$t_g = t_0 + t_D + \frac{1}{S\,m} \ln\left[1 + S\,m\,(t_0\,k_w\,e^{-S\varphi_0} - t_D)\right] \tag{4.7}$$

where, for convenience, the gradient program has been shifted to compensate t_D, so that the linear ramp starts at $t = 0$.

For the NK model, the following analytical solution is found:

$$t_g = \frac{\ln H\,[1 + c(a - mt_0)] - B(a - mt_0)}{m(B - c\ln H)} \tag{4.8}$$

where H is given by:

$$H = m\,B\ln k_w\left(t_0 - \frac{t_D}{k_w}\right) + e^{\frac{B(a + mt_D)}{1 + c(a + mt_D)}} \tag{4.9}$$

Equations (4.7–4.9) are valid for solutes eluting along a single gradient ramp. With multi-linear gradients, the solution requires calculating each linear segment separately. In this case, Equation (4.6) needs to be adapted to include the different gradient segments. The contributions (Equations (4.4) or (4.5)) of those segments where the solute still remains inside the column after the completion of the segment(s) must be added:

$$t_0 - t_{\text{ext}} = \int_0^{t_D} \frac{dt}{k(\varphi_0)} + \int_{t_D}^{t_1} \frac{dt}{k(\varphi_1(t))} + \ldots + \int_{t_{n-1}}^{t_g - t_0} \frac{dt}{k(\varphi_n(t))} =$$

$$= \frac{t_D}{k_0} + \sum_{i=1}^{n-1} \left(I(t_{i+1}) - I(t_i) \right) + I(t_g - t_0) - I(t_{n-1})$$

(4.10)

Finally, the gradient retention time (t_g) in the last term is worked out. A more practical solution is finding numerically t_g from Equation (4.6) by applying root-finding methods [24].

The next section introduces the basis of the systematic calculation of the uncertainties, in the prediction of retention times under isocratic and gradient elution associated to any experimental design of arbitrary configuration and complexity.

4.3.2. Construction of isocratic and gradient error maps

The error propagation theory allows the calculation of the uncertainty associated to the prediction of any complex expression $F(x_1, x_2, \ldots)$, as a function of the uncertainties of the variables (e.g., s_{x1}, $s_{x2}, \ldots$), through the general equation [16]:

$$s_F^2 = s_{x_1}^2 \left(\frac{\partial F}{\partial x_1} \right)^2 + s_{x_2}^2 \left(\frac{\partial F}{\partial x_2} \right)^2 + \ldots$$

(4.11)

The application of Equation (4.11) to the calculation of uncertainties in the prediction of isocratic and gradient retention times, by propagation of the uncertainties in the regressed parameters obtained from a set of isocratic or gradient experiments, is next explained.

4.3.2.1. Calculation of uncertainties

The calculation of uncertainties is based on the combination of Jacobian matrices, which gather the partial derivatives of the retention model with regard to its n_{par} parameters (c_1, c_2..., e.g., $c_1 = k_w$ and $c_2 = S$ for Equation (4.1)), for each available experimental condition i. Two Jacobian matrices are involved:

(i)   **J**$_{train}$, which contains the partial derivatives of the retention model, corresponding to the experiments used to fit the model (i.e., the training experimental design),

(ii) **J**$_{pred}$, which contains the partial derivatives of the retention model for the experiment (or experiments) whose uncertainty (and eventually, retention) is being predicted.

As can be seen, **J**$_{train}$ and **J**$_{pred}$ include contributions from the solute model and the design geometry. The first one participates through the parameters c_i in the respective models, and the second, through the values of the parameters describing the experiments. Note that each (training or predicted) gradient experiment i is defined by a specific set of parameters in Equation (4.3): a (or φ_0) and m.

Generically, both Jacobian matrices can be written as follows:

$$
\mathbf{J} = \begin{pmatrix}
\left[\dfrac{\partial t_g}{\partial c_1}\right]_{i=1} & \left[\dfrac{\partial t_g}{\partial c_2}\right]_{i=1} & \cdots & \left[\dfrac{\partial t_g}{\partial c_{n_{par}}}\right]_{i=1} \\[2ex]
\left[\dfrac{\partial t_g}{\partial c_1}\right]_{i=2} & \left[\dfrac{\partial t_g}{\partial c_2}\right]_{i=2} & \cdots & \left[\dfrac{\partial t_g}{\partial c_{n_{par}}}\right]_{i=2} \\[2ex]
\vdots & \vdots & \ddots & \vdots \\[2ex]
\left[\dfrac{\partial t_g}{\partial c_1}\right]_{i=n_{exp}} & \left[\dfrac{\partial t_g}{\partial c_2}\right]_{i=n_{exp}} & \cdots & \left[\dfrac{\partial t_g}{\partial c_{n_{par}}}\right]_{i=n_{exp}}
\end{pmatrix}
\tag{4.12}
$$

In this work, t_g refers to the retention time independently of the elution mode (isocratic experiments are a particular case of gradients with $m = 0$). Thus, isocratic and gradient values were inspected with the same methodology. The two contributions mentioned above (model and run) are reflected in the structure of the Jacobian matrices. Thus, $\mathbf{J}_{train}$ (the Jacobian matrix associated to the training design) includes as many columns as fitted parameters (n_{par}) in the solute model, and as many rows as experiments in the training design (n_{exp}). There are necessarily several rows in $\mathbf{J}_{train}$ because the models should involve a number of experiments equal or larger than parameters in the retention model (the degrees of freedom must be zero or positive). Meanwhile, $\mathbf{J}_{pred}$ may have several rows or only one. In this case, and following the nomenclature conventions in Chemometrics, the Jacobian matrix will be a row vector written in lower case and bold characters: $\mathbf{j}_{pred}$, corresponding to the experiment whose uncertainty is being predicted (with $n_{exp} = 1$).

For the Snyder model, the partial derivatives needed for building $\mathbf{J}_{train}$ and $\mathbf{j}_{pred}$ are given by:

$$\frac{\partial t_g}{\partial c_1} = \frac{\partial t_g}{\partial k_w} = \frac{t_0 \, e^{-S\varphi_0}}{1 + \alpha - S \, m \, t_D} \tag{4.13}$$

$$\frac{\partial t_g}{\partial c_2} = \frac{\partial t_g}{\partial S} = \frac{1}{S^2 \, m} \left[\frac{\alpha - S \, m \, t_D - S \, \varphi_0 \, \alpha}{1 + \alpha - S \, m \, t_D} - \ln\left(1 + \alpha - S \, m \, t_D\right) \right] \tag{4.14}$$

where:

$$\alpha = S \, m \, t_0 \, k_w \, e^{-S\varphi_0} \tag{4.15}$$

For the NK model, the partial derivatives are too complex, and numerical approximations are preferable. These derivatives are calculated by departing slightly the parameters from the nominal values, and inspecting the effects on the

retention time of the predicted gradient ($\hat{t}_g$). This inspection can be done by examining the variations in $\hat{t}_g$ unilaterally (i.e., either increasing or decreasing the examined parameter) for saving operations:

$$\frac{\partial[\hat{t}_g(c_1,...,c_i...)]}{\partial c_i} = \lim_{h \to 0} \frac{\hat{t}_g(c_1,...,c_i+h...) - \hat{t}_g(c_1,...,c_i...)}{h} \qquad (4.16)$$

Equation (4.16) includes two sources of error, namely the truncation and round-off errors [25]. The former comes from the elimination of the highest terms in the Taylor series expansion, in which Equation (4.16) is based:

$$\frac{\hat{t}_g(c_1,...,c_i+h...) - \hat{t}_g(c_1,...,c_i...)}{h} = \frac{\partial[\hat{t}_g(c_1,...,c_i...)]}{\partial c_i} + \frac{1}{2}h\frac{\partial^2[\hat{t}_g(c_1,...,c_i...)]}{\partial c_i^2} + ... \quad (4.17)$$

The round-off error has several contributions, the most important being the magnitude of h along the approximation to zero. The truncation and round-off errors can be less significant than the uncertainties in the calculation of $\hat{t}_g$ by solving Equation (4.6) with root-finding methods. Indeed, the predicted gradient retention times are affected by the uncertainty associated to the resolution of the fundamental equation of gradient elution, and this uncertainty can be more serious than the truncation and round-off errors.

In this work, a variation of the Ridders' method [25,26] has been applied for the calculation of the numerical derivatives in the Jacobian matrices. This method is based on extrapolating the central difference to zero (compare with Equation (4.16)):

$$\frac{\partial[\hat{t}_g(c_1,...,c_i...)]}{\partial c_i} = \lim_{h \to 0} \frac{\hat{t}_g(c_1,...,c_i+h...) - \hat{t}_g(c_1,...,c_i-h...)}{2h} \qquad (4.18)$$

Along the calculation, we have monitored some quality measurements related to the noise level of the derivative, when h is too close to zero (see Section 4.5.3.1). Once the Jacobian matrices have been obtained, the variance in the predicted retention time for the solute under study is calculated as follows [16]:

$$s_{\text{pred}}^2 = s_{\text{PE}}^2 \left(\mathbf{j}_{\text{pred}} \, (\mathbf{J}_{\text{train}}^{\text{T}} \, \mathbf{J}_{\text{train}})^{-1} \, \mathbf{j}_{\text{pred}}^{\text{T}} \right) \tag{4.19}$$

Note that $\mathbf{j}_{\text{pred}}$ is written in bold lower case, since it refers to a single experimental condition (a row vector with the partial derivatives). In Equation (4.19), s_{PE}^2 is the square of the so-called pure experimental error:

$$s_{\text{PE}} = \sqrt{\frac{\sum_{i=1}^{n_{\text{exp}}} (t_{\text{g},i} - \bar{t}_{\text{g}})^2}{n-1}} \tag{4.20}$$

while s_{pred}^2 corresponds to the variance of retention times that one would measure with replicated experiments at the centre of the design. In regression problems, s_{PE} is often replaced by the standard error in predictions (*SEP*), which is measured from the scattering of the experimental data around the regression curve (i.e., retention model):

$$SEP = \sqrt{\frac{\sum_{i=1}^{n_{\text{exp}}} (\hat{t}_{\text{g},i} - t_{\text{g},i})^2}{n_{\text{exp}} - n_{\text{par}}}} \tag{4.21}$$

In the above equation, $\hat{t}_{\text{g},i}$ and $t_{\text{g},i}$ are the predicted and experimental retention times at the ith experimental condition in the training experimental design. The validity of Equation (4.21) is subjected to the absence of correlations between errors and retention times (absence of bias).

4.3.2.2. Relative uncertainty maps associated to a training experimental design

In Equation (4.19), $\mathbf{J}_{train}$ quantifies the richness of the information provided by the data used to build the retention models, whereas $\mathbf{j}_{pred}$ quantifies the level of difficulty of the experimental condition to be predicted, so that a prediction too prone to error will have high $\mathbf{j}_{pred}$ values. Predictions of uncertainty associated to a given experimental design can be comprehensively inspected from the $\mathbf{J}_{pred}$ matrix (instead of a $\mathbf{j}_{pred}$ vector), by changing gradually (scanning) the predicted experimental conditions and keeping $\mathbf{J}_{train}$ constant.

The uncertainty values (s_{pred}) can be plotted as a function of the variable systematically changed along the scan (i.e., the mobile phase composition in isocratic elution and the slope in gradient elution), giving rise to "prediction uncertainty maps". We will consider two types of scan of predicted conditions and representations:

(i)   isocratic maps, with a comprehensive scan of mobile phases of increasing solvent concentration ($s_{pred} = F(\varphi)$), and

(ii) gradient maps, where the effects of increasing systematically the slope of the gradient (m) up to reach a maximal value, φ_{max} (set as target to be reached at $t = t_G$), are examined ($s_{pred} = F(m)$).

Both maps allow a comprehensive inspection of the predictive capability of any experimental design under a statistical perspective. For an easier analysis of the results, relative uncertainties in predictions ($s_{R,pred}$) are used:

$$s_{R,pred} = 100 \frac{s_{pred}}{\hat{t}_g} \tag{4.22}$$

which will be plotted as a function of φ or m, instead of the absolute uncertainties (s_{pred}, Equation (4.19)).

4.4. Experimental

4.4.1. Reagents

In this study, a set of 14 sulphonamides was considered: (1) sulphaguanidine, (2) sulphanilamide, (3) sulphadiazine, (4) sulphathiazole, (5) sulphapyridine, (6) sulphamerazine, (7) sulphamethazine, (8) sulphamethizole, (9) sulphamonomethoxine, (10) sulphachloropyridazine, (11) sulphamethoxazole, (12) sulphisoxazole, (13) sulphadimethoxine, and (14) sulphaquinoxaline (Sigma, Roedermark, Germany). Stock solutions containing 100 μg/mL of each compound were prepared with nanopure water (obtained with a purification system of Adrona B30 Trace, Burladingen, Germany), assisted with an ultrasonic bath (from Elmasonic, Singen, Germany).

Chromatographic runs were carried out in both isocratic and gradient modes, using mobile phases prepared with HPLC-grade acetonitrile (Scharlau, Barcelona, Spain) and anhydrous sodium dihydrogen phosphate (Fluka, Germany), until reaching a 0.01 M concentration level with nanopure water. The pH was fixed at 3.0 by addition of HCl and NaOH (Scharlau), both 0.01 M. Duplicate injections were carried out.

All solutions were filtered through 0.45 μm Nylon membranes from Micron Separations (Westboro, MA, USA), before injection into the chromatographic system.

4.4.2. Apparatus and column

The analysis was carried out with an HP1100 chromatograph (Agilent, Waldbronn, Germany), composed of the following modules: quaternary pump, autosampler equipped with 2 mL vials, thermostated column compartment, and UV-Vis detector monitoring at 254 nm. The injection volume was 20 μL, and the

mobile phase flow rate was kept constant at 1.0 mL/min. A Zorbax Eclipse XDB-C18 column (150 mm × 4.6 mm) with 5 μm particle size (Agilent) was used in the analyses. All injections were performed under controlled temperature conditions (25 ºC). The dead time of the system under different mobile phases was determined experimentally through the injection of KBr (from Acros Organics, Fair Lawn, NJ, USA). The dwell time, 1.16 min, was measured with an acetone gradient. The extra-column time was 0.12 min.

A pH-meter (model MicropH 2002, Crison, Barcelona) and a glass membrane electrode (model 8102, Orion, Barcelona), containing a Ag/AgCl reference electrode with 3.0 M KCl solution as salt bridge, were used to measure the pH.

4.4.3. Software

For the acquisition of signals, an OpenLAB CDS LC workstation (Agilent, revision B.04.03) for data acquisition was used. MICHROM [27] was applied for data processing. All other calculations were carried out with home built-in functions written in Matlab 2016b (The MathWorks Inc., Natick, MA, USA).

4.5. Results and discussion

This work develops and validates the methodology described in Section 4.3 to evaluate the quality of experimental designs. The performance is illustrated by investigating comprehensively the properties of five common training experimental designs.

4.5.1. Designs under evaluation

Figures 4.1 and 4.2 show the five training designs considered in this work, all of them containing five runs:

(i) Design ISO1, which consists of a set of isocratic experiments gradually concentrating at low organic solvent contents (10, 13, 15, 20 and 25% acetonitrile), to sample better the solvent range giving rise to high retention.

(ii) Design ISO2, a set of isocratic experiments strongly focused on the domain of low elution strength (10–14% acetonitrile).

(iii) Design G1, with a set of gradients at constant gradient time, t_G (60 min + t_D).

(iv) Design G2, with a set of gradients reaching the same final concentration, φ_F (25% acetonitrile). This design and design G1 are representative of the type of runs used frequently to model gradient data.

(v) Design G3, where the start of the ramps is arranged in three levels, and both t_G and φ_F are varied. This design is intended to combine, to a certain extent, the advantages of the former four designs.

In order to interpret the results, the retention times of the 14 solutes in all runs of each design are overlaid. Some designs were not able to provide the required information for modelling some solutes, because the degrees of freedom were insufficient (df = 0 or < 0). In this work, two similar experiments where the t_g values differed in less than 0.1 min were considered as equal.
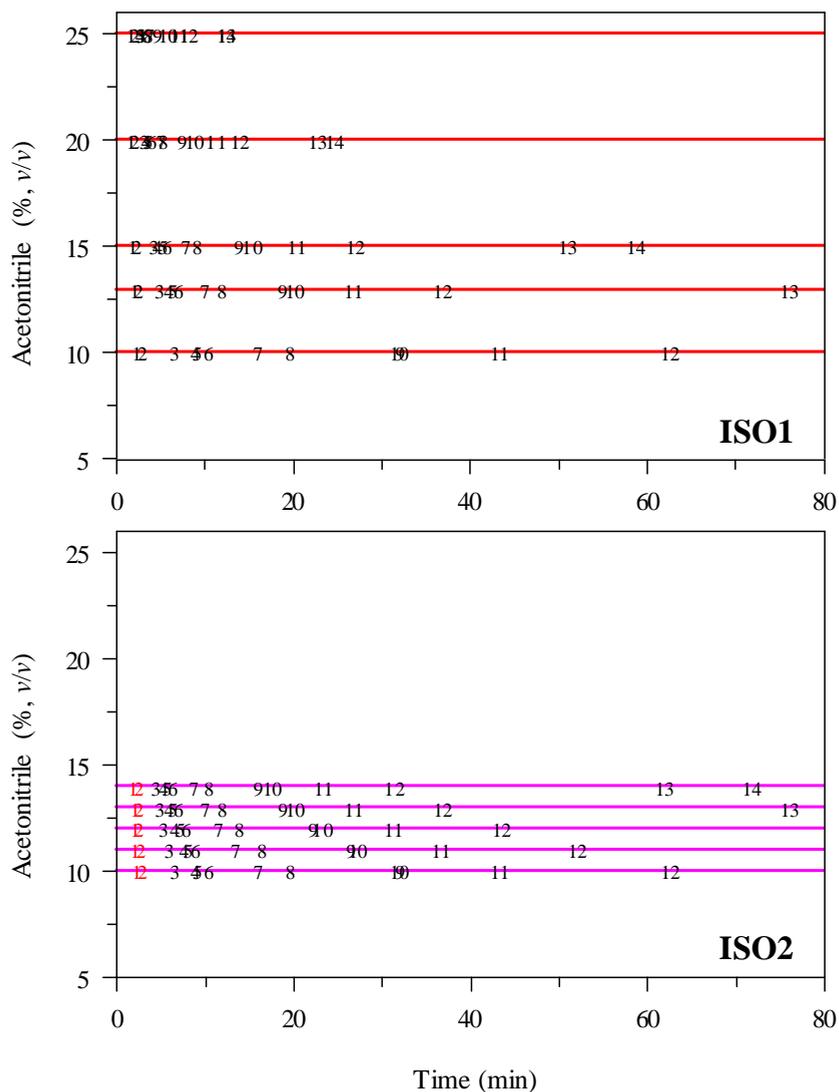
Figure 4.1. Isocratic training experimental designs under evaluation, indicating (overlaid) the solute retention times within each run. Problematic solutes for design ISO2 (see text) were: (1) sulphaguanidine, and (2) sulphanilamide (design ISO1 presented no problems). Some hydrophobic solutes eluting in certain conditions beyond 80 min are not plotted. See Section 4.4.1 for solute identification codes, and Section 4.5.1 for more details.
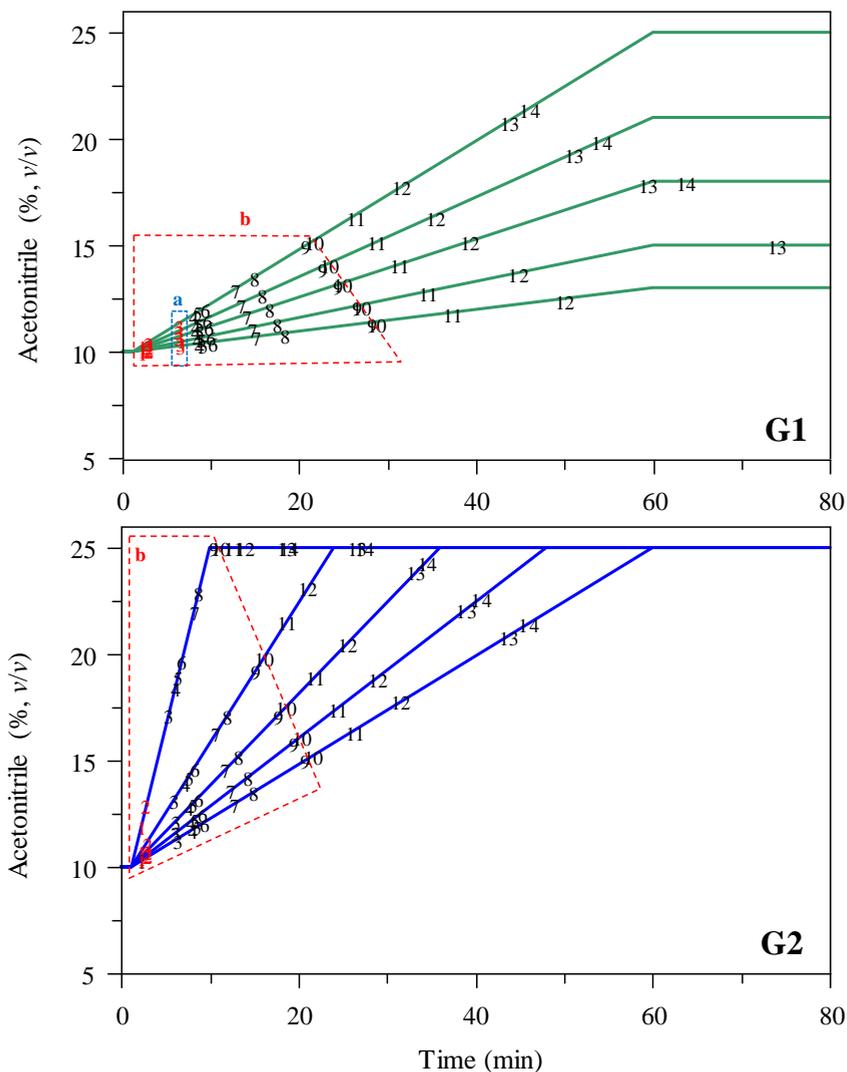
Figure 4.2. Gradient training experimental designs under evaluation. Problematic solutes were: (1) sulphaguanidine, (2) sulphanilamide, and (3) sulphadiazine. Specifically, the coverage was critical for solutes 1 to 3 in G1, 1 and 2 in G2, and 1 in G3. See Figure 4.1 for other details.
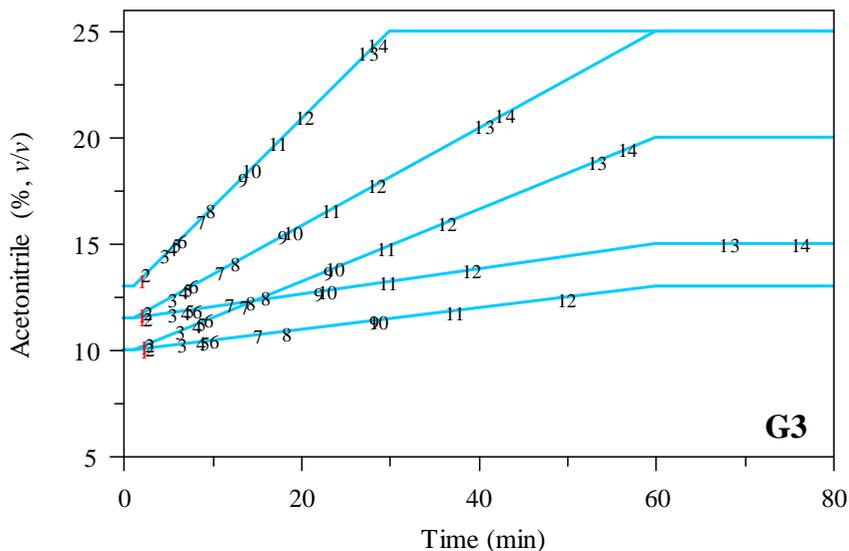
Figure 4.2 (continued).

4.5.2. *Modelling of retention*

The practical use of isocratic experimental designs is limited by the excessive retention times obtained for the most hydrophobic solutes at low solvent contents. Nevertheless, there is a general agreement about the fact that this type of design provides the richest information about solute retention, and the calculated parameters are considered the most reliable. Accordingly, for developing this study, the model parameters for each solute were obtained by fitting the experimental data from the five runs of design ISO1 to the LSS and NK models (Equations (4.1) and (4.2), respectively). Tables 4.1 and 4.2 provide the regressed parameters, the uncertainties in their determination, and other fitting statistics for both models. For the fittings and derived calculations along this work, the modifier concentration was expressed as volumetric fractions (*v/v*), although the plots are shown as percentages.

Table 4.1. Regression parameters and estimated standard errors for the Linear Solvent Strength (LSS) model (Equation (4.1)), using the experimental data acquired with the design ISO1. The following fitting statistics[a] are given: adjusted determination coefficient (R^2_{adj}), mean relative error in predictions (RE) referred to the mean retention, F-Snedecor and standard error in prediction (SEP), which approximates the pure experimental error (s_{PE}) in the absence of lack of fit. The coefficients correspond to fittings carried out with φ expressed as unit fraction (v/v) (also in Table 4.2).

Solutes	LSS model parameters		R^2_{adj}	RE (%)	F–Snedecor	SEP (min)
	k_w	S				
Sulphaguanidine	0.28 ± 0.10	-5.5 ± 0.7	0.98876	5.149	78.1	0.0491
Sulphanilamide	0.64 ± 0.08	-4.4 ± 0.5	0.99283	3.957	85.4	0.0652
Sulphadiazine	2.38 ± 0.11	-9.4 ± 0.8	0.99078	4.884	204.1	0.2281
Sulphathiazole	3.11 ± 0.03	-13.0 ± 0.3	0.98980	5.864	264.5	0.3274
Sulphapyridine	3.04 ± 0.12	-11.9 ± 0.9	0.99078	5.341	266.9	0.3324
Sulphamerazine	3.11 ± 0.11	-11.3 ± 0.9	0.99107	5.155	260.7	0.3750
Sulphamethazine	3.81 ± 0.12	-13.4 ± 1.0	0.99129	5.486	318.3	0.5827
Sulphamethizole	4.22 ± 0.12	-15.3 ± 1.0	0.99232	4.537	403.9	0.6797
Sulphamonomethoxine	4.85 ± 0.12	-16.3 ± 1.0	0.99372	5.073	518.9	1.0262
Sulphachloropyridazine	4.62 ± 0.10	-14.0 ± 0.8	0.99409	4.656	486.5	1.0158
Sulphamethoxazole	4.96 ± 0.03	-14.3 ± 0.3	0.99493	4.378	576.8	1.2840
Sulphisoxazole	5.51 ± 0.02	-16.0 ± 0.2	0.99548	4.288	709.1	1.7471
Sulphadimethoxine	6.9 ± 0.2	-21.2 ± 1.4	0.99286	5.636	541.1	5.3630
Sulphaquinoxaline	7.15 ± 0.02	-21.6 ± 0.2	0.99558	4.481	881.5	4.9273

[a] Fitting statistics:
(see also Equation (4.21))

$$R^2_{adj} = 1 - \frac{\sum_{i=1}^{ne}(\hat{k}_i - k_{exp,i})^2}{\sum_{i=1}^{ne}(k_{exp,i} - \bar{k}_{exp})^2} \times \frac{ne-1}{ne-np-1}$$

$$RE = \frac{\sum_{i=1}^{ne}|\hat{k}_i - k_{exp,i}|}{\sum_{i=1}^{ne} k_{exp,i}} \times 100$$

$$F = (ne-np) \times \frac{\sum_{i=1}^{ne}(\hat{k}_i - \hat{k}_{mean})^2}{\sum_{i=1}^{ne}(k_{exp,i} - \hat{k}_i)^2}$$

In these expressions, ne is the number of runs, np is the number of model parameters (2 for the LSS model), $\hat{k}_i$ and $k_{exp,i}$ are the predicted and experimental retention factors (respectively) for each run i in the design, $\bar{k}_{exp}$ is the mean experimental retention factor, and $\hat{k}_{mean}$ the mean predicted retention factor.

216

Table 4.2. Regression parameters and estimated standard errors for the Neue-Kuss (NK) model (Equation (4.2)), using the experimental data acquired with the design ISO1. The following fitting statistics are given (see Table 4.1 for the mathematical definitions): adjusted determination coefficient (R^2_{adj}), mean relative error in predictions (RE) referred to the mean retention, F-Snedecor and standard error in prediction (SEP), which approximates the pure experimental error (s_{PE}) in the absence of lack of fit.

Solutes	NK model parameters			R^2_{adj}	RE (%)	F–Snedecor	SEP (min)
	k_w	c	B				
Sulphaguanidine	4.40 ± 0.10	5.635 ± 0.010	40.8 ± 0.3	0.99989	0.676	2332	0.0072
Sulphanilamide	4.62 ± 0.03	3.26 ± 0.02	43.84 ± 0.12	1.00000	0.136	38695	0.0025
Sulphadiazine	31.5 ± 0.2	4.046 ± 0.008	37.30 ± 0.06	0.99999	0.235	43960	0.0127
Sulphathiazole	87 ± 4	4.29 ± 0.12	47.8 ± 1.3	0.99999	0.218	88864	0.0144
Sulphapyridine	67.0 ± 1.3	3.90 ± 0.05	41.4 ± 0.3	0.99992	0.562	11168	0.0418
Sulphamerazine	67 ± 2	3.79 ± 0.02	39.1 ± 0.4	0.99989	0.672	7343	0.0575
Sulphamethazine	149 ± 2	3.79 ± 0.02	43.55 ± 0.12	0.99988	0.692	9228	0.0879
Sulphamethizole	237 ± 2	3.68 ± 0.02	46.21 ± 0.11	0.99998	0.313	57212	0.0463
Sulphamonomethoxine	393 ± 4	3.26 ± 0.02	43.84 ± 0.12	0.99990	0.677	13703	0.1625
Sulphachloropyridazine	261 ± 7	3.000 ± 0.005	37.1 ± 0.2	0.99998	0.297	54450	0.0779
Sulphamethoxazole	335 ± 8	2.710 ± 0.015	34.9 ± 0.2	0.99994	0.482	20915	0.1735
Sulphisoxazole	596 ± 12	2.634 ± 0.006	37.1 ± 0.2	0.99995	0.448	30550	0.2172
Sulphadimethoxine	6441 ± 165	4.456 ± 0.009	67.4 ± 0.4	0.99997	0.332	70975	0.3690
Sulphaquinoxaline	4808 ± 59	3.301 ± 0.002	53.91 ± 0.15	0.99999	0.205	214489	0.2518

The assayed retention models show strong differences:

(i) The LSS model yields a rather modest performance, with $R^2_{\mathrm{adj}} > 0.99$ and relative errors in prediction of retention times around 5% (Table 4.1).

(ii) The NK model contrasts strikingly, with an excellent performance: $R^2_{\mathrm{adj}} > 0.9999$, and relative errors in prediction as small as 0.7% and below (Table 4.2).

The modifier concentration range (10 to 25% acetonitrile) was selected attending to the retention times in isocratic elution of sulphonamides, which were wished to be preferably below one hour. This short solvent range should have favoured the LSS model. In spite of this, the statistics indicate that this equation indeed does not describe faithfully the experimental retention behaviour. Nevertheless, it has been included in this work owing to its importance and extensive use, but one should expect bias, uncertain predictions and potential problems.

For a given solute, the selection of a representative condition from which obtaining replicates is not feasible, and this prevents measuring experimentally s_{PE} with Equation (4.20), and then calculating the relative uncertainties in predictions ($s_{\mathrm{R,pred}}$ in Equation (4.19)). The reason lays in the extreme variations in retention time between solutes (and between mobile phases for the same solute), characteristic of chromatographic runs. For this reason, the magnitude of s_{PE} was established from the scattering around the retention model, using all available experiments and Equation (4.21). The last column in Tables 4.1 and 4.2 provides the standard error in predictions (*SEP*), which as commented, approximates the pure experimental error with Equation (4.19).

4.5.3. Relative uncertainty maps associated to a training experimental design

Equation (4.19) allows predicting the standard error associated to any chromatographic run, with the information provided by any training set. This section illustrates the computation of the relative uncertainty maps ($s_{R,pred}$ plots).

4.5.3.1. Calculation of Jacobian matrices

The sets of parameters for the LSS and NK models (see Tables 4.1 and 4.2) were directly obtained by fitting the experimental retention data as a function of φ, for each model and solute. It should be taken into account that, when the design is not appropriate, the solute parameters obtained by regression will be biased, and this bias will distort the calculation of the Jacobian matrices, masking the influence of the design under evaluation. For this reason, in the treatment developed for this work, this influence was minimised through the use of the model parameters obtained from design ISO1 for the calculation of $\mathbf{J}_{train}$ and $\mathbf{J}_{pred}$. As already commented (Section 4.2), isocratic designs are the most reliable in terms of accuracy in predictions, but they are less used owing to the long retention times, characteristic of mobile phases of low elution strength.

For the LSS model, $\mathbf{J}_{train}$ and $\mathbf{J}_{pred}$ can be readily calculated with Equations (4.13–4.15), which are valid for those conditions (i.e., rows in the Jacobian matrices) where the solute elutes along the ramp in the gradient. These expressions require modifications if the solute elutes isocratically beyond the end of the ramp ($t_g > t_G$), or within the dwell time. For the NK model, the calculation through the use of algebraic derivatives is not possible with reasonably simple mathematical expressions, even in the most ideal case where the solute(s) elute(s) within the ramp.

Owing to all these limitations, we decided to carry out the evaluation of the Jacobian matrices by numerical procedures, since these can be applied to any isocratic or gradient experiment, independently of the complexity or the instant along the elution program where the solute leaves the column. In the numerical evaluation, the derivatives were estimated by the Ridders' method (see Section 4.3.2.1), from predicted gradient retention times, $\hat{t}_g$. In turn, the $\hat{t}_g$ values were worked out from Equation (4.10), using analytical antiderivatives (Equations (4.4) and (4.5)). For this purpose, the Newton-bisection method was applied, following an approach explained elsewhere [24]. In that report, we gave answer to the efficient calculation of $\hat{t}_g$ in situations where algebraic solutions of Equation (4.10) are not feasible, because combinations of models and gradient programs lacking antiderivatives are involved.

In the estimation of the Jacobian matrices, the derivatives were obtained by decreasing exponentially the parameter h in Equation (4.18), according to $h(c_i) = c_i\,(1+0.5^\lambda)$, λ being the sequence of natural numbers. Three measurements were monitored along the contraction of h (see Equations (4.16) to (4.18)): (i) the partial derivative $(\partial \hat{t}_g / \partial c_i)$, (ii) the predicted retention time $\hat{t}_g$, and (iii) parameter c_i. As far as the differences between two consecutive estimations of $(\partial \hat{t}_g / \partial c_i)$, and between the values of $\hat{t}_g$ and c_i in the same iteration are all of them larger than 1000-fold the precision of the machine, the new value of $(\partial \hat{t}_g / \partial c_i)$ was updated. The final result of the partial derivative is taken from the last $(\partial \hat{t}_g / \partial c_i)$ value, where the three parameters indicated above still fulfilled the validity condition.

4.5.3.2. Construction of uncertainty maps

The evaluation of the uncertainties implied two types of designs, namely training and sampling. Both designs are combined in Equation (4.19), and when s_{pred}^2 is plotted versus the factor scanned in $\mathbf{J}_{\text{pred}}$ (i.e., the composition in isocratic designs or the slope in gradient designs), an absolute uncertainty map is obtained. This kind of map explores the predictive performance of a training design using the experiments of the sampling design. In this work, five training designs have been investigated (see Section 4.5.1). As the product $(\mathbf{J}^T_{\text{train}} \cdot \mathbf{J}_{\text{train}})^{-1}$ in Equation (4.19) was constant in a given map, it was calculated only once for each training design, which allowed saving operations. In contrast, $\mathbf{J}_{\text{pred}}$ (associated to the sampling design) needs to be calculated for the scan of conditions under evaluation. Note that we are referring here to more than one predicted condition, and hence, the Jacobian of prediction is a matrix: $\mathbf{J}_{\text{pred}}$.

For a given training design (e.g., any of the designs in Figures 4.1 and 4.2), we have studied the effects in both isocratic and gradient predictions, using specific sampling designs for each type of elution from which the $\mathbf{J}_{\text{pred}}$ term in Equation (4.19) is calculated. Figure 4.3a shows the isocratic sampling design, constituted by 16 isocratic runs. It allows inspecting the information provided by the training design, when isocratic mobile phases are predicted (in the 10 to 25% v/v acetonitrile range, in 1% increments). Figure 4.3b depicts the gradient sampling design, with 31 gradients showing progressively larger slope evenly distributed, at a constant angular increment of 3º. Among the 31 gradients, 16 complete the ramp at $t_G = 60 + t_D$ min (each of them at a different target concentration, φ_F), whereas the other 15 gradients have variable t_G, although they reach the same φ_F value: 25% acetonitrile. The overlaid lines crossing the isocratic and gradient runs show the retention times of each solute for each run.
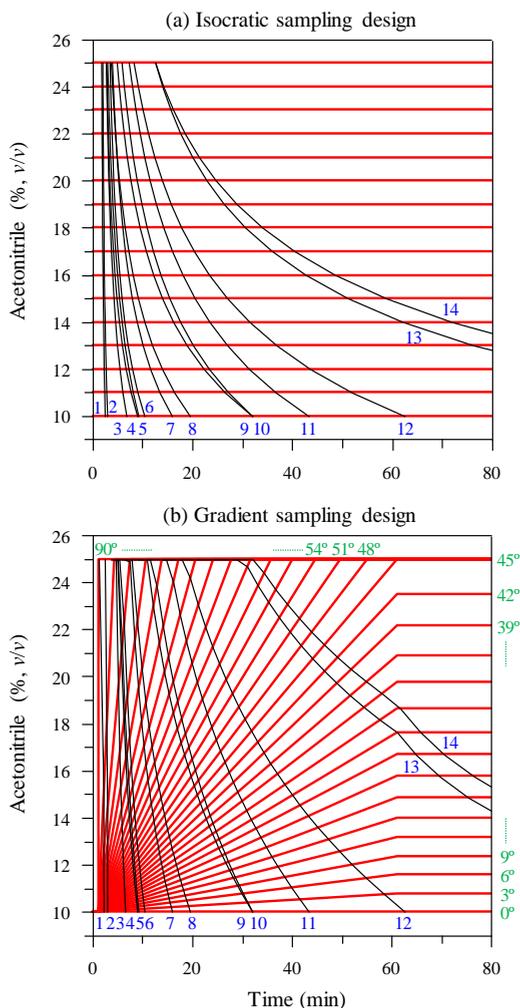
Figure 4.3. Isocratic (a) and gradient (b) sampling experimental designs (red lines) used to assay the performance of the five training designs in Figures 4.1 and 4.2. Black lines indicating the retention time of each solute under each scanning run are overlaid (the solute code is indicated, see Section 4.4.1). In (b), there are two types of gradients with slopes at a constant angular increment of 3°: (i) gradients with slopes between 0° and 45°, where the target concentration (φ_F) is varied and the gradient time is constant ($t_G = 60 + t_D$ min), and (ii) gradients with slopes between 48° and 90°, where φ_F is constant and t_G varies.

The absolute uncertainties are strongly variable and depend on solute retention, with an order of magnitude similar to the *SEP* values (see Table 4.2). For this reason, the results henceforth are represented in terms of relative standard uncertainty in prediction ($s_{R,pred}$), which provides more meaningful and interpretable results, even in the comparison of uncertainties for isocratic and gradient elution. Figure 4.4 shows, as an example, the gradient uncertainty maps for sulphachloropyridazine with the five training designs under study, using the NK model and an increasing integration accuracy in the estimation of $\hat{t}_g$. Two abscissa scales are shown. The upper one corresponds to the angular increments of 3º, yielding to a uniform gradient distribution (see Figure 4.3b). The lower scale shows the equivalent values when the slope is measured as increment in solvent concentration (*v/v*, expressed as volumetric fractions) per minute. Note that this abscissa axis may seem exponential, but it is not.

Typical uncertainty maps for gradient elution show U-patterns, with increments at both extremes (corresponding to the flattest and steepest gradients) and minor errors in between, but not always both extremes are visible. A detailed analysis (including solute polarity, both elution modes and the two retention models, LSS and NK) is done in the next section. A critical point for creating the uncertainty maps is the accuracy level in the calculation of $\hat{t}_g$, which will be referred as Δ. It must be reminded that $\hat{t}_g$ is obtained by numerical methods, working out this variable from Equation (4.10). The magnitude of Δ severely influences the calculation time and affects significantly the accuracy of the derivatives in the Jacobian matrices.
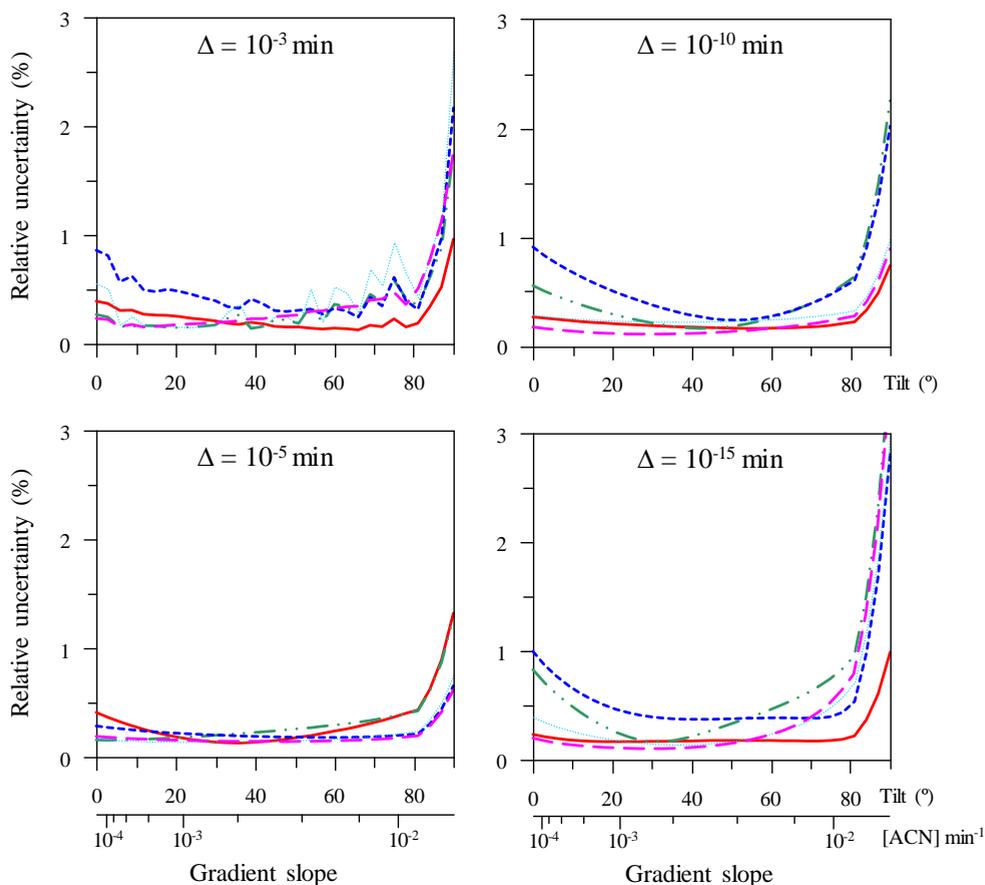
Figure 4.4. Effect of the accuracy (Δ) in the calculation of t_g for sulphachloro-pyridazine. Relative uncertainty ($s_{R,pred}$) maps show the consequences of calculating the Jacobian matrices with progressively more accurate retention times ($\Delta = 10^{-3}$, 10^{-5}, 10^{-10} and 10^{-15} min), for each of the five training designs shown in Figures 4.1 and 4.2. Design identification: ISO1 (continuous red line), ISO2 (long dashed magenta), G1 (dot-dot-dashed green), G2 (short dashed blue), and G3 (dotted cyan).

Figure 4.4 illustrates the consequences of using progressively smaller Δ values. It should be indicated that, for the simulation of chromatograms and resolution estimations, $\Delta = 10^{-3}$ min is perfectly adequate, and can be computed swiftly without the need of root-finding methods [24]. However, as can be observed, the U-pattern at $\Delta = 10^{-3}$ min is very noisy, and the noise is still visible at the $\Delta = 10^{-5}$ level. Beyond $\Delta = 10^{-5}$ min, root-finding methods are needed, and the curves become smoother, although still somewhat biased. Progressively (levels not shown), they become better defined up to remain constant beyond $\Delta = 10^{-13}$. Similar results were observed for the other solutes and designs. Accordingly, for the next studies, $\Delta = 10^{-15}$ min was selected in the evaluation of $\hat{t}_g$.

4.5.4. Evaluation of the designs

We will focus this section on the NK model (Equation (4.2)). For the interpretation of the results, one should take into account that a good determination of a retention model requires:

(i)   large and varied values of retention times (which are related to the magnitude of the interactions with the column, associated to the k_w term in the NK model), and

(ii) varied φ values (dependence with the composition, associated to the B and c terms) provided by the design. Therefore, the participation of each composition along solute migration in gradient designs must be sufficient, and both t_R and φ ranges should be as wide as possible.

Among the 14 solutes initially considered, two of them (sulphaguanidine and sulphanilamide) did not allow a proper calculation, owing to their excessively short retention. For these solutes, the effective number of degrees of freedom in

the fitting to the NK model of the data of each experimental design was insufficient (i.e., the differences in retention were close to 0.1 min and even smaller between related runs), which made the computation of uncertainties less reliable, with the exception of design ISO1. For the success of the proposed methodology, solutes as fast as sulphaguanidine and sulphanilamide would require dedicated designs, including slower eluents and, for this reason, these solutes were excluded in the next discussion. In the case of sulphadiazine, design G1 also led to retention times not sufficiently different (marked with an "a" label in Figure 4.2), which was translated in inflated uncertainties.

4.5.4.1. Isocratic vs. gradient predictions

Figures 4.5 and 4.6 show the expected $s_{R,pred}$ values when each of the five training designs in Figures 4.1 and 4.2 are used to predict isocratic (Figure 4.5) and gradient (Figure 4.6) retention times with the NK model. The plots are sorted by decreasing solute polarity, from sulphadiazine (with the largest polarity) to sulphaquinoxaline (with the lowest; the solute order in Section 4.4.1 follow the sequence of decreasing polarity). A first observation is that, when the performance of a design is inspected in equivalent subplots (i.e., for a given design and solute, isocratic and gradient performance plots), the first point of each uncertainty curve is coincident, because the first composition in the isocratic sampling design (10%, Figure 4.3a) is also the first experiment in the gradient sampling design (whose slope is zero, see Figure 4.3b, gradient 0º).
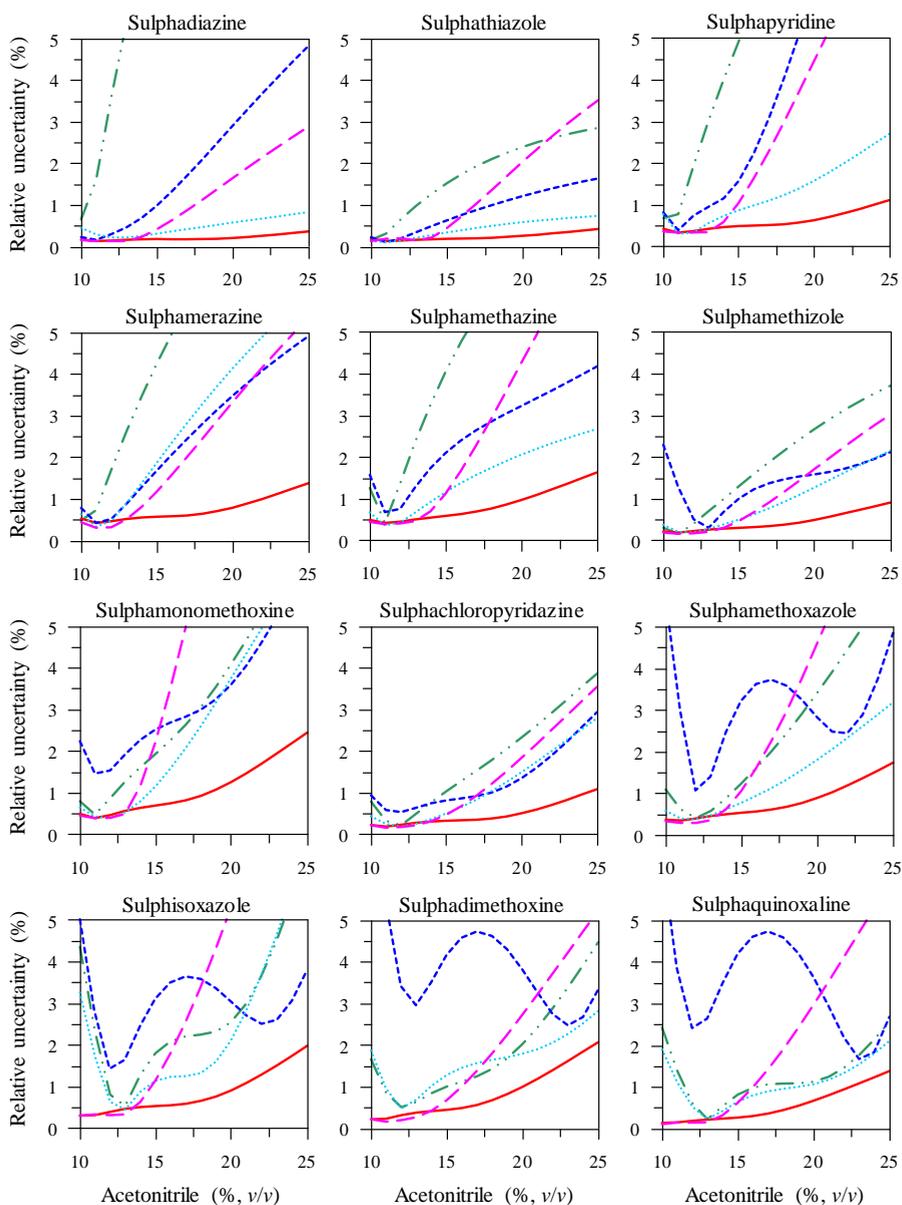
Figure 4.5. Relative uncertainty maps in the prediction of retention times for isocratic experimental conditions ($s_{R,pred}$ as a function of mobile phase composition), using the five training experimental designs (Figures 4.1 and 4.2) and the Neue-Kuss model. See Figure 4.4 for design identification.
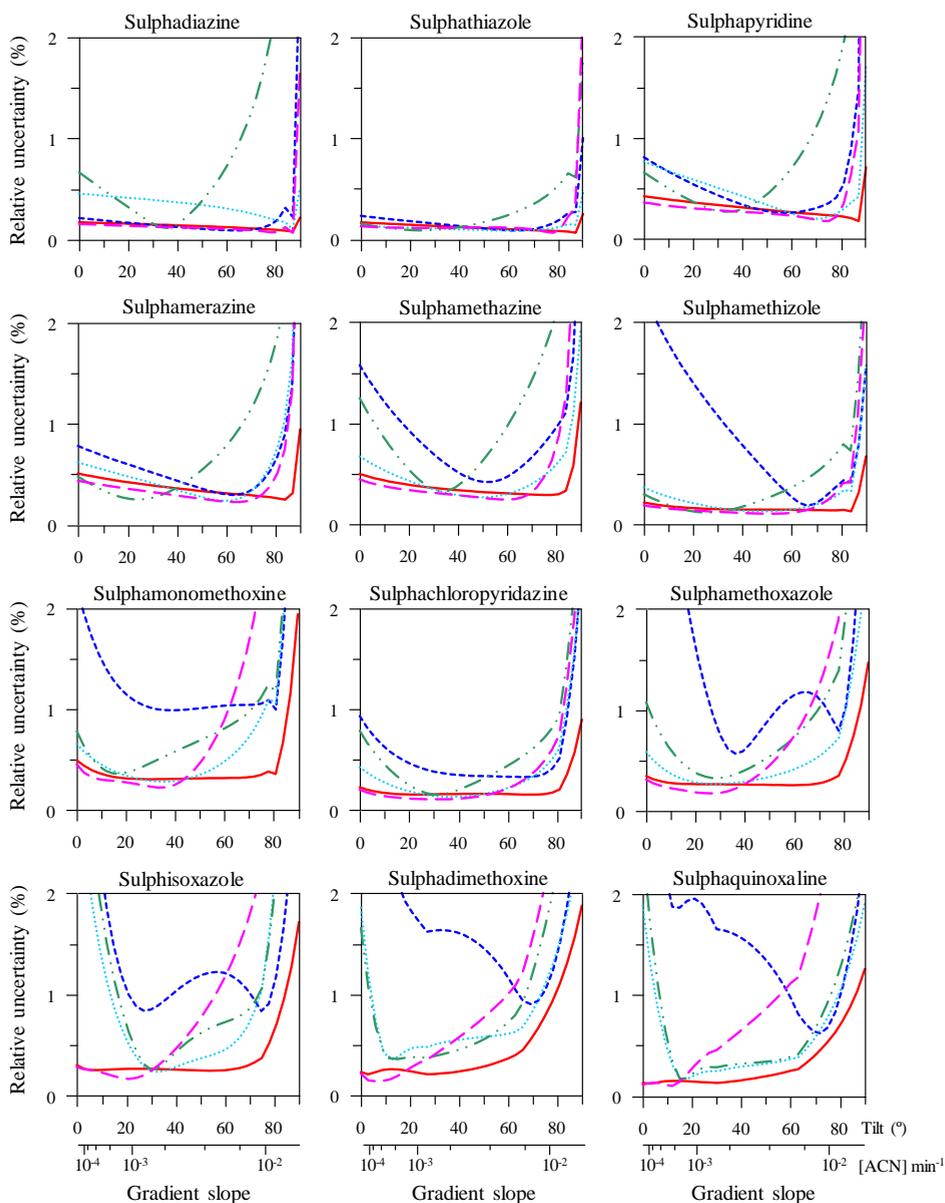
Figure 4.6. Relative uncertainty maps in the prediction of retention times for gradient experimental conditions ($s_{R,pred}$ as a function of the gradient slope), using the five training experimental designs (Figures 4.1 and 4.2) and the Neue-Kuss model. See Figure 4.4 for design identification.

Moreover, the $s_{R,pred}$ value obtained with the highest composition in the isocratic sampling design (25% acetonitrile) again tends to coincide with the equivalent error of the gradient with the strongest slope. The reason is that the 90º gradient is close to an isocratic mobile phase containing 25% acetonitrile (see again Figure 4.3b) for sufficiently retained solutes. In this gradient, the effect of the 10% acetonitrile composition along the isocratic step associated to the dwell time is negligible for retained enough solutes.

In the intermediate regions within the plots, $s_{R,pred}$ varies in a specific way, depending on the solute and training design. In spite of this variability, the curves change systematically as the solute polarity decreases, with a consistence that gives support to both the calculations and the results. The $s_{R,pred}$ values are generally larger when isocratic conditions are predicted (note that the scale in the isocratic and gradient plots is not the same). The magnitude of the minimal uncertainty in equivalent isocratic and gradient plots tends to be similar, but it deteriorates quicker in isocratic scans (even when the training design is adequate). Gradients are, thus, predicted with intrinsically smaller uncertainties for any training experimental design.

4.5.4.2. Isocratic predictions

Generally speaking and for the NK model, design ISO1 is the best for isocratic predictions (lower curves in Figure 4.5), with a maximal $s_{R,pred}$ value of around 2%. Moreover, considering the dependence with the solvent composition along the scan, ISO1 is the least affected among all designs, with uniform prediction uncertainties in the whole domain.

As could be expected, design ISO2 presents similar uncertainties to design ISO1 but only with the slowest eluents. Its performance decays abruptly when the predictions are carried out exceeding the covered domain (10–14%

acetonitrile). Beyond this range and for slow or intermediate solutes (which imply extrapolations), design ISO2 shows a performance similar to the gradient training designs (G1 to G3), although with a more intense deterioration, reflecting the absence of information about the faster compositions. For the fastest solutes, gradient designs G1 and G2 perform even worse.

Design G2 (set of gradients at variable t_G) performs better than design G1 (variable φ_F) for the fastest solutes (sulphadiazine to sulphamonomethoxine). Observe that these solutes elute in wider concentration ranges with design G2 (see "b" label in Figure 4.2), which implies that the fittings are benefitted from the richer information about the highest solvent compositions with regard to design G1. This situation is reversed for the most hydrophobic solutes, since design G1 allows that slower eluents participate in solute retention in a larger extent, before the gradients reach higher eluent concentrations. For this reason, design G1 is more informative for the five most retained solutes (compare the curves for designs G1 and G2 in Figure 4.5). With regard to ISO1, both designs G1 and G2 are insufficiently informative for the fastest eluents (even worse in the case of design G1): the more polar the solute, the more severe the situation.

Finally, it can be observed that design G3, proposed for gathering the good qualities of isocratic and gradient designs, provides indeed a reasonably good performance for all kinds of solutes. Accordingly, it is the most recommended configuration for modelling, with the exception of design ISO1.

4.5.4.3. Gradient predictions

For gradient predictions (Figure 4.6), the superiority of design ISO1 is identically noteworthy. Design ISO2 is, as could be expected, even slightly better than ISO1 in the prediction of gradients with small slopes. This advantage is lost as the gradient slope increases, and the loss in predictive capability becomes

more important for the most hydrophobic solutes, where the weight of the highest eluent compositions is larger. This is translated in a progressive shift of the raise in uncertainty towards lower solvent concentrations.

Taking into account the parallelism between the gradient sampling design (Figure 4.3b), and the training designs G1 and G2 depicted in Figure 4.2, it is not surprising that design G1 performs better at smaller slopes for all solutes, whereas design G2 is better for the strongest slopes. The best gradient training design in the prediction of gradients is again design G3, only surpassed by design ISO1. The magnitude of $s_{R,pred}$ for design G3 is below 1% for intermediate gradients and it is generally as good as 2%, except for unfavourable cases.

4.5.5. Neue-Kuss vs. Linear Solvent Strength model

This section compares the capability of both retention models, in terms of predictive error. The LSS model presents an important drawback: the existence of lack of fit (LOF). Figure 4.7 represents the relative error in the prediction of retention times (see definition for RE in the caption of Figure 4.7), for all solutes as a function of the eluent composition ($RE = F(\varphi)$).In the absence of LOF, these plots should exhibit random patterns. However, as can be observed, the curves for the LSS model (Figure 4.7a), far from showing that expected randomness (well appreciable in the NK curves, Figure 4.7b), present remarkable correlations with φ, denoting an important LOF component in the uncertainty assessments. The presence of LOF confers to the results less reliability, and the real errors can be larger than those expected, with a different pattern. It should be noted that the error propagation methodology does not include the consequences of the LOF, and drawing conclusions beyond general terms (e.g., uncertainty magnitude or design order) is not prudent.
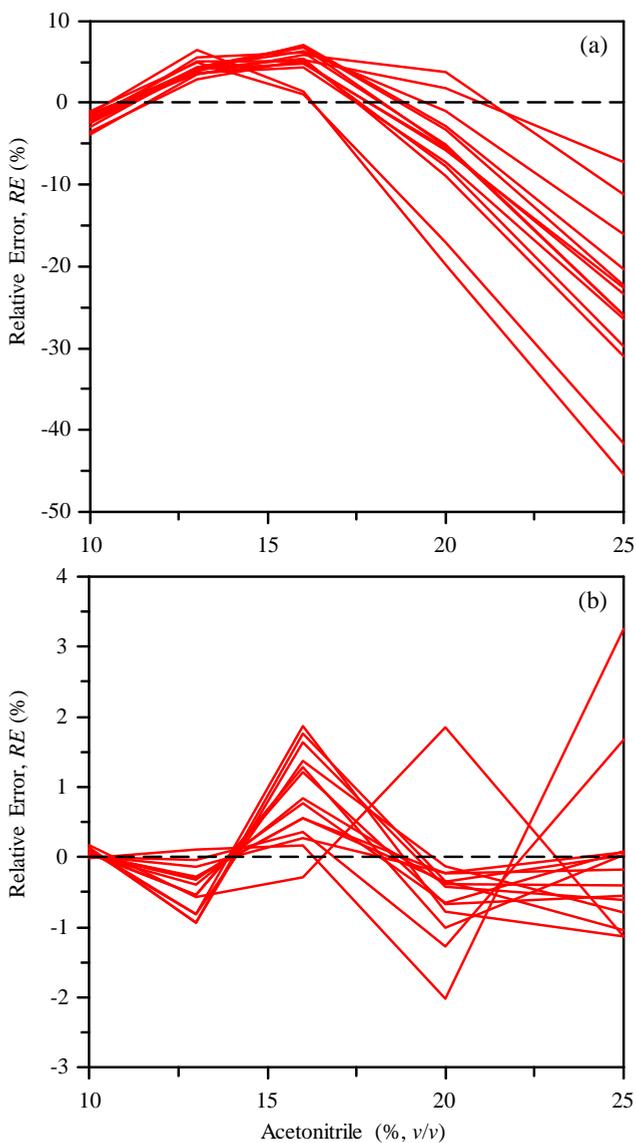
Figure 4.7. Lack of fit plot, evidencing possible correlations between the relative prediction error and the solvent composition (φ), when the fitting is carried out with: (a) the LSS and (b) NK models. The relative errors are calculated as $RE = \dfrac{\hat{t}_g - t_{g,\exp}}{t_{g,\exp}} \times 100$, for each composition and solute.

The simplicity of the LSS model is translated in simpler relative uncertainty maps, with smooth variations that apparently give rise to a maximum in isocratic predictions (Figure 4.8), and a minimum in gradient predictions (Figure 4.9). These variations, however, can be calculation artifacts, consequences of the LOF. Nevertheless, it can be concluded that, globally, the uncertainties are higher for the LSS model than for the NK model (compare Figures 4.8 and 4.9 with Figures 4.5 and 4.6).

Along the sequence of decreasing solute polarity, the relative performance of design ISO1, with regard to the other designs, is strikingly different in both retention models. Thus, with the LSS model, design ISO1 begins offering worse performance than the gradient training sets (G1 to G3) for the four fastest solutes; it presents an intermediate performance (similar for the three gradient designs) for the next five solutes; and becomes the best design for the remaining three solutes, which have the lowest polarity. This trend can identically be observed in both isocratic and gradient predictions (see Figures 4.8 and 4.9). Remember that design ISO1 was the best design for the NK model in all instances.
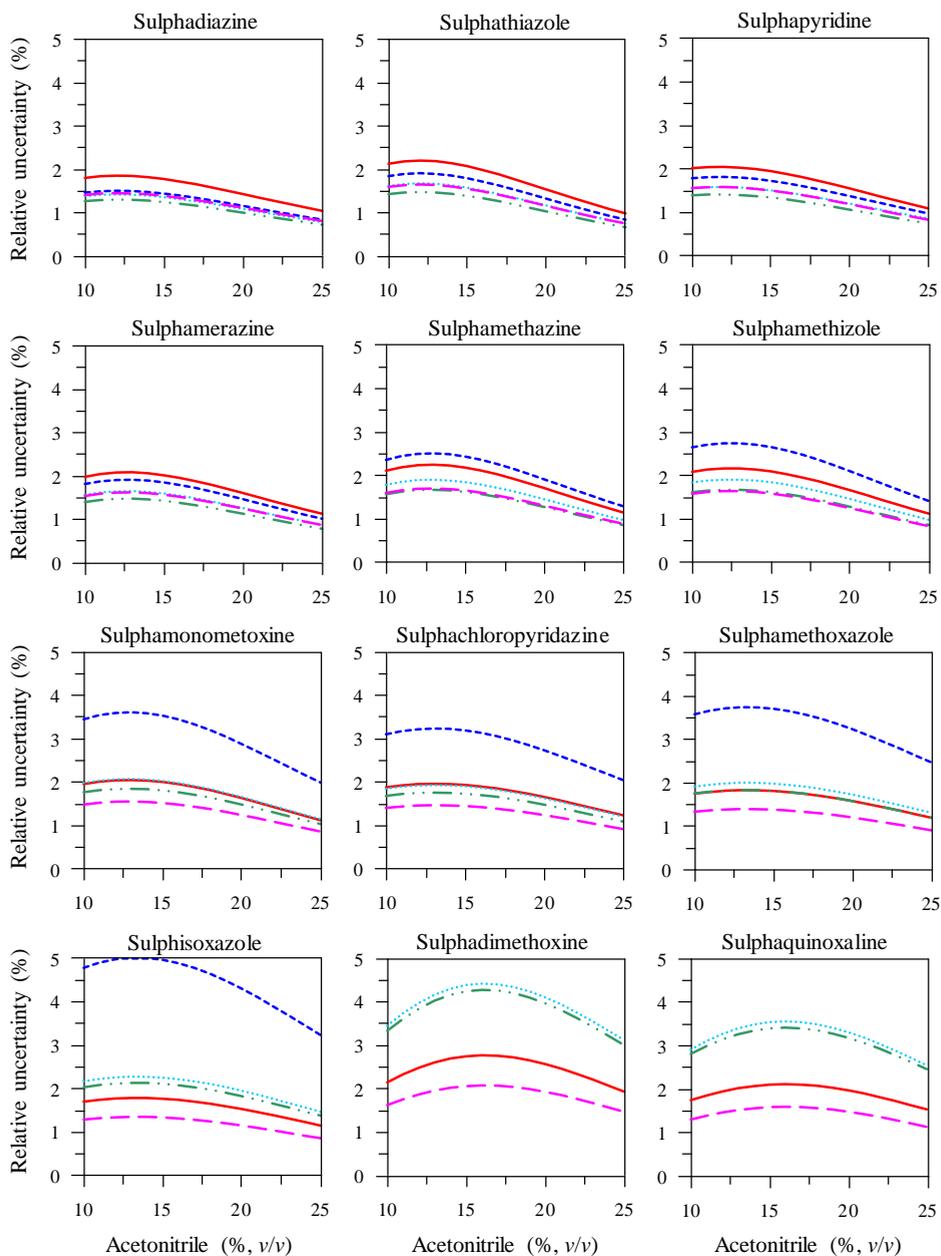
Figure 4.8. Relative uncertainty maps in the prediction of retention times for isocratic experimental conditions, using the five training experimental designs and the LSS model. See Figure 4.4 for design identification.
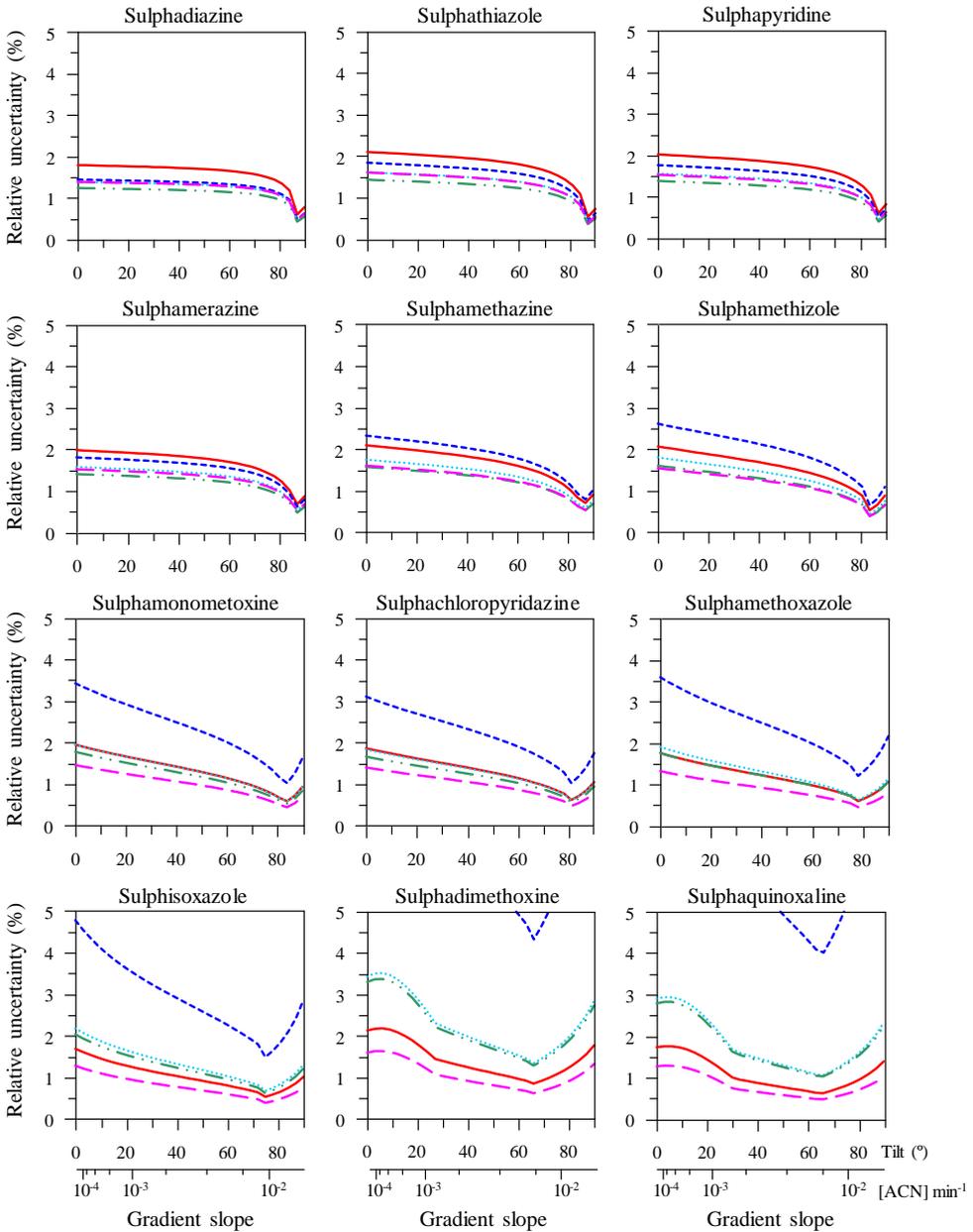
Figure 4.9. Relative uncertainty maps in the prediction of retention times for gradient experimental conditions, using the five training experimental designs and the LSS model. See Figure 4.4 for design identification.

The anomalous behaviour observed for the LSS model is a consequence of the LOF and can be explained by analysing the results of design ISO2 in Figures 4.8 and 4.9, where the uncertainties are comparable (or better) to those found with the three gradient designs (G1 to G3) in all cases. The ISO1 design, in contrast, performs worse than the three gradient designs for fast solutes. In fact, accurate predictions for the fastest solutes require sampling slower solvent compositions more exhaustively, and when this is done (design ISO2), the uncertainties are effectively smaller. Indeed, the narrower concentration range of design ISO2 makes the effects of LOF less visible in the sampled range (10–14% acetonitrile). Note that design ISO2 does not provide information on compositions of high elution strength (15–25% acetonitrile). Therefore, the worse performance of design ISO1 must be attributed to the LOF. Also, the predicted uncertainty at higher organic solvent contents is not safe.

Finally, among the three gradient designs, design G2 is the one offering the worst performance, and design G3 becomes progressively equivalent to design G1 (which is the best of the three) along the polarity sequence. These results are consistent with those found in Sections 4.5.4.2 and 4.5.4.3.

Along the study, the different uncertainty curves associated to each design and solute were interpreted, based on three main principles:

  (i)   each composition along a gradient design must participate significantly in the solute retention,
 (ii)   the compositions scanned along the solute elution must be wide, and
(iii)   the retention times found along a given design must reflect that diversity.

For the two most polar solutes (sulphaguanidine and sulphanilamide), which were scarcely retained for most training designs, these premises were not fulfilled.

4.6. Conclusions

This work introduces a methodology oriented to study comprehensively the performance of arbitrary experimental designs (e.g., single or multi-isocratic, linear or multi-linear gradients, mixed isocratic-gradient designs, or any other), based on the determination of relative uncertainties in the prediction of retention ($s_{R,pred}$). The methodology is tested with five common experimental designs, shown in Figures 4.1 and 4.2 (ISO1, ISO2, G1, G2 and G3), and 14 sulphona-mides of diverse polarity, considering two common retention models: the Linear Solvent Strength (LSS) and Neue-Kuss (NK) equations. The NK model was more accurate and did not show evidences of lack of fit (LOF).

The evaluation of the predictive performance of a design is based on the systematic calculation of the uncertainties associated to the estimation of retention times, when a regular distribution of isocratic and gradient expe-rimental conditions (i.e., scanning or sampling designs) are predicted from the design used for modelling (i.e., training design). The treatment requires the cal-culation of Jacobian matrices associated to the training and sampling designs. The Jacobian matrices can be calculated algebraically in designs involving the LSS model and isocratic runs or single linear gradients. This calculation is not practical with the NK model in gradient elution, even for the simplest gradients. These limitations forced to carry out the evaluation of the Jacobian matrices by numerical procedures. The results are represented in terms of $s_{R,pred}$, which provides more meaningful and interpretable results than the absolute uncertainties (s_{pred} in Equation (4.19)), even in the comparison of isocratic and gradient experiments. The observed systematic changes in the uncertainty plots, as the solute polarity decreases, give support to the reliability of the calculations, as well as the results themselves. Relative uncertainty maps are valid in the

absence of LOF. Otherwise, the results should be analysed with caution, and only general conclusions may be derived.

Isocratic training designs covering the whole solvent domain (ISO1), which are widely accepted as maximally informative, are confirmed as the best in both isocratic and gradient predictions, when the equation is unbiased (NK model). However, in the presence of LOF (LSS model), design ISO1 performs worse than gradient designs for the fastest solutes. A training design focused to lower elution strength mobile phases (ISO2) gives rise to similar uncertainties to design ISO1 in the well sampled region. However, it yields larger increments in the prediction error than gradient training designs (G1 to G3) when predictions imply extrapolations. An interesting feature is that the better performance of design ISO1 with the NK model is not observed with the LSS model for the fastest solutes. Within the narrower ranges covered by design ISO2, there is no appreciable LOF, and the predictions in that domain would be correct. The wider solvent ranges in ISO1 give rise to LOF and an abnormal inversion in the curve order.

Among the gradient training designs, design G2 (t_G variable) performed better for faster solutes than design G1 (φ_F variable). This situation is reversed for solutes with lower polarity, because design G1 allows the compositions with smaller elution strength participate in solute migration in a larger extent, which is translated in richer information on the effect of these compositions. Both gradient designs (G1 and G2) performed worse than design ISO1 for the compositions with higher organic solvent content, because these participate in the design in a lower extent. A third gradient design was proposed (G3) to gather the advantages and solve some of the pitfalls of designs ISO1, G1 and G2. Design G3 can be considered as an extension of the other two gradient designs at the start of the gradient, where there is lack of information for fast solutes. Therefore, it represents a balance between the isocratic and gradient designs for fast solutes.

As expected, design G3 was found to offer comparable performance to design G2 for the fastest solutes and to design G1 for the slowest ones. In the case of the LSS model, the differences were less relevant, although more noticeable for the fastest solutes.

As a general conclusion, gradients are predicted with intrinsically smaller uncertainties, independently of the training experimental design. In addition, gradients are more insensitive than isocratic predictions with regard to the type of training design used. The minimal error is comparable in magnitude, but it is less affected by changes in the scanning variable when gradients are predicted. In contrast, in isocratic sampling designs, the uncertainty worsens quickly when the conditions are departed from the compositions with minimal error. As a consequence of the smaller errors, and larger extension of the minimal error region, the characteristic U-pattern is more visible in the uncertainty maps in gradient predictions than in isocratic ones. All these considerations explain the abnormally good performance of gradient predictions, even with biased models. Anyway, the magnitude of the uncertainties is larger for the LSS model than for the NK model.

In further work, the methodology will be applied to the evaluation of multiple designs, including families of related and mixed designs, by investigating the effects in the predictive performance when the designs are varied systematically (e.g., slopes, distributions, number of experiments). A system to calculate the optimal experimental designs according to a certain pattern will be developed as well. The final aim is finding general gradient patterns optimally informative.

4.7. References

[1] S.N. Deming, S.L. Morgan, Experimental Design: A Chemometric Approach, in *Data Handling in Science and Technology*, Vol. 11, Elsevier Science, Amsterdam, 2nd ed.,1993.

[2] D.B. Hibbert, Experimental design in chromatography: A tutorial review, *J. Chromatogr. B* 910 (2012) 2–13.

[3] E. Rozet, P. Lebrun, P. Hubert, Design spaces for analytical methods, *Trends Anal. Chem.* 42 (2013) 157–167.

[4] G. Vishnu Vardhana Rao, V. Srinivasa Rao, K. Ramakrishna, Experimental design based development of a simple and robust RP-HPLC technique for the related substances of dapoxetine hydrochloride in bulk drugs, *J. Chem. Pharm. Res.* 7 (2015) 639–646.

[5] L.K. Shekhawat, A. Godara, V. Kumar, A.S. Rathore, Design of experiments applications in bioprocessing: Chromatography process development using split design of experiments, *Biotechnol. Prog.* 35 (2018) e2730 (doi:10.1002/btpr.2758, Wiley Online Library).

[6] C. Muscat Galea, D. Didion, D. Clicq, D. Mangelings, Y. Vander Heyden, Method optimization for drug impurity profiling in supercritical fluid chromatography: Application to a pharmaceutical mixture, *J. Chromatogr. A* 1526 (2017) 128–136.

[7] A.P. Schellinger, P.W. Carr, Isocratic and gradient elution chromatography: A comparison in terms of speed, retention reproducibility and quantitation, *J. Chromatogr. A* 1109 (2006) 253–266.

[8] L.R. Snyder, J.J. Kirkland, J.L. Glajch, *Practical HPLC Method Development*, John Wiley & Sons, New York, 2nd ed., 1997.

[9]   R. Carlson, J.E. Carlson, Design and Optimization in Organic Synthesis, in *Data Handling in Science and Technology*, Vol. 24, Elsevier Science, New York, 1992.

[10] P.K. Sahu, N.R. Ramisetti, T. Cecchi, S. Swain, C. Patroa, J. Panda, An overview of experimental designs in HPLC method development and validation, *J. Pharm. Biomed. Anal.* 147 (2018) 590–611.

[11] Y.B. Ji, Q.S. Xu, Y.Z. Hu, Y. Vander Heyden, Development, optimization and validation of a fingerprint of Ginkgo biloba extracts by high-performance liquid chromatography, *J. Chromatogr. A* 1066 (2005) 97–104.

[12] W. Li, H.T. Rasmussen, Strategy for developing and optimizing liquid chromatography methods in pharmaceutical development using computer-assisted screening and Plackett-Burman experimental design, *J. Chromatogr. A* 1016 (2003) 165–180.

[13] T. Sivakumar, R. Manavalan, C. Muralidharan, K. Valliappan, Multi-criteria decision making approach and experimental design as chemo-metric tools to optimize HPLC separation of domperidone and pantoprazole, *J. Pharm. Biomed. Anal.* 43 (2007) 1842–1848.

[14] V.R. de Almeida Borges, A.F. Ribeiro, C. de Souza Anselmo, L.M. Cabral, V.P. de Sousa, Development of a high performance liquid chromatography method for quantification of isomers β-caryophyllene and α-humulene in copaiba oleoresin using the Box-Behnken design, *J. Chromatogr. B* 940 (2013) 35–41.

[15] M.J. Cózar Bernal, A.M. Rabasco, M.L. González Rodríguez, Develop-ment and validation of a high performance chromatographic method for determining sumatriptan in niosomes, *J. Pharm. Biomed. Anal.* 72 (2013) 251–260.

[16] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. DeJong, P.J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics, in *Data Handling in Science and Technology*, Vol. 20B, Elsevier Science, Amsterdam, 1997.

[17] P.F. de Aguiar, B. Bourguignon, D.L. Massart, Comparison of models and designs for optimisation of the pH and solvent strength in HPLC, *Anal. Chim. Acta* 356 (1997) 7–17.

[18] J.R. Torres Lapasió, S. Pous Torres, J.J. Baeza Baeza, M.C. García Álvarez-Coque, Optimal experimental designs in RPLC at variable solvent content and pH based on prediction error surfaces, *Anal. Bioanal. Chem.* 400 (2011) 1217–1230.

[19] G. Vivó Truyols, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Error analysis and performance of different retention models in the transference of data from/to isocratic/gradient elution, *J. Chromatogr. A* 1018 (2003) 169–181.

[20] G. Vivó Truyols, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Estimation of significant solvent concentration ranges and its application to the enhancement of the accuracy of gradient predictions, *J. Chromatogr. A* 1057 (2004) 31–39.

[21] J.A. Navarro Huerta, A. Gisbert Alonso, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Benefits of solvent concentration pulses in retention time modelling of liquid chromatography, *J. Chromatogr. A* 1597 (2019) 76–88.

[22] U.D. Neue, H.J. Kuss, Improved reversed-phase gradient retention modeling, *J. Chromatogr. A* 1217 (2010) 3794–3803.

[23] P. Nikitas, A. Pappa-Louisi, Expressions of the fundamental equation of gradient elution and a numerical solution of these equations under any gradient profile, *Anal. Chem.* 77 (2005) 5670–5677.

[24] S. López Ureña, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Enhancement in the computation of gradient retention times in liquid chromatography using root-finding methods, *J. Chromatogr. A* 1600 (2019) 137–147.

[25] W.H. Press, B.P. Flannery, S. Teukolsky, W.T. Vetterling, *Numerical Recipes in FORTRAN 77: The Art of Scientific Computing*, Cambridge University Press, New York, 1992.

[26] C.J.F. Ridders, Accurate computation of F'(x) and F'(x)F''(x), *Adv. Eng. Softw.* 4 (1978) 75–76.

[27] J.R. Torres Lapasió, M.C. García Álvarez-Coque, J.J. Baeza Baeza, Global treatment of chromatographic data with MICHROM, *Anal. Chim. Acta* 348 (1997) 187–196.

CHAPTER 5

ESTIMATION OF PEAK CAPACITY
BASED ON PEAK SIMULATION

5.1. Abstract

Peak capacity (P_C) is a key concept in chromatographic analysis, nowadays of great importance for characterising complex separations as a criterion to find the most promising conditions. A theoretical expression for P_C estimation can be easily deduced in isocratic elution, provided that the column plate count is assumed constant for all analytes. In gradient elution, the complex dependence of peak width with the gradient program implies that an integral equation has to be solved, which is only possible in a limited number of situations. In 2005, Uwe Neue developed a comprehensive theory for the calculation of P_C in gradient elution, which is only valid for certain situations: single linear gradients, absence of delays and extra-column effects, Gaussian peaks and constant plate count. Going beyond these limitations implies resolving algebraic expressions that unfortunately cannot be integrated. In this work, P_C is predicted for multiple situations based on peak simulation. The approach is more general and can be applied for situations out of the scope of the Neue outline, such as complex multi-linear gradients, including asymmetrical peaks. The plots of P_C versus retention time of the last eluted solute give rise to Pareto fronts, and can be useful for the probabilistic enhancement of peak resolution in situations where complex multi-analyte samples are processed.

5.2. Introduction

Peak capacity (P_C) is a theoretical key concept in chromatographic analysis, defined as the maximal number of peaks that ideally can be completely resolved in a pre-established time window (i.e., between specific peaks or within a given time range), under specific experimental conditions. On the one hand, this parameter allows establishing whether a given separation system will be able to separate a sample of certain complexity. On the other, it allows discriminating the experimental conditions that offer more chances to the separation, since the total number of observed peaks in a complex mixture increases with the P_C value. The concept was first developed for isocratic elution [1–5]. Later it attracted attention in gradient elution, especially for samples that contain many components where achieving complete resolution is problematic [6–8]. The estimation of P_C has been described in detail by several authors, up to recent time [9–12]. P_C is nowadays of great importance for measuring the potential of specialised separation modes, such as two-dimensional liquid chromatography [13–17].

In liquid chromatography, chromatograms tend to have uneven peak distributions, with overlapped peaks and large gaps [8,18–25]. The condition of consecutive location of slightly overlapped peaks for the compounds in a complex sample hardly happens in practice. Some examples appear in Refs. [26–28]. Therefore, P_C rarely can be strictly measured from experimental chromatograms showing regular, consecutively overlapped peaks. On the other hand, the approaches reported in the literature to estimate P_C assume ideal peaks, and peak widths or efficiencies are often considered unchanged with retention time. The outlines proposed by Giddings [2] and Grushka [3] for isocratic elution considered changes in peak width with retention time, provided that the column plate count is assumed constant for all analytes, which

is unreal in practice even for related compounds. If a bandwidth of four standard deviation units (4σ) is assumed for the peak boundaries, P_C can be defined as [6]:

$$P_C = 1 + \int_{t_0}^{t_R} \frac{dt}{4\sigma(t)} \tag{5.1}$$

where t_0 and t_R are the dead and retention times, respectively, and $\sigma(t)$ is a function describing the dependence of the standard deviation with time along the elution program. In isocratic elution, Equation (5.1) gives rise to:

$$P_C = 1 + \frac{\sqrt{N}}{4} \ln \frac{t_2}{t_1} \tag{5.2}$$

N being the system plate number, and t_1 and t_2 the retention time for the first and last peaks that define the selected window, respectively. In previous work, a modification of the Giddings' and Grushka's equations for isocratic elution, taking into account changes in the efficiency and asymmetries, was proposed [10].

The adaptation of Equation (5.1) to gradient elution is more complex. Finding an analytical solution is only possible in very limited cases, since it implies the need of including the gradient retention factor (or time) inside the integral. A comprehensive theory for P_C calculation for linear gradients was developed by Neue [6,9,29]. The reported final expression is:

$$P_C = 1 + \frac{\sqrt{N}}{4} \frac{1}{G+1} \ln \left(e^{S\Delta c} \frac{G+1}{G} - \frac{1}{G} \right) \tag{5.3}$$

being

$$G = S\,m\,t_0 \tag{5.4}$$

where S is the elution strength (i.e., the slope of the relationship between the logarithm of the retention factor and the solvent content, in the linear solvent strength theory, LSST), m is the slope of a linear gradient, and Δc is the change in solvent concentration along the ramp at t_G (the time at which the ramp is finished, namely, the gradient time). Equations (5.3) and (5.4) allow anticipating easily P_C *in silico*.

The Neue treatment presents some limitations: it is only valid for single linear gradients, it does not consider extra-column contributions, nor delays associated to the tubing (i.e., the dwell time), it assumes that the plate count is uniform for all analytes, and does not take into account the asymmetry of chromatographic peaks. Another important limitation is that the underlying retention models must be expressed as the product $k(t) = k^* f(t)$, $k(t)$ being the retention factor in gradient elution, k^* the retention factor at the start of the linear gradient, and $f(t)$ a function of the dependence of the retention with the solvent composition, combined with the change in solvent composition with time along the gradient. Going beyond these limitations implies resolving algebraic expressions that unfortunately cannot be integrated. Similarly to Neue, Snyder and Dolan proposed an equation to calculate peak capacity, based on the LSST, which includes the gradient compression factor [30,31]. This equation is valid for linear gradients, and considers a constant efficiency and symmetrical peaks.

For more complex gradients, P_C should be approximated in a given elution window, by dividing the window size Δt by the average peak width ($\overline{w}$) in that domain (i.e., size of the retention time window measured in peak width units). Considering a elution window between the retention times for an unretained compound (t_0) and the last eluted solute ($t_{R,last}$) [8,32,33]:

$$P_C = 1 + \frac{t_{R,last} - t_0}{\overline{w}} \tag{5.5}$$

Equation (5.5) will obviously offer large uncertainty, owing to the disparity in peak widths. More accurate estimations can be achieved by splitting the chromatogram in regions between consecutive peaks along the elution window [31]:

$$P_C = 1 + 2 \sum_{i=1}^{n-1} \frac{t_{R,i+1} - t_{R,i}}{w_{i+1} + w_i} \tag{5.6}$$

where i and $i+1$ refer to the consecutive peaks.

In this work, P_C is predicted based on chromatographic peak simulation using the retention and peak profile properties of a set of related compounds. Most tools for prediction and simulation are the same as those used in the interpretive optimisation of separation conditions in liquid chromatography, which have been demonstrated to offer excellent accuracy. The approach is valid for a variety of situations, including existence of extra-column effects and the application to multi-linear gradients. It allows obtaining accurate measurements of P_C in different time windows, even when there are variations in solute properties along the chromatogram and the peaks are asymmetrical. It is thus possible to make predictions of P_C in cases where Equation (5.1) has no algebraic solution. To illustrate the approach, a case of study is used consisting of a set of 15 sulphonamides analysed with three columns (C18, phenyl and cyano), eluted using isocratic mobile phases, and linear and multi-linear gradients.

5.3. Theory

The proposed approach to estimate P_C is based on the simulation of chromatograms containing sequential peaks with a small amount of controlled overlap. In order to build the ideal sequence of consecutive peaks from which estimate P_C, the influence of the gradient program and polarity of solutes on the peak profiles along the chromatogram should be considered. The methodology used in this work to simulate chromatograms, in reversed-phase liquid chromatography (RPLC), is described below. For both retention times and peak profiles, a reduced set of experimental data is used to build models, from which predict their variation under new conditions. The details referred to the numerical prediction of P_C are given in Section 5.5.

5.3.1. Prediction of retention

In order to make a more general approach, peak simulation in either isocratic or gradient elution (linear and multi-linear) was carried out using the following equation that relates the retention with the elution program:

$$t_0 - t_{ext} = \int_0^{t_R - t_0} \frac{dt}{k(\varphi(t))} \tag{5.7}$$

where t_{ext} is the time needed for solutes to go through the external tubes and connections, k the retention factor, and $\varphi(t)$ the volume fraction of organic solvent, which is fixed in isocratic elution and changes with time in gradient elution. The time at which the solute reaches the column outlet (i.e., the retention time, t_R) corresponds to the instant at which the summation of terms in Equation (5.7) matches $t_0 - t_{ext}$ [34]. The retention factor was calculated as:

$$k = \frac{t_R - t_0}{t_0 - t_{ext}} \tag{5.8}$$

A logarithmic quadratic retention model, based on the linear model proposed by Bosch and Rosés [35,36], was used to predict the retention:

$$\log k = q + S_1\, P_M^N + S_2\, (P_M^N)^2 \tag{5.9}$$

where P_M^N measures the polarity of the mobile phase. For acetonitrile:

$$P_M^N = 1.00 - \frac{2.13\varphi}{1+1.42\varphi} \tag{5.10}$$

In Equation (5.9), q depends on the type of column, S_1 is related to the elution strength of the solvent, and S_2 accounts for the deviations from the linear behaviour for sufficiently large elution ranges. For the three parameters (q, S_1 and S_2), specific values are needed for describing the retention of each solute in a sample.

Only in very specific cases, does Equation (5.7) have analytical solution, such is the case when both the gradient program and logarithmic retention model are linear. In this work, the retention time was found numerically, outlining the problem by splitting the integral as follows [37]:

$$t_0 - t_{ext} = \int_0^{t_D} \frac{dt}{k_0} + \int_{t_D}^{t_g - t_0} \frac{dt}{k(\varphi(t))} = \frac{t_D}{k_0} + \int_{t_D}^{t_1 - t_{01}} \frac{dt}{k(\varphi(t))} +$$

$$+ \int_{t_1 - t_{01}}^{t_2 - t_{02}} \frac{dt}{k(\varphi(t))} + ... + \int_{t_{i-1} - t_{0i-1}}^{t_i - t_{0i}} \frac{dt}{k(\varphi(t))} + \int_{t_i - t_{0i}}^{t_{i+1} - t_0} \frac{dt}{k(\varphi(t))} \tag{5.11}$$

Each term in Equation (5.11) corresponds to a column fraction through which the solute migrates during a given time interval. In gradient elution, the solute migrates isocratically during a certain time (i.e., dwell time, t_D), before

the gradient front reaches the solute. This first step gives rise to the first term in Equation (5.11). The other terms take into account the progressively longer delay needed for the gradient front to reach the solute as it migrates along the column. If the steps in Equation (5.11) are short enough, the elution will be isocratic along each step, giving rise to:

$$
\begin{aligned}
t_0 - t_{\text{ext}} &\approx \frac{t_D}{k(\varphi(t_0))} + \frac{t_1 - t_D}{k(\varphi(t_1 - \tau_1))} + \frac{t_2 - t_1}{k(\varphi(t_2 - \tau_2))} + \dots + \frac{t_{i+1} - t_i}{k(\varphi(t_{i+1} - \tau_{i+1}))} = \\
&= \frac{t_D}{k(\varphi(t_0))} + \frac{\Delta t}{k(\varphi(t_1 - \tau_1))} + \frac{\Delta t}{k(\varphi(t_2 - \tau_2))} + \dots + \frac{\Delta t}{k(\varphi(t_{i+1} - \tau_{i+1}))}
\end{aligned} \tag{5.12}
$$

where the τ parameters consider the progressive delays along solute elution (see Chapter 3). In this work, the time size for each step was taken as 0.01 min. This criterion is arbitrary and assures an accuracy of around 1–2 s, comparable to the accuracy in peak location. Other details can be found elsewhere [37].

5.3.2. Width modelling and peak simulation

Peak simulation was carried out considering the peak profiles. This section describes the approach used to predict the width and asymmetry of normalised chromatographic peaks, based on the left (A) and right (B) half-widths (i.e., width in the time dimension at each side of the peak apex from the centre to the preceding and following part of the peak, respectively) measured at 10% peak height. The models used to predict the half-widths in isocratic elution describe parabolic trends with a gentle curvature [38]:

$$
A = a_0 + a_1 t_R + a_{11} t_R^2 \tag{5.13}
$$

$$
B = b_0 + b_1 t_R + b_{11} t_R^2 \tag{5.14}
$$

It should be noted that for long elution times, the trends of global half-width versus retention time may give rise to artifacts. For gradient elution, the half-widths can be predicted with enough accuracy based on the Jandera's approximation [39], which is valid for linear and multi-linear gradients showing smooth transitions. The Jandera's approximation states that, in gradient elution, the peak width of a given solute is approximately the same the solute would have if it migrated isocratically at the instant composition when it leaves the column during the gradient. In this work, these isocratic retention times were obtained with Equations (5.11) and (5.12), as a function of the organic solvent content.

To simulate asymmetrical peaks in isocratic and gradient elution, a handy modification of the Gaussian model was used [40]:

$$h(t) = h_0 \exp\left[-\frac{1}{2} \left(\frac{t - t_R}{s_0 + s_1(t - t_R)} \right)^2 \right] \qquad (5.15)$$

where $h(t)$ and h_0 are the height at time t and the maximal peak height, respectively, s_0 is a measurement of the peak width established on a Gaussian basis, and s_1 and higher order terms account for peak distortion. The coefficients of the linear function can be easily calculated from the predicted values of half-widths as follows:

$$s_0 = 0.932 \frac{A_{0.1} B_{0.1}}{A_{0.1} + B_{0.1}} \qquad (5.16)$$

$$s_1 = 0.466 \frac{B_{0.1} - A_{0.1}}{A_{0.1} + B_{0.1}} \qquad (5.17)$$

Equation (5.15) can fit almost any peak, from tailing to fronting. However, the function does not work properly when the polynomial takes zero or negative values. Also, after reaching a minimum value, the predicted signal may grow outside the peak region. This is especially troublesome for the prediction of chromatograms, where the signals of individual peaks separated in time should be added to give composite signals, as is the case of study. The artefacts are more conspicuous for strong asymmetrical signals ($B/A > 2.5$) and for simulations involving long time windows. To solve this problem, the baseline at both sides of the peak should be properly restored. For this purpose, we have used a mixed exponential-PMG1 function [41], where the outer peak regions of the modified Gaussian model are replaced by exponential decays at both sides of the peak:

$$h = k_{1,\text{left}} \exp\{k_{2,\text{left}} (t - t_R)\} \qquad \text{for } t < t_R - A_{0.1} \tag{5.18}$$

$$h = k_{1,\text{right}} \exp\{k_{2,\text{right}} (t - t_R)\} \qquad \text{for } t > t_R + B_{0.1} \tag{5.19}$$

hold to the restriction that the slopes of the Gaussian and exponential functions at the respective connecting points should coincide. The model parameters are calculated as follows:

$$k_{1,\text{left}} = 0.1\, h_0 \exp(k_{2,\text{left}}\, A_{0.1}) \; ; \quad k_{2,\text{left}} = \frac{\sigma_0\, A_{0.1}}{(\sigma_0 - c_1\, A_{0.1})^3} \tag{5.20}$$

$$k_{1,\text{right}} = 0.1 h_0 \exp(k_{2,\text{right}}\, B_{0.1}); \quad k_{2,\text{right}} = -\frac{\sigma_0\, B_{0.1}}{(\sigma_0 + c_1\, B_{0.1})^3} \tag{5.21}$$

5.4. Experimental

The training set used to develop the approaches described in this work consisted of 15 sulphonamides: (1) sulphacetamide, (2) sulphachloropyridazine, (3) sulphadiazine, (4) sulphadimethoxine, (5) sulphaguanidine, (6) sulpha-merazine, (7) sulphamethazine, (8) sulphamethizole, (9) sulphamethoxazole, (10) sulphamonomethoxine, (11) sulphanilamide, (12) sulphapyridine, (13) sulphaquinoxaline, (14) sulphathiazole, and (15) sulphisoxazole, all from Sigma (St. Louis, MO, USA). These compounds were analysed with three analytical columns containing 5 μm particles from ACE (Aberdeen, Scotland, UK): C18 (9 cm long), phenyl (5 cm), and cyano (11 cm), which were available in our laboratory, but other set of related solutes and columns of any other length could be used as well. Sulphonamides were eluted using mixtures of acetonitrile (Scharlau, Barcelona, Spain) and water buffered at pH 3.5 with 0.01 M dihydrogen phosphate anhydrous from Sigma (Roedermark, Germany) and HCl from Scharlau. The mobile phase compositions (acetonitrile percentage, *v/v*) in the experimental designs were the following: C18 (10, 13, 15, 17 and 20), phenyl (10, 13, 15 and 20), and cyano (10, 13, 15 and 20).

The chromatographic analyses were carried out with an Agilent (Waldbronn, Germany) instrument, equipped with a quaternary pump (Model 1260 Infinity) run at 1 ml/min, an autosampler (Model 1200) with 2 ml vials, a multiple-variable wavelength UV-visible detector (Model 1200), and a temperature controller (Model 1100) fixed at 25 °C. The injection volume was 20 μl, and the detection wavelength, 254 nm. The dead time was determined by injection of KBr from Acros Organics (Fair Lawn, NJ, USA). The system was controlled by an OpenLAB CDS LC ChemStation (Agilent B.04.03).

The mathematical treatment was performed using MATLAB 2018a (The MathWorks Inc., Natick, MA, USA) routines.

5.5. Results and discussion

In this work, P_C is literally estimated according to its definition, as the number of peaks that can be located one after the other with identical overlap conditions, up to fill completely a given elution window. Predictions of P_C carried out considering only the peak features for only one compound eluted at a given condition, or the mean properties for a series of peaks of related compounds, are not sufficiently realistic. For a hypothetical sample with the solute peaks consecutively allocated one after another, the retention characteristics for each peak (q, S_1 and S_2 in Equation (5.9)) must adopt very specific values. In the approach described in this work, P_C is accurately predicted from simulations, using a series of fictitious solutes derived from a training set, whose properties are gradually modulated to accomplish the consecutive elution, where the variation in peak profile (width and asymmetry) is also taken into account.

5.5.1. *Generation of a series of fictitious solutes from the properties of the training set*

The methodology developed for P_C prediction consists of generating (based on the behaviour of experimental peaks) a series of predicted peaks corresponding to fictitious solutes that under the separation conditions (as simple as isocratic elution or as complex as multi-linear gradients with sudden changes) elute consecutively (one after the other), fulfilling the desired overlap condition. The solute properties in the series are obtained from a training set of compounds in a range of polarities, including the solutes of interest (e.g., sulphonamides, amino acids, β-blockers, phenols, steroids, or flavonoids,

etc.). The process is carried out by building correlations between the parameters in Equation (5.9), obtained for the standards in the training set as follows:

$$S_1 = a + b\, q \tag{5.22}$$

$$S_2 = c + d\, q \tag{5.23}$$

A similar correlation between the slope and intercept has been commented in the literature for the logarithmic linear model of $\log k$ versus φ, for structurally related compounds [42–44]. We have checked that the use of $P_{\mathrm{M}}^{\mathrm{N}}$ instead of φ (Equation (5.9)) gives rise to better correlations between the regression parameters (see also Ref. [36]). Solute retention in RPLC increases with the molecular size and hydrophobicity; therefore, S_1 should be larger for later eluting solutes. The linear regression coefficients of Equations (5.22) and (5.23) are related to the similarity of solutes: the more similar, the larger the regression coefficient of the correlations.

The experimental designs used to build the above correlations comprised the range 10–20% acetonitrile for sulphonamides, for the three columns (C18, phenyl and cyano). This range was the most appropriate to avoid excessively short or long retention times for these compounds. Figure 5.1 shows the established correlations for S_1 and S_2 versus q. The chromatographic data for the 15 sulphonamides were obtained with five, four and four isocratic mobile phases, for the C18, phenyl and cyano columns, respectively (see Section 5.4). The fitted values for S_1, S_2 and q are provided in Table 5.1.
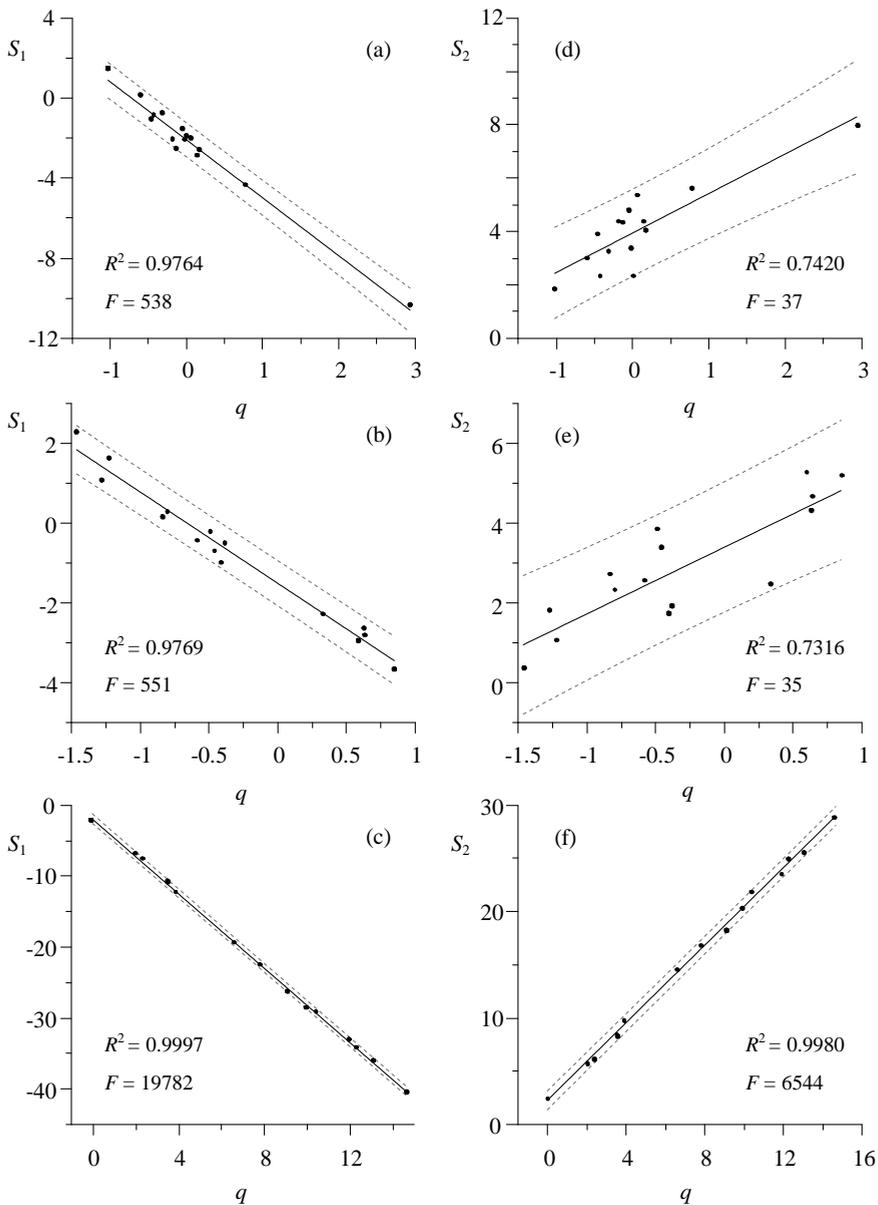
Figure 5.1. Correlation between the parameters in the logarithmic quadratic retention model (Equation (5.9)), applied to a mixture of 15 sulphonamides separated using: (a,d) 9 cm C18 column, (b,e) 5 cm phenyl, and (c,f) 11 cm cyano. The 95% confidence intervals and regression straight-lines are given.

Table 5.1. Regression parameters in Equation (5.9) for the set of 15 sulphonamides, eluted in the 10–20% (v/v) acetonitrile range.

Compounds	C18 column			Phenyl column			Cyano column		
	q	S_1	S_2	q	S_1	S_2	q	S_1	S_2
Sulphacetamide	-0.4283	-0.8925	2.3248	-0.3842	-0.4945	1.8546	9.9448	-28.5952	20.4231
Sulphachloropyridazine	-0.3161	-0.7596	3.2448	0.8504	-3.6245	5.1044	11.9497	-33.1213	23.6137
Sulphadiazine	-0.0186	-2.1103	3.3616	-1.4648	2.2749	0.3080	3.5178	-11.0386	8.5255
Sulphadimethoxine	-0.0533	-1.5415	4.7679	0.5895	-2.9269	5.2908	12.3152	-34.2380	25.0387
Sulphaguanidine	2.9421	-10.3511	7.9161	-0.4098	-0.9840	1.6786	-0.0295	-2.3933	2.5213
Sulphamerazine	0.1714	-2.6038	4.0324	-1.2258	1.6316	1.0053	3.5303	-10.8815	8.3781
Sulphamethazine	0.7768	-4.3605	5.5718	-0.8019	0.2865	2.2590	2.3407	-7.6559	6.2127
Sulphamethizole	-0.1816	-2.0949	4.3639	-0.8354	0.1579	2.6501	6.6078	-19.4493	14.6527
Sulphamethoxazole	-1.0252	1.4099	1.8618	0.6304	-2.6089	4.2324	13.0848	-36.0121	25.6332
Sulphamonomethoxine	-0.4586	-1.1031	3.8978	-0.4600	-0.6890	3.3162	7.8319	-22.6073	16.8655
Sulphanilamide	0.0025	-1.9418	2.3388	0.3339	-2.2644	2.3998	9.1051	-26.2776	18.3804
Sulphapyridine	0.1440	-2.8942	4.3553	-0.5872	-0.4250	2.4918	2.0097	-6.9729	5.7928
Sulphaquinoxaline	0.0552	-2.0402	5.3379	-0.4903	-0.2038	3.7626	10.4243	-29.2352	21.8967
Sulphathiazole	-0.1331	-2.5581	4.3213	-1.2795	1.0865	1.7420	3.8874	-12.4248	9.8681
Sulphisoxazole	-0.5968	0.1215	2.9982	0.6385	-2.7761	4.5896	14.6680	-40.4882	28.9424

It can be observed that the data used to evaluate S_1 (Figures 5.1a to c) are in general less scattered than for S_2 (Figures 5.1d to f), owing to the smaller magnitude of S_2. For the C18, phenyl and cyano columns, the model parameters in Equations (5.22) and (5.23) were: $a = -2.079 \pm 0.108$, -1.517 ± 0.078, and 1.969 ± 0.162, $b = -2.888 \pm 0.124$, -2.286 ± 0.097, and -2.625 ± 0.019, $c = 3.960 \pm 0.208$, 3.393 ± 0.227, and 2.332 ± 0.0194, and $d = 1.470 \pm 0.240$, 1.676 ± 0.282, and 1.815 ± 0.022, respectively.

Figures 5.2a and b shows the satisfactory correlations achieved for sulphonamides in a wider range (5–30% acetonitrile for the C18 column). Correlations for the separation of 17 o-pthalaldehyde-N-acetyl-cysteine (OPA-NAC) amino acid derivatives, eluted in the range 5.0–25.5% acetonitrile, and 25 phenols in the range 15–60%, are also given (Figures 5.3a and b, and 5.4a and b, respectively). The scattering observed in the plots is correlated to the variability in the molecular structure and in its translation in terms of retention.

The correlations between the model parameters were valid enough to obtain S_1 and S_2 from the q parameter. These correlations were used to generate, with enough accuracy, peaks for sets of fictitious related compounds so that they fulfilled the P_C definition. The values of S_1 and S_2 for different solutes can be finely tuned by setting values of q in Equations (5.22) and (5.23). Once the three parameters in Equation (5.9) are established, the retention times are calculated by numerical integration using Equation (5.12). The range of q values to estimate P_C should be extended so that the retention time domain for the generated peaks covers the selected time window. Probably, the time range between the first and last peak for the available standards in the training set (used to generate the series) will not comprise the whole time domain where P_C should be estimated. This is the case of the set of sulphonamides used in this work.
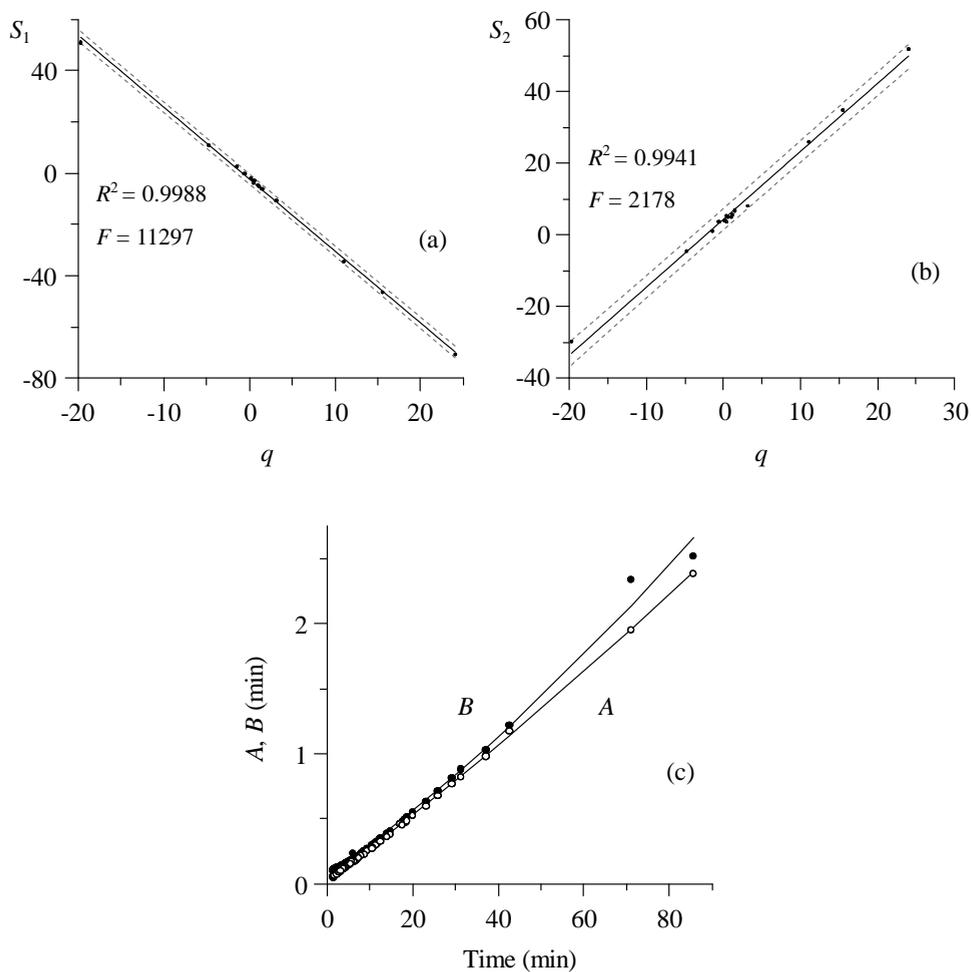
Figure 5.2. (a,b) Correlations between the parameters in Equation (5.9), and (c) isocratic half-width plots, for the set of 15 sulphonamides eluted with 5–30% acetonitrile in a 9 cm C18 column. Regression lines (for the three plots) and 95% confidence intervals (dashed lines in a and b) are given.
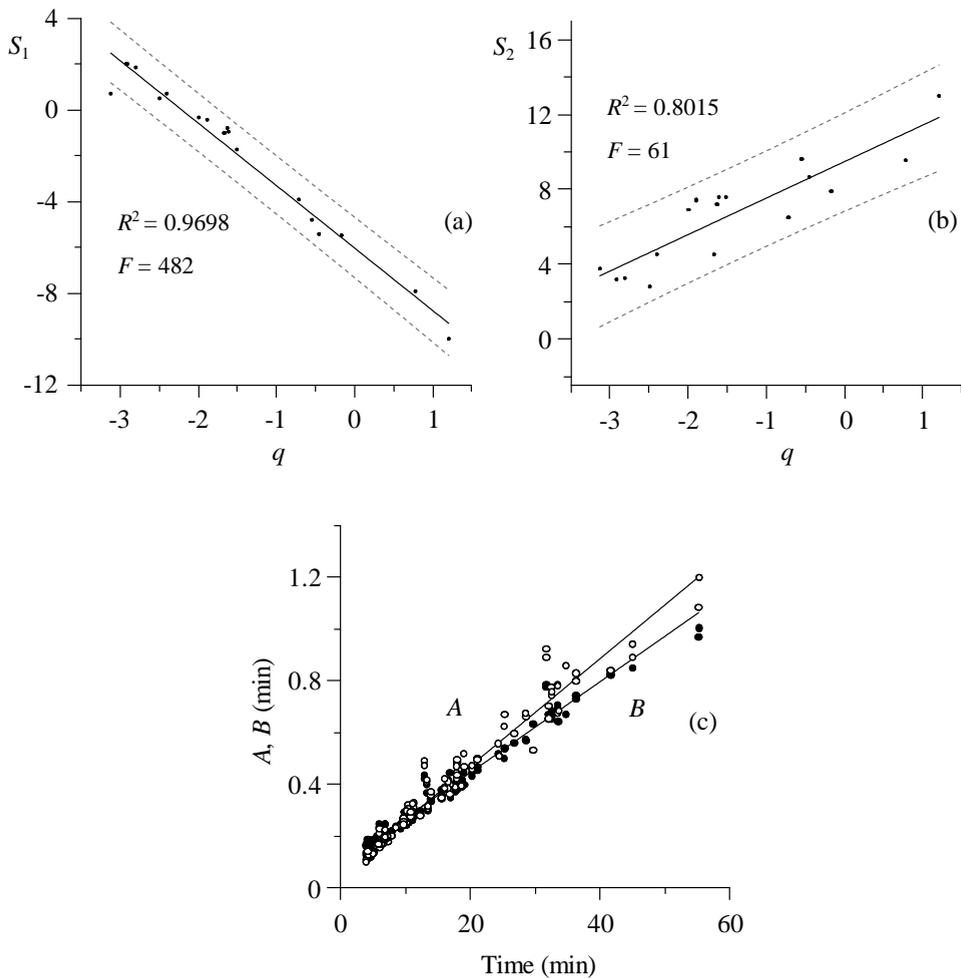
Figure 5.3. (a,b) Correlations between the parameters in Equation (5.9), and (c) isocratic half-width plots, for a set of 17 OPA-NAC amino acid derivatives, eluted with 5.0–27.5% acetonitrile in a 25 cm C18 column. Regression lines (for the three plots) and 95% confidence intervals (dashed lines in a and b) are given.
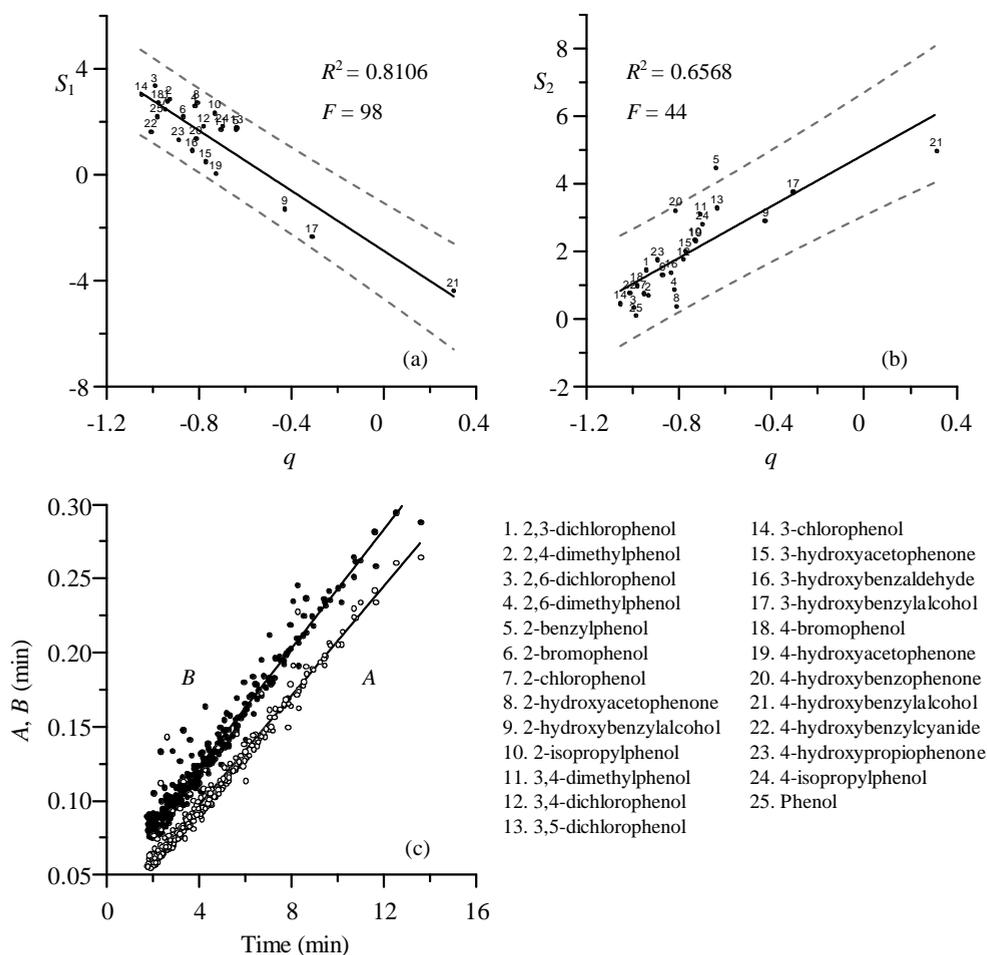
Figure 5.4. (a,b) Correlations between the parameters in Equation (5.9), and (c) isocratic half-width plots, for a set of 25 phenols, eluted with 15–60% acetonitrile in a 15 cm C18 column. Regression lines (for the three plots) and 95% confidence intervals (dashed lines in a and b) are given.

We have considered as minimal retention time a value slightly above the dead time ($1.05 \times t_0$), whereas the maximal time was well above the retention time of the last eluted solute. The adoption of $1.05 \times t_0$ for the retention time of the first peak in the series tries to reduce the impact of biases in S_1 and S_2 for fast solutes. For these solutes, the retention parameters are more uncertain due to the low magnitude of the retention and the influence of refractometric signals and other disturbances, which affect the modelling. The peak for the least retained standard(s) may or not be affected by these problems.

As commented, q is the factor used to modulate the sequence of fictitious solutes emerging from the column one after the other. The extreme values of this parameter (q_{min} and q_{max}), corresponding to the fastest and slowest fictitious solutes in the series, were established using a minimisation iterative procedure explained in Section 5.5.3. It should be noted that the calculated q values are specific for a given column and elution conditions (i.e., isocratic or gradient program).

5.5.2. Prediction of chromatograms considering peak asymmetry

The generation of a series of consecutive peaks not only needs the calculation of retention times, but also the peak width for each fictitious solute. A satisfactory prediction of P_C should also include non-ideal effects, such as the peak asymmetry and extra-column contributions. As explained in Section 5.3.2, in this work peak simulation was performed based on the isocratic values of the left (A) and right (B) half-widths estimated at 10% peak height ratio, using the Jandera's approximation for gradient elution [39].
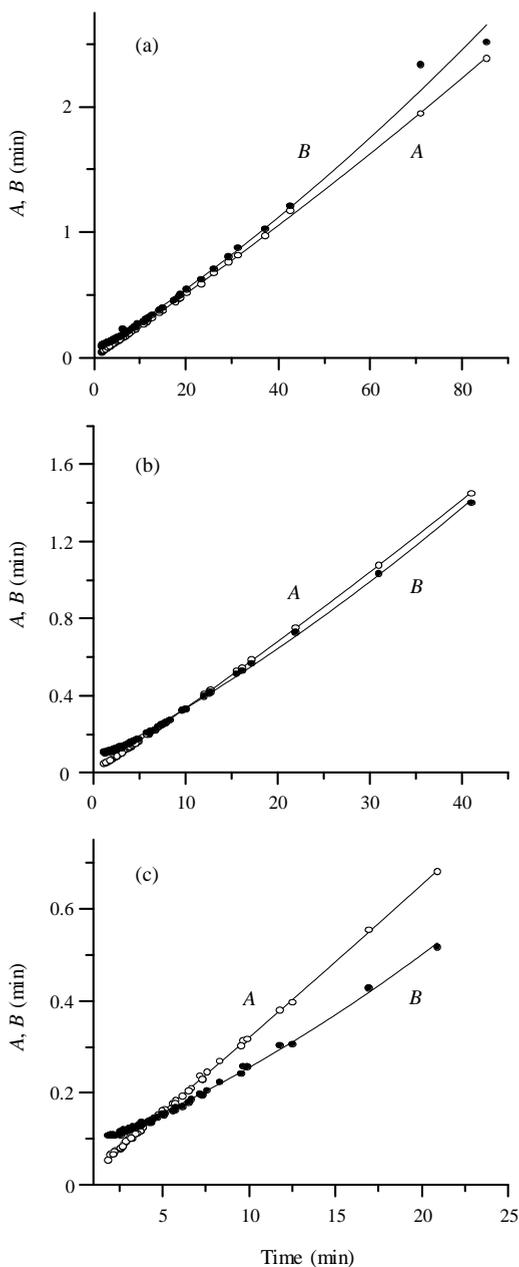
Figure 5.5. Isocratic half-width plots for different columns (column length is given): (a) C18 (9 cm), (b) phenyl (5 cm), and (c) cyano (11 cm). The corresponding parabolas, fitted by least squares, are overlaid.

The obtained correlations for the A and B half-widths versus isocratic retention times (Equations (5.13) and (5.14)) for the set of 15 sulphonamides are shown in Figure 5.5 for the three assayed columns, including the data for all available mobile phases and solutes for each column. The same model is valid for the whole set of fictitious solutes, since there is no significant difference in the kinetics of interaction with the column for the different sulphonamides. The half-width plots for the amino acid derivatives (Figure 5.3c), and phenols (Figure 5.4c), are given as an additional example.

The achieved correlations give directly the values for A and B in isocratic elution at any retention time. For gradient elution, the prediction of half-widths according to the Jandera's approximation implies: (i) obtaining the instant solvent composition when the solute leaves the column, (ii) predicting the isocratic retention time at that instant composition, and (iii) obtaining A and B from the fitted Equations (5.13) and (5.14). For the construction of chromatographic peaks, using the predicted retention times and half-widths, a modified non-Gaussian model was used (see Section 5.3.2).

The prediction of P_C is specific for a given elution program. It must be taken into account that it can include the initial region of the chromatogram (down to t_0), and go beyond the last eluted peak. In both cases, the calculation involves extrapolations, and consequently, the results may be affected by some error in these extreme regions.

5.5.3. Sequential construction of chromatograms to estimate P_C

The construction of a sequence of consecutive peaks fulfilling the P_C definition (i.e., with a pre-fixed degree of overlap) is described below. The process starts by generating a large number of peaks corresponding to fictitious solutes of intermediate properties, which will be called "reference peaks series" (Steps 1 to 3 below). This collection of peaks must cover comprehensively the time window where P_C has to be estimated using the elution conditions under study. These peaks are used as precursors to generate intermediate new peaks to build the "sequence of consecutive peaks" that fulfil accurately the P_C definition (i.e., meet the connection condition at the required height) (Steps 4 to 6).

The retention times for the auxiliary reference peaks series are calculated once their q values (Equation (5.9)) are established. The reference peaks series is used to build half-widths *vs*. retention time dependences (Equations (5.13) and (5.14)), for the applied elution program, from which interpolating the half-widths of each new peak at other retention times. For these interpolations, piecewise cubic Hermite interpolation polynomials were applied [45]. This will finally give rise to the sequence of consecutive peaks fulfilling exactly the desired overlap condition for the P_C definition. The retention time and half-widths for the intermediate peaks will be referred as t_g, A_g and B_g. The connection points constituting the boundaries between each pair of consecutive peaks in the sequential series of peaks are usually established in terms of standard deviation. In this work, a peak width value of 4σ was adopted, since this is the value generally used in the literature for estimating P_C. Since the simulations involve the construction of asymmetrical peaks from predicted left and right half-width values at 10% peak height ratio, the associated standard

deviations (σ_A and σ_B) were obtained by dividing the value of each half-width by 2.145 (see Section 5.5.2 for more details).

The approach developed to generate the sequence of consecutive peaks for P_C prediction will be called "consecutive peaks approximation" (CPA). The steps to be followed are next explained. The procedure is illustrated with the assistance of Figure 5.6, which shows the predicted chromatogram for the elution of the 15 sulphonamide standards using a multi-linear gradient giving rise to good resolution.

Step 1: The first step consists in accommodating the range of q values for the reference peaks series, so that it covers the retention time range in which P_C must be estimated. The extreme values should be modified up to reach a minimal q_{min}^* value that gives rise a retention time of $1.05 \times t_0$ and a maximal q_{max}^* value that matches the retention time of the last eluted solute, or any other below or above it. Each of these q^* values were obtained iteratively using a unidimensional search method, based on the Simplex algorithm. For example, for finding the lower extreme value (q_{min}^*), the search started with three points: q_{min} (for the least retained solute in the series of sulphonamide standards), q_{max} (for the most retained sulphonamide) and $(q_{max} + q_{min})/2$. After obtaining the retention times associated to these three starting q values, a new point is generated closer to q_{min}^*, discarding the least favourable value of the starting point. Once the set of three points is close to q_{min}^*, the algorithm collapses the search by reducing the searching distance, up to reach q_{min}^*. In this way, the algorithm moves the set of three points towards the target time.
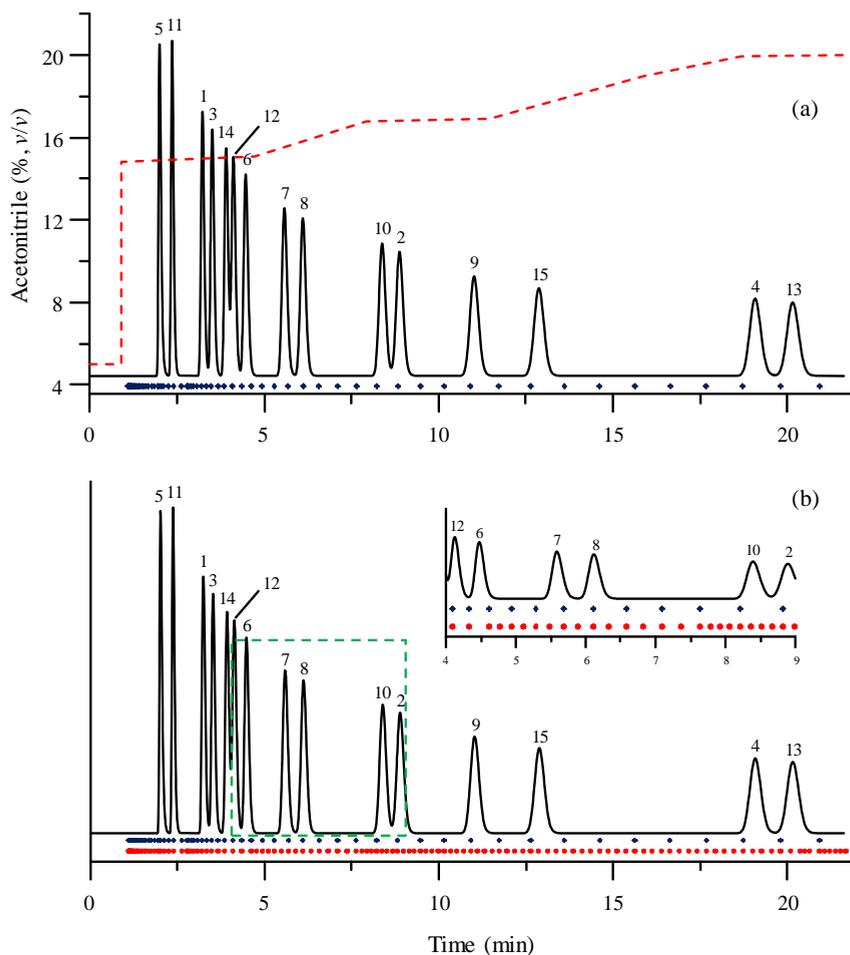
Figure 5.6. Generation of the series of peaks according to the correlations in Figure 5.1 for the C18 column, using a multi-linear gradient. (a) Initial distribution of the reference peaks series for fictitious solutes (crosses, Step 1 of the CPA approach), (b) inclusion of intermediate fictitious solutes in the original distribution (dots, Step 2), and (c) final chromatogram fulfilling the P_C definition (see text for details). The chromatogram of the mixture of 15 sulphonamides is shown in (a)–(c) as thick line, and the gradient program is overlaid in (a) as dashed line. A magnified view of the region between 4 and 9 min is shown in (b).
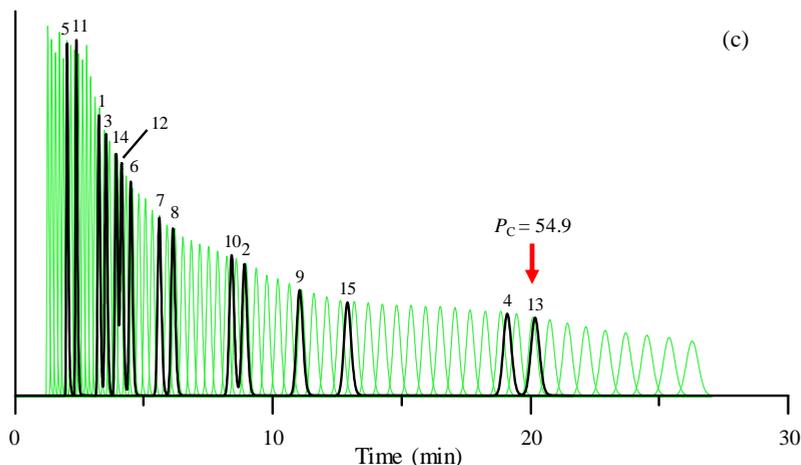
Figure 5.6 (continued).

Step 2: The extended values of q are used to generate an arbitrary number of peaks (e.g., 100) by increasing regularly q in the range between q_{min}^* and q_{max}^*. For isocratic elution, an exponential distribution of retention times is obtained, since Equation (5.9) is logarithmic. For gradient elution, this distribution is altered according to the gradient program. The retention times that represent the location of the initial reference peaks series, for an arbitrary multi-linear gradient, are represented in Figure 5.6a as crosses below the chromatogram of the 15 sulphonamide standards.

Step 3: It should be noted that the distribution of the initial reference peaks series along the elution time window is not satisfactory, particularly for multi-linear gradients (see crosses under the peaks in Figure 5.6a, which represent the retention times of the initial fictitious peaks). In order to sample more uniformly the separation space, a cyclic addition of new peaks is carried out. This correction is particularly important when the gradients include segments with steep slopes. The addition process consists of filling the largest gaps

between consecutive peaks at interpolated q values. The retention time for each new peak is obtained by averaging the q values for two consecutive solutes, then calculating the corresponding S_1 and S_2 parameters, and finally, the corresponding retention time. The addition of peaks is carried out starting by the solute pair showing the largest gap between their retention times. The process is repeated successively for the smaller gaps, until the maximal separation between solutes reaches a pre-established value, or the number of added fictitious solutes reaches a pre-fixed number. The obtained reference peaks series samples comprehensively the variation of the peak properties along the elution program. The process is illustrated as the row of dots below the row of crosses in Figure 5.6b.

Step 4: The construction of the sequence of consecutive peaks fulfilling the P_C definition starts by placing the peak at the dead time with its expected width, obtained with Equations (5.13) and (5.14) for $t = t_0$. When more than one peak can be located in the time window between t_0 and $1.05 \times t_0$, the peaks adopt the profile of the peak at $1.05 \times t_0$. This implies a certain simplification, which may slightly underestimate P_C. However, this decision avoids the drawbacks related to insufficiencies in the models close to the dead time (i.e., the existence of non-linearities in the S_1 and S_2 versus q relationships, Equations (5.22) and (5.23)).

Step 5: The peaks forming the sequence of consecutive peaks, whose retention is beyond $1.05 \times t_0$, are built based on the half-widths *vs.* retention time dependences for the applied elution program. The process used to generate these peaks is sequential. The addition of a certain peak i to the sequence of consecutive peaks beyond $1.05 \times t_0$ is next described.

The retention time of peak i is first approximated from the properties (t_g, A_g and B_g) of the previously added peak $i-1$, according to:

$$t_{g,i} = t_{g,i-1} + p\,\frac{B_{g,i-1} + A_{g,i-1}}{2.145} \qquad (5.24)$$

The p value is a variable that governs the overlap level for the sequence of consecutive peaks ($p = 2$ implies a distance between peaks of 4σ). Here, $t_{g,i}$ is a starting value, because it depends on the $A_{g,i-1}$ value of the previous peak. Therefore, $A_{g,i}$ should be updated to obtain the correct distance for the new peak from the previous one. The updating process is carried out based on the A_g versus t_g dependence in an iterative fashion:

$$t_{g,i}^{*} = t_{g,i-1} + p\,\frac{B_{g,i-1} + A_g(t_{g,i})}{2.145} \qquad (5.25)$$

If the retention times obtained in two consecutive iterations match each other, the peak is accepted in the chromatogram of the sequence for P_C evaluation. New peaks are then added following this procedure up to cover the retention time window under study.

Figure 5.6c shows the sequence of consecutive peaks for P_C evaluation, built according to the rules above, drawn as thin lines. The predicted chromatogram for the 15 sulphonamide standards under the assayed gradient, based on particular retention models (Equation (5.9)) for each compound fitted from experimental values, is overlaid with thick lines. As observed, the simulated peaks for P_C evaluation behave according to the peaks for the 15 sulphonamides obtained from particular models. The similarity between the sequence of fictitious peaks and the chromatogram for sulphonamides demonstrates the reliability of the proposed approach. Equivalent figures will be shown below for other elution conditions.

The treatment outlined above provides accurate values of P_C at the locations where each peak in the chromatogram connects with the next one. The time at which peak i ends (t_i) is:

$$t_i = t_{g,i} + p \frac{B_i}{2.145} \tag{5.26}$$

At this time, i consecutive peaks with the established overlap condition (p value) have been fully eluted; therefore, $P_C = i$. Based on the t_i versus $P_{C,i}$ relationship, it is possible to define a continuous function, and from this to estimate P_C either between the dead time and a certain time value, or associated with any other time range.

5.5.4. Isocratic elution

5.5.4.1. Prediction of P_C for different column types and lengths and mobile phase compositions

The CPA approach described in Section 5.5.3 was applied to evaluate P_C for three columns (C18, phenyl and cyano), based on the elution of the 15 sulphonamide standards, using isocratic, linear and multi-linear gradients. Although in isocratic elution the prediction of P_C is well known, it is interesting to check first the agreement of the results obtained with the proposed approach and the conventional calculation of P_C. For this purpose, the difference in separation performance offered by the three columns under isocratic elution was examined, attending to the P_C values in the elution window between the dead time and the retention time of the last eluted sulphonamide. The P_C values for the three stationary phases at different column lengths and mobile phase compositions are given in Table 5.2. The mean efficiency in each situation for the set of 15 sulphonamides is also indicated.

Table 5.2. Estimation of P_C in the range between the dead time and the retention time for the last eluted solute (sulphaquinoxaline), under several isocratic conditions.

	Column			P_C		$t_{R,last}$ (min)	Plate number		
	Column	Length (cm)	% ACN	CPA	Eq. (5.2) [a]		N_{mean}	N_{min}	N_{max}
Stationary phase type	C18	9	15.0	53.16	52.90	28.70	3960	1100	6200
	Phenyl	5	11.5	48.41	47.90	29.50	2880	650	4150
	Cyano	11	8.6	50.21	53.26	29.60	4870	2300	6400
	Phenyl	9	14.5	40.95	44.77	29.56	2565	800	4200
	Cyano	9	8.0	51.90	55.38	28.8	4600	2100	6400
Solvent effect	C18	9	10.0	74.21	75.21	86.12	4590	1300	6150
	C18	9	15.0	53.16	52.90	28.70	3960	1100	6200
	C18	9	20.0	36.82	36.19	12.49	3210	1000	5600
Column length effect	C18	2	15.0	31.65	31.02	6.48	1660	104	5400
	C18	5	15.0	46.60	46.39	16.00	3215	430	6100
	C18	15	15.0	55.33	55.00	48.25	4105	1650	6200
	C18	25	15.0	55.49	55.26	81.00	4030	1530	6200

[a] Estimated with the mean plate number.

A first comparison was made for the column lengths used in the modelling step: 9, 5 and 11 cm for the C18, phenyl and cyano columns, respectively. The same column length (9 cm) was next considered to make the results for the three stationary phases more comparable. In all cases, the solvent composition was adapted to get similar retention time for the last eluted sulphonamide (compound number 13, sulphaquinoxaline) of around 29 min. As observed, the P_C values were similar for the C18 and cyano columns, and significantly smaller for the phenyl column.

The effect of the column length and mobile phase composition on the P_C values for the C18 column is also shown in Table 5.2. At smaller elution strength, the number of observed peaks increases. It should be noted that the peak width for related compounds, as is the case of sulphonamides, depends only on the retention times and not on the mobile phase composition (see Figure 5.5, where the data obtained for several mobile phases and all assayed solutes are overlaid). This also holds for the P_C vs. retention time trends: the differences in P_C observed in Table 5.2, as the organic solvent content in the mobile phase decreases, are due to the exponentially longer retention time for the last eluted solute under isocratic conditions, when the elution strength decreases. Meanwhile, a change in column length gives rise to an increase in P_C that tends to a final asymptotic value ($P_C \approx 55.5$), once the extra-column contributions become negligible. This asymptotic value is also observed for the mean efficiency of the set of sulphonamides.
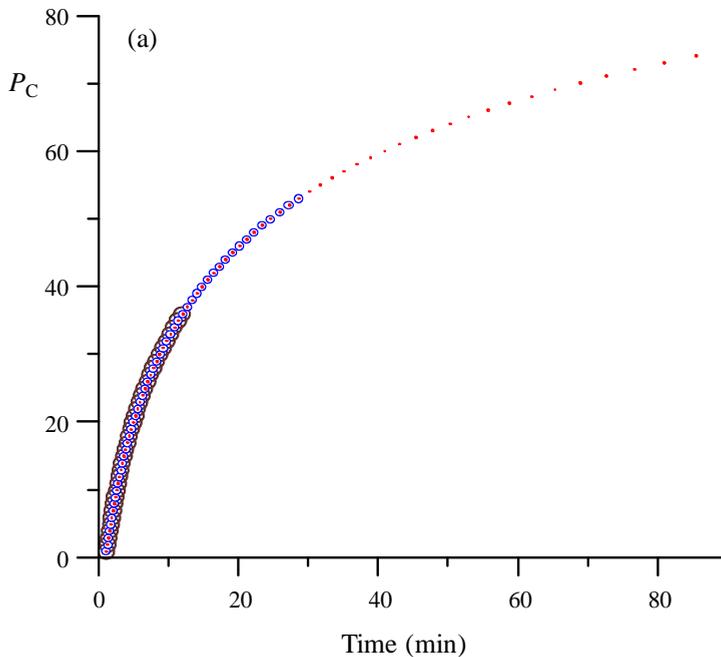
Figure 5.7. Predicted P_C values under different conditions according to the CPA approach, corresponding to the elution of 15 sulphonamides eluted from a C18 column, measured between the dead time and the retention time for the most retained compound (sulphaquinoxaline): (a) Isocratic elution using mobile phases of 20% acetonitrile (black big circles), 15% (blue small circles), and 10% (red dots). (b) Gradient elution considering different programs, where the retention time for sulphaquinoxaline is pointed out. The linear gradient is depicted in Figure 5.9a, the multi-linear gradient (45 min) in Figure 5.9b, and the multi-linear gradient (20 min) in Figure 5.6a.
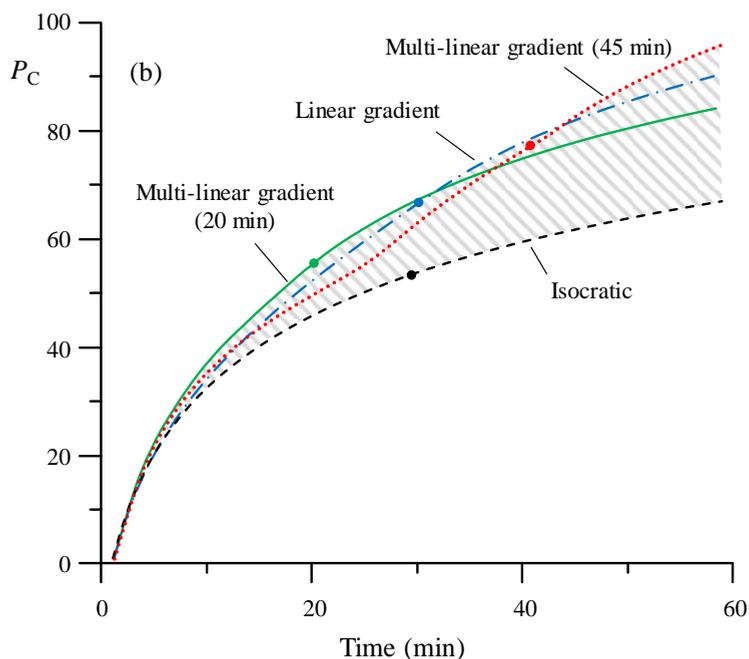
Figure 5.7 (continued).

Figure 5.7a depicts the P_C values for time windows of increasing length, measured from the dead time, for 20, 15 and 10% acetonitrile using the C18 column. The last point in each series corresponds to the elution of the last eluted sulphonamide. The plot illustrates how all P_C trends for different mobile phase compositions actually correspond to a common behaviour. This is another consequence of the common trends of the half-width plots (see Figure 5.5a for the C18 column).

Figure 5.8 shows the sequence of consecutive peaks for the fictitious solutes built to evaluate P_C in isocratic elution for the three columns (thin lines). The predicted chromatogram for the set of 15 sulphonamides is overlaid with thick lines. The good agreement between the sequence of consecutive fictitious peaks and the peaks of the sulphonamide standards is noteworthy.
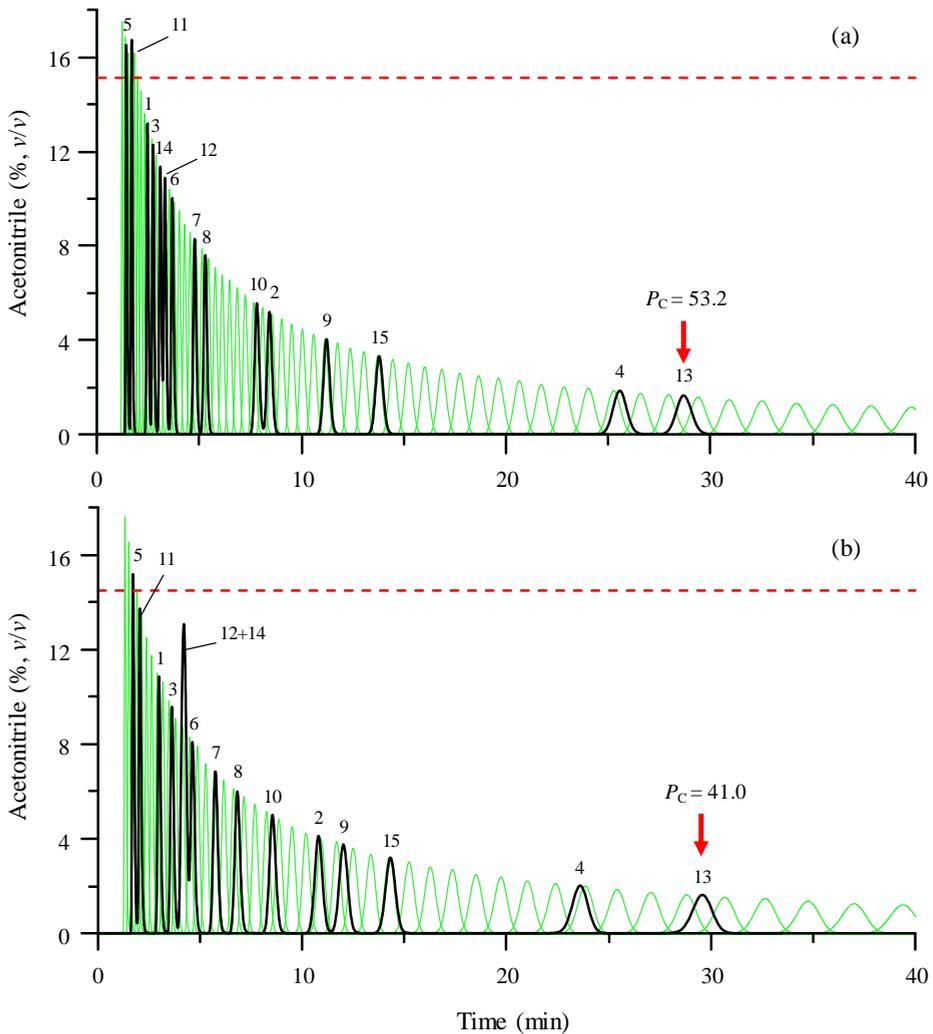
Figure 5.8. Simulated chromatograms for P_C estimation (green thin line) for the elution of 15 sulphonamides separated using isocratic elution and different columns, all of them of 9 cm: (a) C18, (b) phenyl, and (c) cyano. The chromatogram of the mixture of 15 sulphonamides is also shown (black thick line). The mobile phase composition is depicted in each chromatogram as red dashed line.
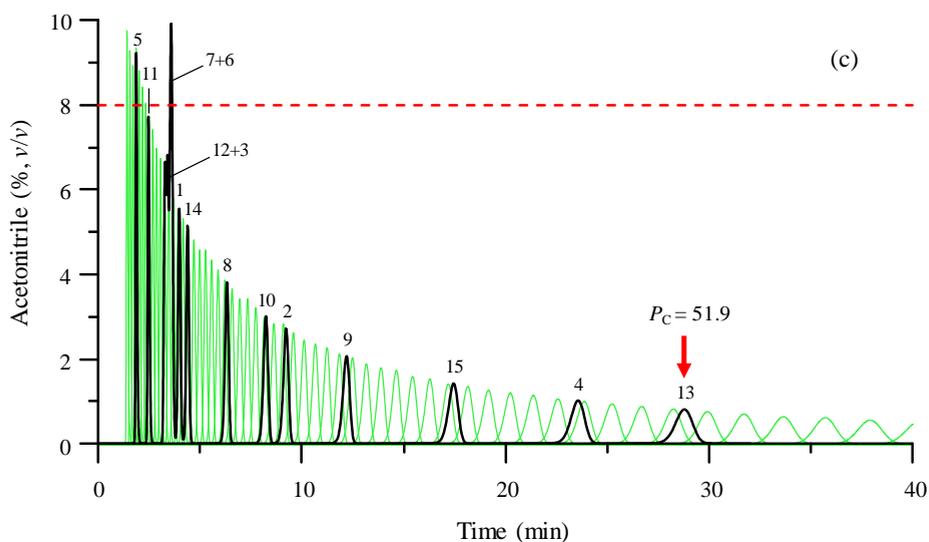
Figure 5.8 (continued).

5.5.4.2. *Validation of CPA by comparison with Equation (5.2)*

The P_C values estimated with CPA were compared with the corresponding values found using Equation (5.2). This makes use of the efficiency value from only one compound eluted with a particular mobile phase, or the mean efficiency from a few compounds eluted with one or several mobile phases. As observed in Table 5.2, the efficiency of sulphonamides inside the elution window exhibits significant variability. Therefore, P_C prediction using the efficiency from only one compound will change with the selected compound. Also, the column or system efficiency, given by the asymptotic value at sufficiently long retention times, will lead to overestimations of P_C. This overestimation will be severe for the situations of more practical interest, where the elution window covers from the dead time to retention times close to the analysis time.

An example that shows the effect of the variability of the efficiencies on the prediction of P_C with Equation (5.2) is next given for the 9 cm C18 column, using 15% acetonitrile in the time window between the dead time and the location of the last eluted solute. In these conditions, P_C would range between 74.2 and 48.2, based on the efficiencies for the most and least retained sulphonamides. The approach that builds a sequence of neighbour peaks in contact with each other (CPA) gives a unique value of $P_C = 53.2$.

The dependence of P_C with the selected peak, for the classical prediction, can be compensated, to a certain extent, by averaging the efficiencies of solutes sampling the elution window. The mean efficiency can be estimated from experimental chromatograms for a set of compounds. An alternative is using the mean efficiencies from simulated data obtained according to the methodology explained in Section 5.3 for the prediction of retention times and peak profiles. This allows exploring the performance under unassayed conditions. The values of P_C estimated with Equation (5.2), indicated in Table 5.2, were obtained following this approach. As observed, the P_C values obtained using mean efficiencies agree satisfactorily with the values obtained with CPA, for all tested conditions.

5.5.5. Linear gradients

As commented in Section 5.2, Neue et al. reported, in successive articles, a methodology to evaluate P_C for linear gradients, under a number of assumptions and simplifications [6,9,29]. The final equation for RPLC (Equation (5.3)) is valid in the absence of extra-column effects, and the time window for estimating P_C is restricted to the linear gradient ramp (i.e., the calculation should involve consecutive solutes eluting along the ramp). Similar to Equation (5.2), a unique efficiency value is used, which may lead to over- or

under-estimations. The algebraic treatment is only valid for solutes whose retention is described by equations of the type $k(t) = k_0 \cdot f(t)$, which include the linear retention model between the logarithm of the retention factor and solvent content, φ. The model that incorporates a quadratic term in φ, which also fulfils the mathematical expression $k_0 \cdot f(t)$, cannot be processed. The final expression proposed by Neue for RPLC linear gradients makes use of a unique elution strength value, which strictly corresponds to only one eluted solute. It should be noted that for a series of consecutive solutes, the elution strength varies.

For comparison purposes, the CPA approach for P_C prediction was applied to the last eluted solute still under the ramp for different linear gradients, starting from the dead time. Table 5.3 shows three situations where the acetonitrile content was increased from 10 to 20%, with gradient times of $t_G = 8$, 15 and 30 min. For these conditions, the last eluted solutes were sulphamonomethoxine (solute 10), sulphisoxazole (solute 15), and sulphaquinoxaline (solute 13), respectively. The retention times, elution strength (S_1), dead time, and width for each solute are also indicated in Table 5.3. Comparing the P_C values obtained with CPA and the Neue approach, it may be observed that the discrepancies are larger as the gradient slope increases (i.e., shorter t_G values). The sequence of consecutive peaks according to CPA, and the predicted chromatogram for the set of 15 sulphonamide standards analysed with a linear gradient are shown in Figure 5.9a. This chromatogram should be compared with that in Figure 5.8a, where the elution is isocratic.

Table 5.3. Estimation of P_C according to different approaches for a linear 10–20% acetonitrile gradient, using different gradient times.

t_G (min)	8	15	30
Δc	0.10	0.10	0.10
Solute[a]	Sulphamonomethoxine	Sulphisoxazole	Sulphaquinoxaline
t_R solute (min)	9.03	15.41	29.28
S_1	16.66	16.11	21.30
t_0 (min)	1.041	1.045	1.055
$2\,(\sigma_A + \sigma_B)$	0.2317	0.3040	0.4150

<div align="center">Peak capacity</div>

CPA	35.86	48.50	65.75
Neue, Eq. (5.3)	42.23	53.21	69.16
Eq. (5.5)[b]	34.47	47.44	69.01
Windows, Eq. (5.6)[c]	37.70	49.13	66.26

[a] Last eluted solute. [b] Calculated with the mean width for all peaks in the chromatogram. [c] Calculated in windows for consecutive peaks.
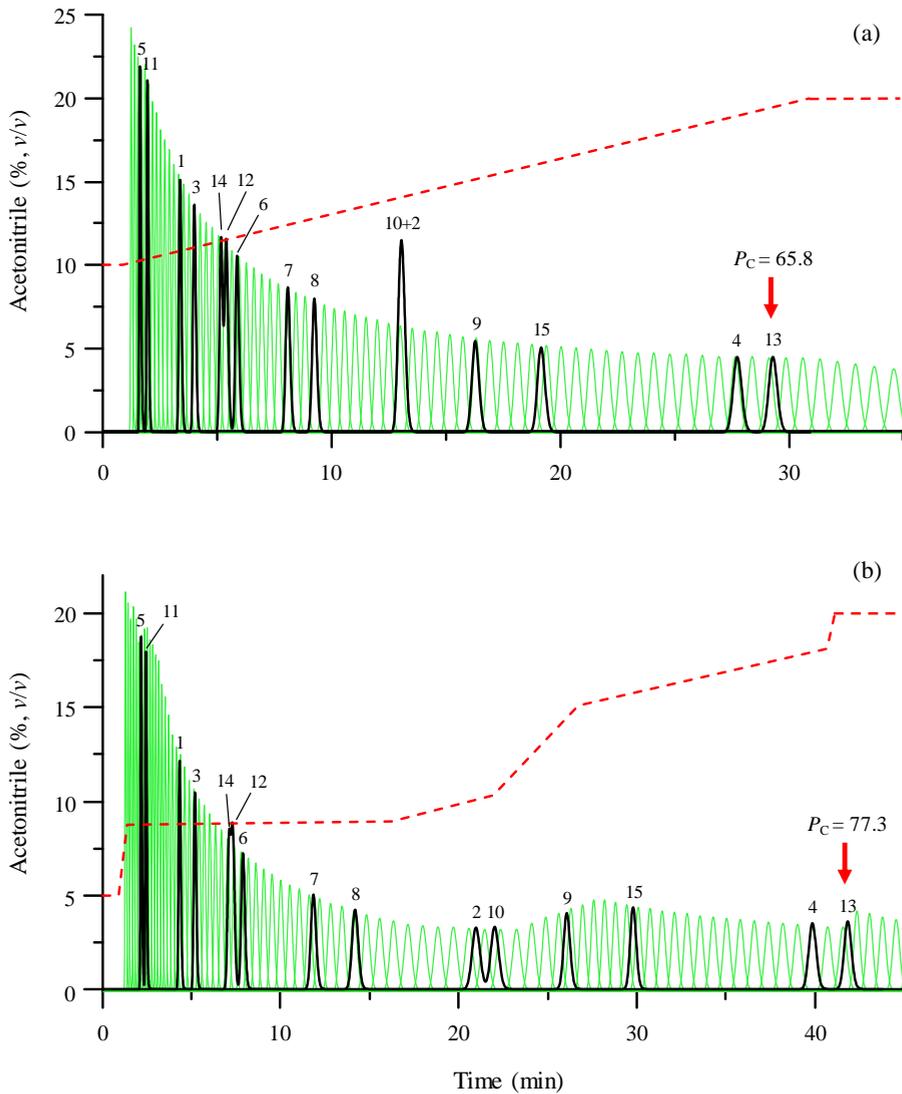
Figure 5.9. Simulated chromatogram for P_C estimation (green thin line) for the elution of 15 sulphonamides separated using the C18 column (9 cm) under a: (a) linear gradient, and (b) multi-linear gradient. The chromatogram of the mixture of 15 sulphonamides is also shown (black thick line), and the gradient programs are overlaid as red dashed line.

Table 5.4 shows the effect of the transition from gradient to isocratic elution on P_C. Gradients in the 10–20% range for progressively larger t_G values are considered up to 2500 min. The P_C prediction for isocratic elution is also given. The predictions were performed for windows between the dead time and the retention time for the last eluted solute (which was always sulphaquinoxaline). As observed, the P_C value for flatter gradients tends to 74.1 peaks, which is the value obtained for isocratic elution for 10% acetonitrile.

Table 5.4. Estimation of P_C according to the CPA approach for a 10–20% linear acetonitrile gradient, using different gradient times.

t_G (min)	P_C	t_R (min) last solute
30	66.66	29.28
60	69.24	39.87
100	70.95	48.47
150	71.79	55.28
200	72.48	59.88
250	72.65	63.24
2500	73.87	82.20
Isocratic at 10% ACN	74.10	85.53

P_C can alternatively be estimated by considering an average peak width for the whole chromatogram (Equation (5.5)). A more elaborate estimation splits the chromatogram in windows defined by each pair of consecutive peaks, where an average peak width is calculated independently (Equation (5.6)). In the latter case, P_C is the summation restricted to the different time windows. In practice, Equations. (5.5) and (5.6) are applied to experimental chromatograms [31], but the estimations can be also carried out using simulated chromatograms. Table 5.3 lists the P_C predictions for the three linear gradients with different t_G values, using the predicted retention times and peak widths for the 15 sulphonamides. In all instances, the P_C values obtained with CPA are closer to those obtained with the windows treatment (Equation (5.6)).

5.5.6. Multi-linear gradients

The algebraic prediction of P_C for gradient elution (Equations (5.3) and (5.4), Neue approach) is very limited. These equations cannot be applied to multi-linear gradients, and to linear gradients where the solutes behave according to the logarithmic quadratic model, or even linear gradients where the calculation is extended beyond the end of the linear ramp. For such situations, Equations (5.5) and (5.6) can be used for either experimental or simulated chromatograms. The prediction of P_C can be also accomplished by applying CPA.

Figures 5.6c and 5.9b show two examples of multi-linear gradients where the sequence of consecutive peaks was built with CPA, applying the Jandera's approximation for peak simulation. The corresponding P_C values according to CPA and using Equation (5.6), which makes the estimation by windows, were 54.9 and 56.3 for the gradient in Figure 5.6c, and 77.3 and 78.4 for the gradient

in Figure 5.9b, respectively. The agreement between both approaches is very satisfactory.

As expected, the gradient program affects the peak width, delaying the peak broadening as the retention time increases. This effect is evidenced by comparing the chromatograms obtained using isocratic elution (Figure 5.8), linear gradient (Figure 5.9a), and the two multi-linear gradients (Figures 5.6 and 5.9b). Peak compression is more evident in cases where the gradient slope is more strongly increased (Figure 5.9b).

5.5.7. Optimisation based on P_C

An issue that makes the comparison of elution conditions in terms of P_C harder is that the calculation can be extended to any value of retention time, and the larger this time, the larger the P_C value (see Figure 5.7). This means that there is no specific value of P_C for a given separation condition, but a relationship between P_C and the time domain at which it has been estimated, giving rise to curves as those plotted for time ranges starting in the dead time. Figure 5.7b depicts P_C curves for four representative situations corresponding to isocratic elution (Figure 5.8a), a linear gradient (Figure 5.9a), and the two multi-linear gradients (Figures 5.6c and 5.9b).

The comparison among different separation conditions has been overcome in the literature, to a certain extent, by setting a fixed time value to establish P_C, or alternatively, calculating P_C by time unit. The results in this work points out, however, that the most meaningful comparison is carried out by referring the calculation to a common compound, preferably one of the most strongly retained. In the assayed conditions, the last eluted compound was always sulphaquinoxaline. The P_C value calculated at the retention time of this compound is marked in Figure 5.7b with a point on the four P_C curves under

study. The existence of a boundary of P_C values is outlined for the three assayed gradients. This suggests that the system cannot provide P_C values outside the region limited by the isocratic trend and the upper common boundary for gradient elution. This will be confirmed in Figure 5.10.

Figure 5.10 depicts the results for: (i) a systematic scanning of isocratic conditions ranging between 10 and 20% acetonitrile (◊), (ii) a systematic scan of linear gradients in the same range using a gradient time of 60 min, but varying the starting and ending compositions in the gradient (□), and (iii) multi-linear gradients of five nodes along 60 min using the same acetonitrile range (○). Since the systematic exploration of multi-linear gradients is not possible, the population of gradients correspond to those evolved along an optimisation using genetic algorithms, where the objective was to get the best resolution.

Each symbol depicted in Figure 5.10 corresponds to the P_C value for the time range between the dead time and the retention time for the most retained compound. The plot shows the existence of a region comprising all possible P_C values for a given system and set of compounds. As commented above, all isocratic separations with the same column follow the same trend (Figure 5.7a). Also, the multi-linear gradients yield the experimental conditions offering the highest P_C at minimal elution time. The upper boundary of the population of gradients can in this way be considered as a Pareto optimal in terms of maximal P_C and minimal time for the last eluted solute. It can be observed that the region depicted for the linear gradients also has a Pareto optimal boundary.
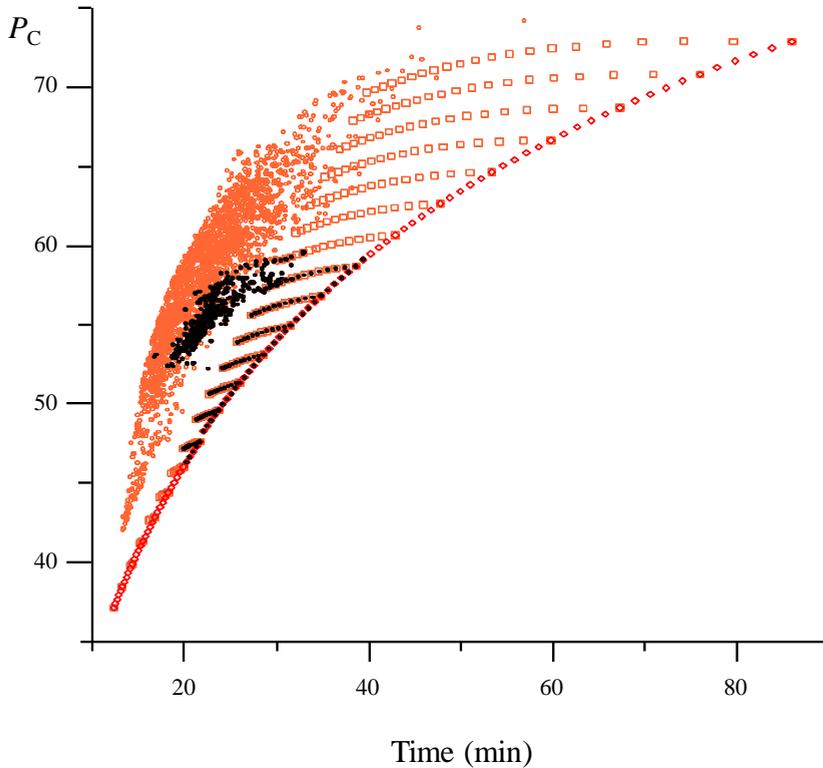
Figure 5.10. Pareto plot corresponding to the P_C estimation from the dead time to the retention time of sulphaquinoxaline (most retained compound), for chromatograms obtained under different conditions: isocratic (◊), linear (□), and multi-linear (○). The chromatograms with peak purity above 0.75 are marked with (●).

The separation conditions (either isocratic, or using linear or multi-linear gradients) offering an acceptable resolution for the set of 15 sulphonamide standards, measured as the product of peak purities > 0.75, are also overlaid in Figure 5.10 (●). The peak purity is the peak area free of overlapping, calculated as [46]:

$$p_i = 1 - \frac{a_\mathrm{o}}{a_\mathrm{T}} \tag{5.27}$$

where a_o is the area under the peak of a given compound overlapped by the hypothetical chromatogram built with the peaks of the other compounds in the sample, and a_T the total peak area.

It is interesting to note that the best conditions in terms of resolution appear in a relatively small region in the plot of P_C *vs*. retention time of the last eluted solute. Figure 5.10 indicates that for the set of 15 sulphonamides, the smallest P_C yielding enough resolution is above 45. An optimisation based on P_C becomes meaningful only for very complex samples. In samples with a small number of compounds, the specific resolution requirements of each peak should be addressed. Thus, in our example, the best separation conditions in terms of resolution are far from those giving rise to the maximal P_C.

5.6. Conclusions

The difficulty of introducing the dependence between peak width and retention, in the integral equation of P_C (Equation (5.1)), makes its calculation through algebraic solutions only possible in ideal and relatively simple situations. This work explores an alternative approach for P_C evaluation (CPA), based on peak simulation, which is applicable to any elution program, from

isocratic to complex multi-linear gradients, considering a non-linear dependence of the retention with the modifier content. The approach generates a sequence of consecutive peaks that fulfil rigorously the P_C definition, based on correlations between the parameters of Equation (5.9) on the one hand, and the half-widths versus the retention time, on the other.

CPA requires previous modelling of the chromatographic behaviour using a training set of structurally-related standards with varying polarity to calculate and correlate their retention and peak profile parameters. For the prediction of chromatographic peaks, CPA is able to process combinations of retention models and gradient programs giving rise to non-integrable expressions in Equation (5.1). Moreover, it allows taking into account effects that can be hardly incorporated in Equation (5.1), such as extra-column contributions and delays of different nature in the solute migration. The inclusion of peak asymmetry in the calculations is also possible, and the peak overlapping level can be easily modulated (e.g., baseline resolved, connection at 13% peak height ratio or at any other height).

CPA provides a more accurate measurement of P_C compared to classical approaches, and also allows its estimation under conditions where previous methods cannot be applied. Since P_C is predicted in a wide range of experimental conditions, without the need of performing new experiments, it is possible to optimise it in very diverse situations. Along the manuscript, practical guidance is given to analysts interested in implementing the approach.

The reliability of the P_C estimation using CPA was validated through the agreement of its results with: (i) the results found with Equation (5.2) for isocratic elution using mean efficiencies, and (ii) the results provided with Equation (5.6) for gradient elution estimating P_C by windows, where the retention times were obtained from predictions using Equation (5.9) fitted for

each particular compound. The good performance of CPA encourages exploring the consequences of special gradient programs in terms of P_C. In contrast with CPA, the estimation of P_C by windows is particularly troublesome in situations where the number of available peaks for standards is insufficient and there are strong variations of slope between the elution of two consecutive peaks. The nature of the calculations made by CPA allows a more comprehensive inspection of the changes along the whole chromatogram.

Any measurement of P_C requires taking a decision about the compounds to be measured, the column and the elution conditions, and all of them should be compatible. Therefore, the same limitations of the CPA approach (and even increased) are present in any of the current P_C estimation methods: classical approaches make use of the efficiency value from only one compound eluted with a particular mobile phase, or the mean efficiency from a few related compounds eluted with one or several mobile phases. It should also be noted that P_C relies on the peak width and asymmetry, which depend on the kinetics of the interactions with the stationary phase. Also, the range of modifier concentrations the method is applicable to will depend on the polarity of the set of training compounds.

Although the approach to calculate P_C has been developed for RPLC, it could be adapted to other chromatographic modes provided that an appropriate retention model is available. The final aim of this study is contributing to the development of a probabilistic enhancement of peak resolution in situations where multi-analyte samples are processed, such as the case of chromatographic fingerprints.

5.7. References

[1]   C.G. Horváth, S.R. Lipsky, Peak capacity in chromatography, *Anal. Chem.* 39 (1967) 1893.

[2]   J.C. Giddings, Maximum number of components resolvable by gel filtration and other elution chromatographic methods, *Anal. Chem.* 39 (1967) 1027–1028.

[3]   E. Grushka, Chromatographic peak capacity and the factors influencing it, *Anal. Chem.* 42 (1970) 1142–1147.

[4]   J.M. Davis, J.C. Giddings, Statistical theory of component overlap in multicomponent chromatograms, *Anal. Chem.* 55 (1983) 418–424.

[5]   Y. Shen, M.L. Lee, General equation for peak capacity in column chromatography, *Anal. Chem.* 70 (1998) 3853–3856.

[6]   U.D. Neue, Theory of peak capacity in gradient elution, *J. Chromatogr. A* 1079 (2005) 153–161.

[7]   Y. Guo, S. Srinivasan, S. Gaiki, Y. Liu, Measuring peak capacity of reversed-phase columns for small molecule compounds under gradient elution, *Chromatographia* 68 (2008) 19–25.

[8]   A. Soliven, I.A. Haidar Ahmad, M.R. Filgueira, P.W. Carr, Optimization of gradient reversed phase chromatographic peak capacity for low molecular weight solutes, *J. Chromatogr. A* 1273 (2013) 57–65.

[9]   U.D. Neue, Peak capacity in unidimensional chromatography, *J. Chromatogr. A* 1184 (2008) 107–130.

[10] S. Pous Torres, J.J. Baeza Baeza, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Peak capacity estimation in isocratic elution, *J. Chromatogr. A* 1205 (2008) 78–89.

[11] L.W. Potts, P.W. Carr, Approximate and exact equations for peak capacity in isocratic high-pressure liquid chromatography, *Anal. Chem.* 83 (2011) 7614–7615.

[12] P. Foley, D.M. Blackney, E.J. Ennis, Peak capacity and peak capacity per unit time in capillary and microchip zone electrophoresis, *J. Chromatogr. A* 1523 (2017) 80–89.

[13] P. Jandera, K. Novotná, L. Kolárová, J. Fisher, Phase system selectivity and peak capacity in liquid column chromatography: The impact on two-dimensional separations, *Chromatographia* 60 (2004) S27–S35.

[14] M. Gilar, A.E. Daly, M. Kele, U.D. Neue, J.C. Gebler, Implications of column peak capacity on the separation of complex peptide mixtures in single- and two-dimensional high-performance liquid chromatography, *J. Chromatogr. A* 1061 (2004) 183–192.

[15] D.R. Stoll, X. Li, X. Wang, P.W. Carr, S.E.G. Porter, S.C. Rutan, Fast, comprehensive two-dimensional liquid chromatography, *J. Chromatogr. A* 1168 (2007) 3–43.

[16] S. Eeltink, S. Dolman, R. Swart, M. Ursem, P.J. Schoenmakers, Optimizing the peak capacity per unit time in one-dimensional and off-line two-dimensional liquid chromatography for the separation of complex peptide samples, *J. Chromatogr. A* 1216 (2009) 7368–7374.

[17] G. Vivó Truyols, Sj. van der Wal, P.J. Schoenmakers, Comprehensive study on the optimization of online two-dimensional liquid chromatographic systems considering losses in theoretical peak capacity in first- and second-dimensions: A Pareto-optimality approach, *Anal. Chem.* 82 (2010) 8525–8536.

[18] C. Desportes, M. Charpentier, B. Duteurtre, A. Maujean, F. Duchiron, Liquid chromatographic fractionation of small peptides from wine, *J. Chromatogr. A* 893 (2000) 281–291.

[19] A. Felinger, M.C. Pietrogande, Decoding complex multicomponent chromatograms, *Anal. Chem.* 73 (2001) 618A–626A.

[20] M.J. Ruiz Angel, R.D. Caballero, E. Simó Alfonso, M.C. García Álvarez-Coque, Micellar liquid chromatography: Suitable technique for screening analysis, *J. Chromatogr. A* 947 (2002) 31–45.

[21] V. Concha Herrera, G. Vivó Truyols, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Limits of multi-linear gradient optimisation in reversed-phase liquid chromatography, *J. Chromatogr. A* 1063 (2005) 79–88.

[22] P. Sandra, G. Vanhoenacker, Elevated temperature-extended column length conventional liquid chromatography to increase peak capacity for the analysis of tryptic digests, *J. Sep. Sci.* 30 (2007) 241–244.

[23] Z. Jiang, N.W. Smith, P.D. Ferguson, M.R. Taylor, Hydrophilic interaction chromatography using methacrylate-based monolithic capillary column for the separation of polar analytes, *Anal. Chem.* 79 (2007) 1243–1250.

[24] A. Ortín, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Finding the best separation in situations of extremely low chromatographic resolution, *J. Chromatogr. A* 1218 (2011) 2240–2251.

[25] C. Ortiz Bolsico, J.R. Torres Lapasió, M.J. Ruiz Ángel, M.C. García Álvarez-Coque, Comparison of two serially-coupled column systems and optimization software in isocratic liquid chromatography for resolving complex mixtures, *J. Chromatogr. A* 1281 (2013) 94–105.

[26] A. Micó Tormos, C. Collado Soriano, J.R. Torres Lapasió, E. Simó Alfonso, G. Ramis Ramos, Determination of fatty alcohol ethoxylates by derivatisation with maleic anhydride followed by liquid chromatography with UV-vis detection, *J. Chromatogr. A* 1180 (2008) 32–41.

[27] A. Micó Tormos, F. Bianchi, E.F. Simó Alfonso, G. Ramis Ramos, Determination of fatty alcohol ethoxylates with diphenic anhydride derivatization and liquid chromatography with spectrophotometric detection: A comparative study with other anhydrides, *J. Chromatogr. A* 1216 (2009) 3023–3030.

[28] M. Beneito Cambra, L. Ripoll Seguer, J.M. Herrero Martínez, E.F. Simó Alfonso, G. Ramis Ramos, Determination of fatty alcohol ethoxylates and alkylether sulfates by anionic exchange separation, derivatization with a cyclic anhydride and liquid chromatography, *J. Chromatogr. A* 1218 (2011) 8511–8518.

[29] M. Gilar, U.D. Neue, Peak capacity in gradient reversed-phase liquid chromatography of biopolymers: Theoretical and practical implications for the separation of oligonucleotides, *J. Chromatogr. A* 1169 (2007) 139–150.

[30] J.W. Dolan, L.R. Snyder, N.M. Djordjevic, D.W. Hill, T.J. Waeghe, Reversed-phase liquid chromatographic separation of complex samples by optimizing temperature and gradient time: I. Peak capacity limitations, *J. Chromatogr. A* 857 (1999) 1–20.

[31] N. Vaňková, J. De Vos, E. Tyteca, G. Desmet, T. Edge, L. Ĉeslová, P. Ĉesla, S. Eeltink, Effect of gradient steepness on the kinetic performance limits and peak compression for reversed-phase gradient separations of small molecules, *J. Chromatogr. A* 1409 (2015) 152–158.

[32] P. Petersson, A. Frank, J. Heaton, M.R. Euerby, Maximizing peak capacity and separation speed in liquid chromatography, *J. Sep. Sci.* 31 (2008) 2346–2357.

[33] Y. Guo, S. Srinivasan, S. Gaiki, Y. Liu, Measuring peak capacity of reversed-phase columns for small molecule compounds under gradient elution, *Chromatographia* 68 (2008) 19–25.

[34] P. Nikitas, A. Pappa-Louisi, Expressions of the fundamental equation of gradient elution and a numerical solution of these equations under any gradient profile, *Anal. Chem.* 77 (2005) 5670–5677.

[35] E. Bosch, P. Bou, M. Rosés, Linear description of solute retention in reversed-phase liquid chromatography by a new mobile phase polarity parameter, *Anal. Chim. Acta* 299 (1994) 219–229.

[36] J.R. Torres Lapasió, M.C. García Álvarez-Coque, M. Rosés, E. Bosch, Prediction of the retention in reversed-phase liquid chromatography using solute-mobile phase-stationary phase polarity parameters, *J. Chromatogr. A* 955 (2002) 19–34.

[37] C. Ortiz Bolsico, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Optimization of gradient elution with serially-coupled columns. Part I: Single linear gradients, *J. Chromatogr. A* 1350 (2014) 51–60.

[38] J.J. Baeza Baeza, M.J. Ruiz Ángel, M.C. García Álvarez-Coque, S. Carda Broch, Half-width plots, a simple tool to predict peak shape, reveal column kinetics and characterise chromatographic columns in liquid chromatography: State of the art and new results, *J. Chromatogr. A* 1314 (2013) 142–153.

[39] P. Jandera, Predictive calculation methods for optimization of gradient elution using binary and ternary solvent gradients, *J. Chromatogr.* 485 (1989) 113–141.

[40] J.R. Torres Lapasió, J.J. Baeza Baeza, M.C. García Álvarez-Coque, A model for the description, simulation and deconvolution of skewed chromatographic peaks, *Anal. Chem.* 69 (1997) 3822–3831.

[41] G. Vivó Truyols, J.R. Torres Lapasió, A.M. van Nederkaassel, Y. Vander Heyden, D.L. Massart, Automatic program for peak detection and deconvolution of multi-overlapped chromatographic signals. Part II: Peak model and deconvolution algorithms, *J. Chromatogr. A* 1096 (2005) 146–155.

[42] K. Valkó, L.R. Snyder, J.L. Glajch, Retention in reversed-phase liquid chromatography as a function of mobile-phase composition, *J. Chromatogr. A* 656 (1993) 501–520.

[43] H.B. Xiao, X.M. Liang, R.C. Lu, Classification of structurally related compounds from Astragalus extract by correlation of the log kw and S, *Chromatographia* 51 (2000) 212–220.

[44] C.F. Poole, *The Essence of Chromatography*, Elsevier, Amsterdam, 2003, pp. 303–304.

[45] C. de Boor, *A Practical Guide to Splines*, Applied Mathematical Sciences Series, Vol. 27, Springer, Berlin, 2001.

[46] S.J. López Grío, G. Vivó Truyols, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Resolution assessment and performance of several organic modifiers in hybrid micellar liquid chromatography, *Anal. Chim. Acta* 433 (2001) 187–198.

CHAPTER 6

SECONDARY CHEMICAL EQUILIBRIA IN REVERSED-PHASE LIQUID CHROMATOGRAPHY

6.1. Abstract

The addition of reagents to an RPLC mobile phase enables the separation of ionisable compounds, inorganic anions and metal ions, using conventional instrumentation, silica-based materials, and hydro-organic mixtures, thanks to a variety of secondary equilibria. This gives rise to several chromatographic modes, whose main features are outlined in this chapter. The effect of the mobile phase pH on the retention of ionisable compounds is described, together with the recommended experimental practice. The mechanism of adsorption of amphiphilic anions or cations on the stationary phase to attract analytes with opposite charge, or suppress the silanol activity, is discussed. Different reagents, such as alkylammonium salts, surfactants (below and above the critical micelle concentration or forming microemulsions), perfluorinated carboxylate anions, chaotropic ions and ionic liquids, are considered. The potential of metal chelation and redox reactions for the determination of metal ions and organic compounds is also summarised.

6.2. Introduction

Theoretically, in reversed-phase liquid chromatography (RPLC) with hydro-organic mixtures as mobile phases, the retention is produced by adsorption on the alkyl-bonded phase; consequently, it is related to compound hydrophobicity: the more hydrophobic the compound, the longer is its retention [1]. RPLC allows the separation of analytes in a wide range of polarities and structures. However, ionised organic compounds and inorganic anions or metals, which are polar, show little or no retention. This has been a challenge in environmental, clinical, and food chemistry throughout the development of RPLC. The situation is even more complex, as there is no ideal support for preparing RPLC stationary phases yet. The vast majority is still prepared with silica, due to its attractive properties: easy derivatisation and control of particle size, porosity, mechanical stability, and incompressibility. However, owing to steric problems in the derivatisation, silanol groups remain on the stationary phase in a non-negligible amount and, when ionised, interact with ionic analytes by ion-exchange, producing attraction or repulsion of cationic and anionic analytes, respectively, which increases and decreases the retention, in some cases excessively. Also, the sorption-desorption kinetics on free silanols is a slow process that yields tailed and broad peaks [2].

In the late 1970s, Horváth and other authors wrote a series of fundamental reports trying to give a solution to the separation of ionisable compounds and inorganic ions, using conventional RPLC instrumentation, silica-based materials, and hydro-organic mixtures [3]. The complexity of the experimental conditions was increased by introducing several reagents (additives) in the mobile phase. This gives rise to secondary reactions on the support or within the mobile phase: dissociation-protonation of ionisable compounds by tuning the pH, ion-exchange processes through adsorption of an ionic lipophilic

reagent on the stationary phase which attracts analytes with an opposite charge or suppress the silanol activity, formation of analyte-reagent ion pairs in the mobile phase, or metal complexation, among others [4].

In conventional RPLC, the solutes of interest are generally eluted in one chemical form and separated via differences in their primary equilibrium constants (i.e., the distribution of the solute between the mobile phase and stationary phase). In the presence of secondary equilibria, the analytes are eluted in more than one form and separated, thanks to differences in their secondary equilibrium constants [5]. Such secondary equilibria can be generically expressed as:

$$A + X \rightleftharpoons AX \qquad (6.1)$$

where A is the analyte or the silanol group on the support, and X is H^+, a lipophilic ion, a ligand, or other added species. The observed retention factor (k) is a weighted average of the retention factors of both chemical forms:

$$k = k_A\,\delta_A + k_{AX}\,\delta_{AX} = \frac{k_A + k_{AX}\,K\,[X]}{1 + K\,[X]} \qquad (6.2)$$

where δ_A and δ_{AX} are the molar fractions of A and AX, [X] is the molar concentration of X in the mobile phase, and K the formation constant (for an acid-base reaction, $\log K = pK_a$, where K_a is the dissociation constant). In practice, the situation can be far more complex, as two or more secondary equilibria may exist simultaneously inside the column.

Secondary equilibria may provide enough selectivity for the separation of mixtures of analytes under intermediate conditions in which comparable amounts of both forms exist. Therefore, they represent a very powerful tool for conventional RPLC to enhance the chromatographic performance (in terms of absolute and relative retention and peak shape). Secondary equilibria have

given rise to new chromatographic modes with an impressive increase in the number of compounds that can be analysed by RPLC. The main features of these modes are outlined next.

6.3. Acid-base equilibria

6.3.1. Changes in retention with pH

Equation (6.2) (with $[X] = [H^+]$) defines a sigmoidal change in the RPLC retention of weak acids and bases as a function of the mobile phase pH, with a pronounced drop around $pH = pK_a$ (referred to the hydro-organic mixture) [6–8]. The height of the transition depends on the hydrophobicity of the neutral species. Acids lose a proton and become ionised when the pH increases, and bases accept a proton when the pH decreases (see Figures 6.1a and d). For polyprotic compounds, the k-pH curve depends on the charge of the different acid-base species.

Small variations in the mobile phase pH at values close to pK_a result in significant changes in retention and selectivity. Therefore, the pH in this region needs to be controlled tightly. Nevertheless, to achieve robust methods, a region scarcely affected by changes in pH is preferable. For weak acids, where a neutral species is obtained at acidic pH, the chromatographic mode is called "ion-suppression chromatography". The analysis of basic compounds is also carried out at an acidic pH to protonate (deactivate) the silanols on the stationary phase. However, separations at low pH are not always feasible, due to long separation times and column instability.
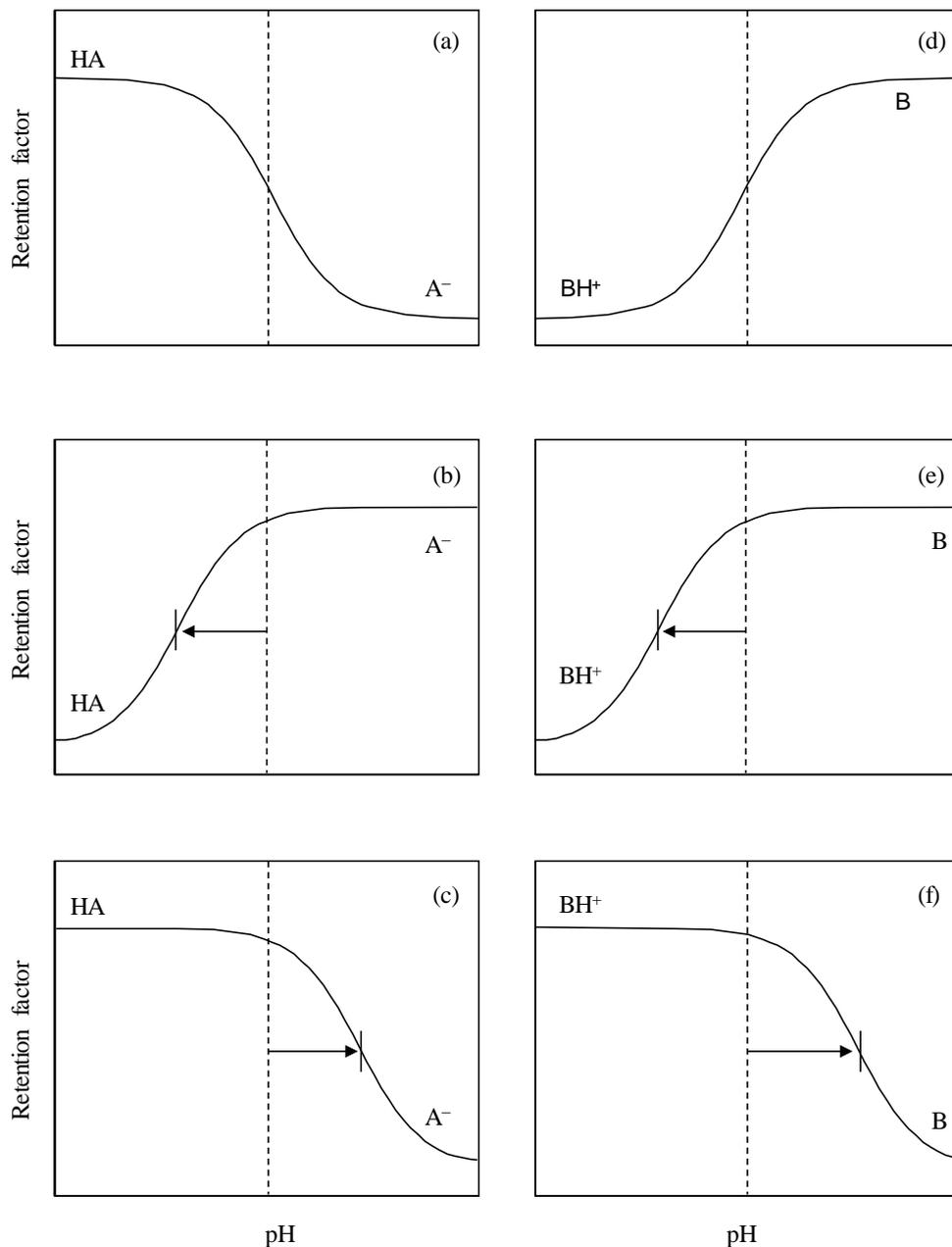
Figure 6.1. RPLC retention versus pH trends of acidic (a to c) and basic (d to f) compounds, without an additive (a,d), and in the presence of cationic (b,e) and anionic (c,f) additives. The arrows indicate the shifts in pK_a .

The retention behaviour of ionisable analytes under organic solvent gradient elution is especially cumbersome. Even using buffered gradients, the variation in mobile phase composition during the programmed gradient can lead to strong changes in the mobile phase pH and pK_a values of both the analyte and the buffer system [9,10].

6.3.2. Buffers and measurement of pH

The working pH range for conventional columns in RPLC is 2.5−7.5. Outside this range, the silica packing can suffer important damage (i.e., hydrolysis of the siloxane bonds below pH = 2, and dissolution of silica above pH = 8). Innovative supports that contain short carbon chains between the silicon atoms, as well as protecting polymer layers, have extended the range to 2−12. In any case, the addition of an appropriate buffer is needed to achieve reproducible retention for ionisable compounds. Common buffers correspond to the acid-base systems of phosphoric, citric, tris(hidroxymethyl)aminomethane (Tris), phthalic, acetic, formic, and ammonium. Phosphoric and citric buffers, which provide control over wide pH ranges, are the most popular. Their main disadvantage is that their inorganic salts may precipitate inside the column if the proportion of organic solvent is too high, particularly with acetonitrile or while maintaining a low column temperature. Only volatile buffers (acetic, trifluoroacetic and formic acids, and their ammonium salts) are compatible with evaporative light scattering (ELS) and mass spectrometry (MS) detection. However, trifluoroacetic acid reduces the sensitivity in MS, particularly when working in the negative ion mode.

The buffering capacity occurs in the range pH = $pK_{a,buffer}$ ± 1. To control the pH appropriately, this should be measured in the hydro-organic mixture, rather than in the aqueous buffer. In principle, the electrode system must be calibrated

with standard buffers prepared using the same solvent composition as the mobile phase (^s_spH scale). As these standards are not commercially available and require careful maintenance, a solution is to measure the pH in the hydro-organic mixture and calibrate the electrode system with aqueous buffers (^s_wpH scale, which can be easily converted to the ^s_spH scale) [11,12]. Column temperature should be controlled, as it affects the degree of ionisation for analytes and buffers [13,14].

Hydrophilic interaction chromatography (HILIC) is nowadays accepted as a complimentary separation mechanism to RPLC for the separation of polar and ionised solutes that are poorly retained in RPLC. Retention has been attributed to partition of the solute between a water layer held on the surface of a polar stationary phase and the bulk mobile phase typically containing a high concentration of acetonitrile. The pH of solutions of formic, phosphoric, trifluoroacetic and heptafluorobutyric acids cover a relatively narrow range when used in water (^w_wpH 1.9–2.8), but a much wider range in 90% acetonitrile (used in HILIC) when the true thermodynamic pH is considered (^s_spH 2.4–5.2). These differences can explain the considerable selectivity changes observed for such buffer systems [15].

6.4. Ion-interaction chromatography

6.4.1. Retention mechanism

An RPLC mode with a broad scope is achieved by adding amphiphilic anions or cations to the hydro-organic mixture [16–18]. The added reagent typically contains a hydrophobic tail that interacts strongly with the bonded chains on the stationary phase, and a charged head projecting out into the

mobile phase to interact with the analytes. The stationary phase modification facilitates the separation of mixtures of ionic and neutral species. The retention is regulated by the nature and concentration of reagent counterion, organic solvent, and competing ions with the same charge as the analyte.

The retention mechanism is not fully understood yet [18–20]. Due to the complexity of the mobile phases, which contain the ionisable or ionic analyte(s), and at least the additive and buffer ions (and their co-ions), it is not easy to sort out their mutual influence on the adsorption behaviour. At the origin of RPLC, bonded phases were considered as equivalent to a "mechanically held liquid phase". Therefore, the theory of the combination of the analyte and lipophilic ions of opposite charge to form an ion pair in the mobile phase, able to partition into the non-polar bulk-liquid stationary phase, is not surprising. Hence, the name "ion-pair chromatography" (IPC) taken from liquid-liquid separations. Experimental facts further suggested a dynamic ion-exchange mechanism, instead, which considers that the lipophilic ion is dynamically distributed between mobile and stationary phases, where it is adsorbed (immobilised), behaving as an ion-exchanger for oppositely charged analytes. This model implies an interaction essentially Coulombic, and pioneered the stoichiometric approach that was followed for decades.

Broader perspectives (non-stoichiometric approaches) consider the ionic analyte as being under the influence of all ions in the chromatographic system. Also, the role of the electrical double layer formed by the lipophilic ion (primary charged ion region) and counterion (diffuse outer region) is envisioned. The analyte is not associated specifically with any charged moiety, and its retention involves its transfer across the double layer. This creates a surface potential, which depends primarily on three parameters: the lipophilic ion surface concentration, and the mobile phase relative permittivity and ionic

strength. The higher the surface concentration, the larger is the effective ion-exchange capacity, and hence the retention of solutes with an opposite charge to the lipophilic ion. This is expected to be spaced over the stationary phase due to repulsion, which leaves much of the surface unaltered and available for the separation of neutral species. The same framework does not hold for small hydrophilic organic and inorganic anions, which probably interact primarily through Coulombic forces. However, in general, other interactions within the mobile phase should not be neglected: the actual mechanism is thus far more complex. A deep insight into the composite mechanistic processes is obscure, as the accurate determination of equilibrium constants is difficult.

IPC is by far the most widely used term for this RPLC mode, but usually it does not describe the real mechanism. Also, this term is usually associated with the addition of small amounts of the lipophilic ion to avoid any excess in the mobile phase. The terms "ion-interaction chromatography" (IIC) or "ion-modified chromatography" have been suggested instead to describe the use of diverse types of ionic additives in RPLC at several concentrations. Other names are also found in the literature, such as "paired-ion chromatography", "hydrophobic chromatography with dynamically coated stationary phase", "surfactant (or soap) chromatography" (referring to the use of ionic detergents as additives), and "hetaeric chromatography" (referring to the use of hetaerons, "counterions").

The adsorbed amphiphilic reagent essentially changes the stationary phase from a non-polar (hydrophobic) to a polar (hydrophilic) charged surface, generating charge sites to serve as ion-exchangers for analytes, positive or negative depending on the nature of the analyte (see Figure 6.2).
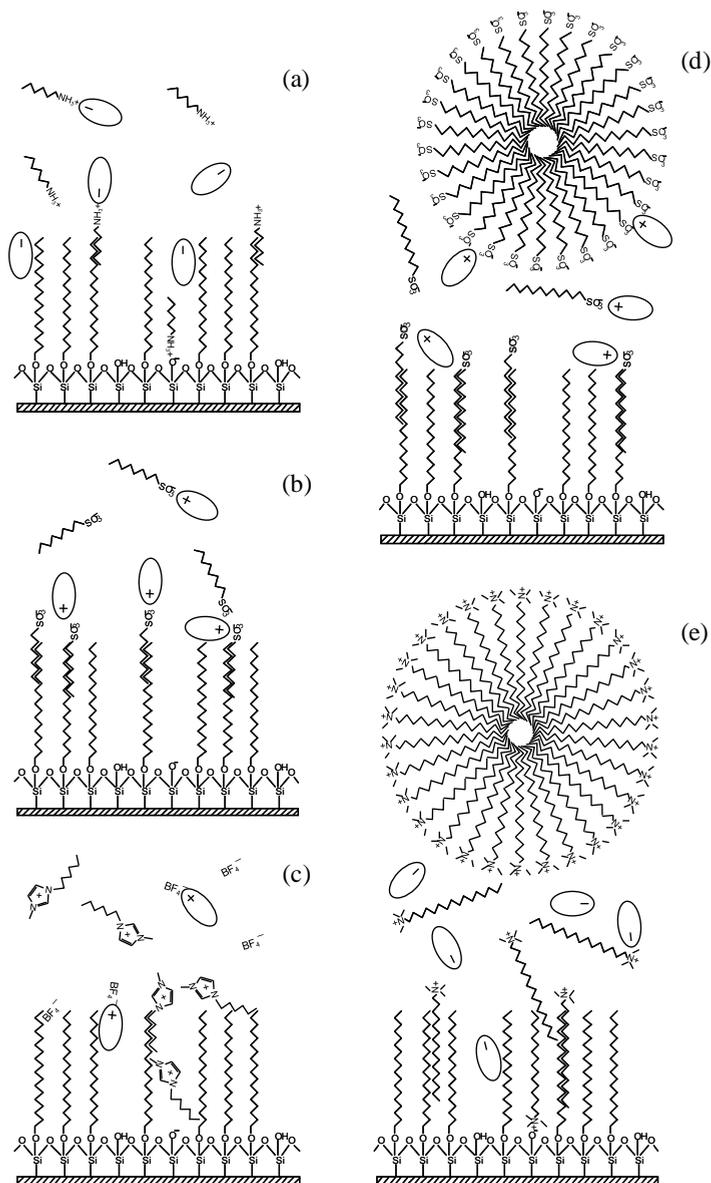
Figure 6.2. Simplified solute environments in a C18 chromatographic system with mobile phases containing: (a) hexylamine, (b) 1-octanesulphonate, (c) 1-hexyl-3-methylimidazolium tetrafluoroborate, (d) sodium dodecyl sulphate, and (e) cetyltrimethylammonium chloride.

The major advantage of the dynamic coating is the possibility of controlling the column ion-exchange capacity by varying the mobile phase composition. A quite distinct alternative is the equilibration of the stationary phase with a highly lipophilic ion. This coating is strongly bound and persists for long periods of subsequent use. The method is known as "permanent coating IIC", and is close to ion-exchange chromatography, where charged groups are covalently bonded to the stationary phase.

6.4.2. Common reagents and operational modes

In principle, any salt containing a lipophilic ion can be used as an IIC reagent. To separate anions, the stationary phase must contain immobilised cations. Conversely, to separate cations, it must contain immobilised anions (see Figures 6.2a and b). Salts of alkylammonium or tetraalkylammonium for anions, and alkyl sulphates or alkylsulphonates for cations (with different alkyl chain lengths) cover most common applications. The longer the alkyl chain, the more hydrophobic the reagent, and stronger the adsorption on the stationary phase. The anion in alkylammonium salts can be inorganic (e.g., chloride, hydroxide, or phosphate), or organic (e.g., salicylate and tartrate). The cation for alkyl sulphate and alkylsulphonate salts is usually sodium or potassium. Newer reagents are perfluorinated carboxylic acids, chaetropic ions and ionic liquids (ILs). New methods may be developed by tailoring the mobile phase composition to suit the retention of a particular analyte, or the separation of a particular mixture.

Popular choices tend to favour the relatively less-lipophilic IIC reagents regarding the separation time. These should be replaced by a more lipophilic ion when the retention is too short. The same column can be converted into an anion-exchanger or a cation-exchanger. The adsorbed layer of lipophilic ion

can be removed by washing the column with an organic solvent such as methanol. On increasing IIC reagent concentration, the retention increases, provided the stationary phase surface remains unsaturated. Meanwhile, on increasing organic solvent concentration, the retention decreases, due to desorption of the reagent and competition equilibria in the mobile phase. Therefore, both IIC reagent and organic solvent should be kept constant in the mobile phase at specified concentrations, in order to maintain a reproducible ion-exchange capacity. It is not essential that the IIC counterion operates as the ion-exchange competing ion. A separate component, such as phosphate, citrate, oxalate, or phthalate, is often added to assist in the elution of strongly retained anions.

The analytes need to be ionised to interact with the IIC counterion. Therefore, the retention of ionisable compounds depends on the pH and pK_a (which changes by interaction of the ionic species with the IIC counterion, see Figures 6.1b, c, e and f). The counterion adsorption onto the column, the interaction between ionic solutes and counterion, and especially the ionisation of solutes and buffer components are temperature dependent; therefore, system reproducibility requires accurate temperature control. Other considerations are the requirement of a longer equilibration time to get a constant counterion coating (especially in gradient elution); the fact that some counterions tend to associate very strongly to the stationary phase changing the initial column properties; the need to saturate the mobile phase with silica for some IIC reagents by inserting a pre-column between the pump and the injection system; and the appearance of system peaks in the chromatograms. Traditional lipophilic reagents are not usually compatible with ELS and MS detection.

6.4.3. Separation of inorganic anions

Surfactant coatings constitute an easy and inexpensive way of converting silica-based RPLC packings into ion-exchangers [21]. Its attractiveness arises from their different ion-exchange capacities and selectivities, by just altering the coating conditions. However, some problems have been described regarding the stability of these coatings: retention times may drift, which forces periodic column regeneration. This has depreciated their use for routine separations. A reproducible behaviour is, however, possible with careful column equilibration to its plateau capacity.

Cationic surfactants with quaternary ammonium groups are frequently used for the separation of inorganic anions. However, coating first with a layer of non-ionic surfactant, then with the cationic surfactant, creates a more efficient column with shorter retention times. On the other hand, when using a surfactant with a single functionality (anionic or cationic), analyte release from the Stern layer to the bulk solution requires a mobile phase with a competing ion to exchange the analyte. If, instead, the stationary phase is coated with a zwitterionic surfactant (with positive quaternary ammonium and negative sulphonate groups close to each other), the analyte experiences simultaneous attraction and repulsion forces. This means that it can be retained by the stationary phase but also be released without the need for a competing ion-exchange ion. This chromatographic mode, termed "electrostatic ion chromatography", constitutes a kind of green chromatography, as the mobile phase can just be pure water or an electrolyte solution, such as $NaHCO_3$ or $Na_2B_4O_7$. The addition of a cationic surfactant to the coating solution containing a zwitterionic surfactant reverses the elution order of monovalent and divalent anions.

Aliphatic amines are also used as cationic ion pair reagents for the analysis of inorganic anions, either metallic (as CrO_4^{2-}, VO_3^-, MoO_4^{2-}, and WO_4^{2-}), and non-metallic (as Cl^-, Br^-, I^-, NO_2^-, NO_3^- and SO_4^{2-}), the retention of which increases. Other applications refer to the analysis of a variety of organic anions.

6.4.4. The silanol effect and its suppression with amine compounds

Nitrogen-containing basic compounds constitute a significant fraction of the drugs used in modern therapy. A large number of compounds of biomedical and biological significance are also bases or zwitterions. However, the RPLC analyses of such compounds with silica-based columns suffer several problems, including long retention, peak tailing, poor efficiency, and strong dependence of retention on sample size. These effects are due to ion-exchange of the cationic analyte on active (dissociated) silanols on the support, the acidity of which raises by the presence of metal impurities [2]. Silanol ionisation cannot be entirely suppressed using mobile phases in the pH range 2.5–7.5. Consequently, much effort has been invested in the chemistry of bonded phases to eliminate metal impurities and residual silanols.

The extreme differences in the behaviour of packing materials of the same type, such as bonded octadecylsilane (ODS), toward basic compounds is due to differences in the silica backbone, type of bonded silane, and coating level –all of them resulting in a varying concentration of surface silanols. The brand-to-brand variation in the selectivity of bonded phase materials is, however, attractive. RPLC would never have reached such broad applicability if only hydrocarbon-like stationary phases were available. With the newer generation of RPLC columns, based on ''ultrapure'' silica and improved bonding technologies, the influence of surface silanols on basic analyte retention is less pronounced. Nevertheless, some tailing still occurs.

At least three solutions to avoid the silanol effect have been suggested [2,22]: reducing the pH to less than 3 to protonate residual silanols (however, using an extreme pH can damage the packing), increasing the pH to obtain neutral analytes (but simultaneously more silanols are dissociated), and masking the electrostatic interaction with IIC reagents (but an additional background for MS detection appears, and the column properties may be permanently altered if the reagent cannot be removed from the stationary phase).

Peak shapes can be improved by using acidic mobile phases containing hydrophobic anions, such as alkyl sulphates or alkylsulphonates, but this is not always successful, and the retention of basic compounds can increase excessively. The use of amines as silanol blockers (suppressors or anti-tailing agents) is also widespread [23,24]. Better silanol suppression is achieved with bulky substituents. Salts of quaternary amines (with alkyl chain lengths usually between 1 and 4), or amines with long alkyl chains (between 4 and 10), seem the best, due to their stronger interactions. The most usual anion is Cl^-. Other options are Br^-, OH^- or PO_4^{3-}, and organic ions such as acetate, salicylate or tartrate. Concomitantly, with the improvement in the peak shape, the adsorbed amine decreases retention. The presence of an anion with adsorption properties on the stationary phase can affect its separation properties.

Another option is to use a suitable combination of two counterions of opposite charge in the mobile phase, such as an alkylsulphonate and an amine. Whereas the alkylsulphonate acts as an IIC reagent, the organic amine masks the residual silanols, yielding an efficient separation within a reasonable time.

6.4.5. *Use of perfluorinated carboxylate anions and chaotropic ions as additives*

Ionisation of silanols and carboxylic groups in amino acids, peptides, proteins and other zwitterionic compounds of biochemical relevance, can be suppressed at low pH. However, this may give rise to early elution (and poor resolution), unless anionic reagents, such as alkylsulphonates or perfluorinated carboxylates, are added. Alkylsulphonates may, however, associate to the stationary phase, making column regeneration difficult. Therefore, perfluorinated carboxylates, which are volatile and thus compatible with ELS and MS detection and suitable for preparative chromatography, are preferable. Among these, trifluoroacetic acid is most commonly used due to its high purity, water solubility, and transparency at 220 nm. Other less common volatile perfluorinated acids are pentafluoropropionic acid and heptafluorobutyric acid.

In addition to perfluorinated carboxylates, other anions (mostly inorganic) are appropriate to separate zwitterions and basic compounds in the low pH region. Longer retention and enhanced peak symmetry are obtained with anions with a less localised charge, higher polarisability, and lower degree of hydration, with the following trend (called the Hofmeister series): $PF_6^- > ClO_4^- > BF_4^- > CF_3COO^- > NO_3^- > Cl^- > CH_3SO_3^- > HCOO^- > H_2PO_4^-$. More lipophilic anions can exhibit performance similar to traditional amphiphilic anions, but with fewer drawbacks. The mechanism of retention for the most hydrophilic anions is not clear, as their adsorption capability is small. This has been explained by considering that basic cationic analytes are usually well-solvated by the aqueous mobile phase and have little affinity for the lipophilic phase. However, they can interact in the mobile phase with the anionic additives to form an ion pair, which produces disruption of the solvation shell. As the ion pair is more lipophilic than the unpaired analyte, it is

more strongly retained by the stationary phase. The ability to increase the disorder of water is called chaotropicity (or chaotropic effect), that depends on the position of the anion in the Hofmeister series. This effect also explains the influence of the nature of buffers on retention.

6.4.6. Use of ILs as additives

Only the anion or the cation is adsorbed on the stationary phase for IIC reagents such as sodium hexanesulphonate and tetrabutylammonium hydroxide. In contrast, reagents such as hexylamine salicylate, butylammonium phosphate, or ILs have a dual character (both cation and anion are adsorbed), which creates a bilayer, positively or negatively charged, depending on the relative strength of the adsorption of cation and anion, respectively (see Figure 6.2c). ILs have many excellent characteristics, such as low volatility, high stability, good solubility and a wide range of structures. Although they are known mainly as green solvents, they behave in RPLC just like dissociated salts [25,26]. ILs are water-stable, soluble in typical RPLC solvents, and at small concentration, the mobile phase viscosity is not altered drastically. Meanwhile, several kinds of intermolecular interactions of ILs (hydrophobic, electrostatic, and other specific interactions with the stationary phase and analytes) are kept.

Most reported applications have been focused on ILs with a large imidazolium (or pyridinium) cation and BF_4^-, PF_6^-, Cl^- or Br^- as anion. The IL cation can interact through specific electrostatic interactions with the silanols on the alkyl-bonded silica surface, competing with the polar group of basic analytes. At the same time, different alkyl groups on the heterocyclic ring or quaternary cation in the IL can interact with the non-polar alkyl groups of the stationary phase through hydrophobic and other unspecific interactions. The observed retention behaviour and peak shape (peak tailing and band

broadening), with resolution enhancements, are a combination of the silanol-masking effect of the cation with the chaotropic character of the anion.

The relative adsorption of the anion and cation of an IL are useful to adjust the selectivity. If the compounds elute too rapidly, an IL with a lyotropic anion, such as PF_6^-, BF_4^-, or ClO_4^-, can be used with a short alkyl-chain imidazolium cation, such as 1-ethyl- or 1-butyl-3-methylimidazolium. If the compounds are highly retained, a long alkyl-chain cation such as 1-hexyl-3-methylimidazolium (the solubility of 1-octyl-3-methylimidazolium is too low to be practical) with an anion of low lyotropy such as Cl^- must be used. If there is no problem with the retention, ILs containing 1-hexyl-3-methylimidazolium and BF_4^- or Cl^- are recommended as the first choice. ILs containing a stronger chaotropic anion, such as PF_6^-, yield excessive retention. A number of IL-based stationary phases with interesting properties have been prepared for the separation of various compounds [27].

6.4.7. Measurement of the enhancement of column performance using additives

Extremely narrow signals would give rise to maximal information quality in RPLC, but owing to solute dispersion the signals are peaks with diverse widths and asymmetries (non-Gaussian peaks are quite common in practice). Peak variance results from several factors with two origins: extra-column contributions (dispersion in the tubing, unions, and detector cell) and column (diffusion and interaction with the support and stationary phase). The magnitude of the latter contributions depends on the column geometry; substrate properties; and the type of interactions among solutes, stationary phase, and mobile phase (i.e., secondary equilibria).

System performance can be conveniently visualised through the correlations between the peak half-widths and the retention times of analytes [28]. For

isocratic elution, the plots are nearly linear. They can be obtained with the half-widths/retention time data for a set of analytes experiencing similar kinetics, eluted with a mobile phase of fixed or varying composition (if the kinetics is not modified). The half-width plots approach is a simple tool that facilitates the characterisation of chromatographic columns.

Different studies have shown that, likely, bulky additives do not interact directly with free silanols by direct association, but the observed effect with basic analytes is produced by coating of the stationary phase with the additive. On the contrary, small additives may block silanol groups by direct electrostatic interaction, but this masking mechanism seems to be less effective. The larger the cation and its adsorption capability, the more intense is the masking of the silanol effect (i.e., the better the peak shape). Meanwhile, the specific nature of the additive does not seem to influence the peak shape. Thus, for instance, the benefits obtained in the presence of amines may be similar or even superior to those obtained in the presence of some ILs used as additives in RPLC [24]. To illustrate this behaviour, several half-width plots corresponding to amines and ILs are depicted in Figure 6.3. In the absence of additive, the slope of the right half-width is significantly larger with regard to the left half-width, which indicates tailing peaks. The three additives (cycloheptylamine, *N,N*-dimethyl-octylamine, and 1-hexyl-3-methylimidazolium), especially the latter two, enhanced the peak shape (see Figures 6.3c and d). This suggests that these additives efficiently hinder access of the basic drugs to the silanols on the column.
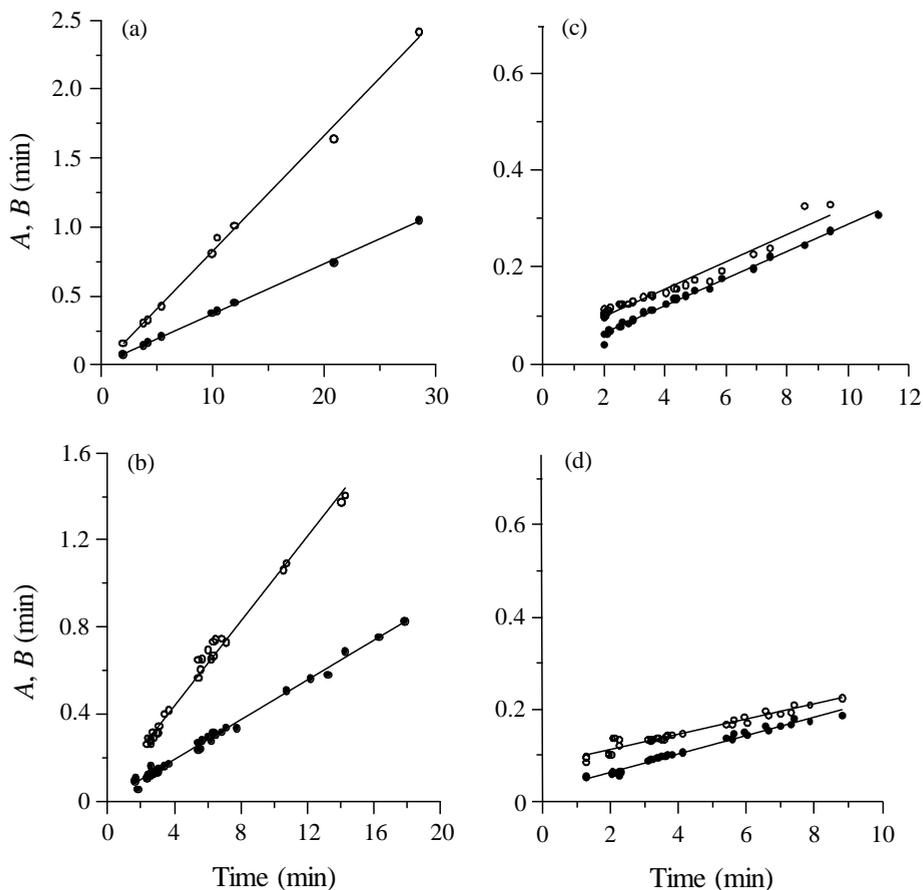
Figure 6.3. Half-widths plots (*A*, left half-width (●) and *B*, right half-width (○)), including the data obtained with nine basic drugs (β-adrenoceptor antagonists). The RPLC mobile phases contained 15% acetonitrile without additives (a), and different amounts of the additives cycloheptylamine (b), *N,N*-dimethyl-octylamine (c), and 1-hexyl-3-methylimidazolium (d).

6.5. Micellar liquid chromatography

6.5.1. An additional secondary equilibrium in the mobile phase

Above a certain concentration of an IIC reagent in the mobile phase, the stationary phase becomes saturated and more reagent remains in the mobile phase. Beyond this threshold, the retention, instead of further increasing, decreases progressively due to a number of secondary effects, such as the displacement of the adsorbed analyte by the IIC counterion, the formation of ion pairs between the analyte and IIC counterion in the mobile phase, or in the case of surfactants, the interaction with dynamic aggregates called "micelles", which are formed above the so-called critical micelle concentration (CMC) [29–31]. Micelles behave as a new phase (a pseudophase) within the mobile phase, which leads into the field of another RPLC mode, named "micellar liquid chromatography" (MLC), (see Figures 6.2d and e). MLC is classified among the pseudophase liquid chromatographic modes, where the mobile phase contains entities that interact with the analytes, such as micelles, cyclodextrins, vesicles, or nanometre-sized oil droplets in oil-in-water microemulsions.

MLC has had more impact than other pseudophase modes. Its unique selectivity is attributed to the ability of micelles to organise solutes at the molecular level. However, the association between the surfactant monomers and the bonded phase (forming a structure similar to the micelle surface) has deep implications with regard to retention and selectivity (see Figure 6.4). The amount of adsorbed surfactant remains constant or is near saturation above the CMC, which is an important feature with regard to robustness. Analytes are separated on the basis of their differential partitioning between the bulk aqueous phase and the micellar aggregates or the surfactant-coated stationary

phase. Therefore, a secondary equilibrium is added to the mobile phase, which can be altered for ionisable compounds by tuning the pH, as shown in Figures 6.1b, c, e, and f. Insoluble species partition via direct transfer from the micelles to the surfactant-modified stationary phase.

Surfactants with ionic, zwitterionic, and non-ionic head groups can be used to separate ionic or neutral analytes that are able to interact with the surfactant. The steric factor can also be important. The anionic sodium dodecyl sulphate (SDS) is by far the most common surfactant in MLC, used in two thirds of the reports, followed by the cationic cetyltrimethylammonium bromide (CTAB) and the non-ionic polyoxyethylene-(23)-dodecyl ether (Brij-35) [29,30]. Brij-35 is also applied to emulate in vitro the partitioning process in biomembranes in a mode called "biopartitioning MLC" [32]. The polar hydrophilic head of the Brij-35 molecule (the polyoxyethylene chain with the hydroxyl end group, which is oriented away from the surface of the stationary phase) increases the polarity of the stationary phase, which remains neutral. The hydroxyl end group of Brij-35 can also interact with polar or moderately polar solutes by formation of hydrogen bonds with hydroxyl and amino groups, increasing their retention [33].
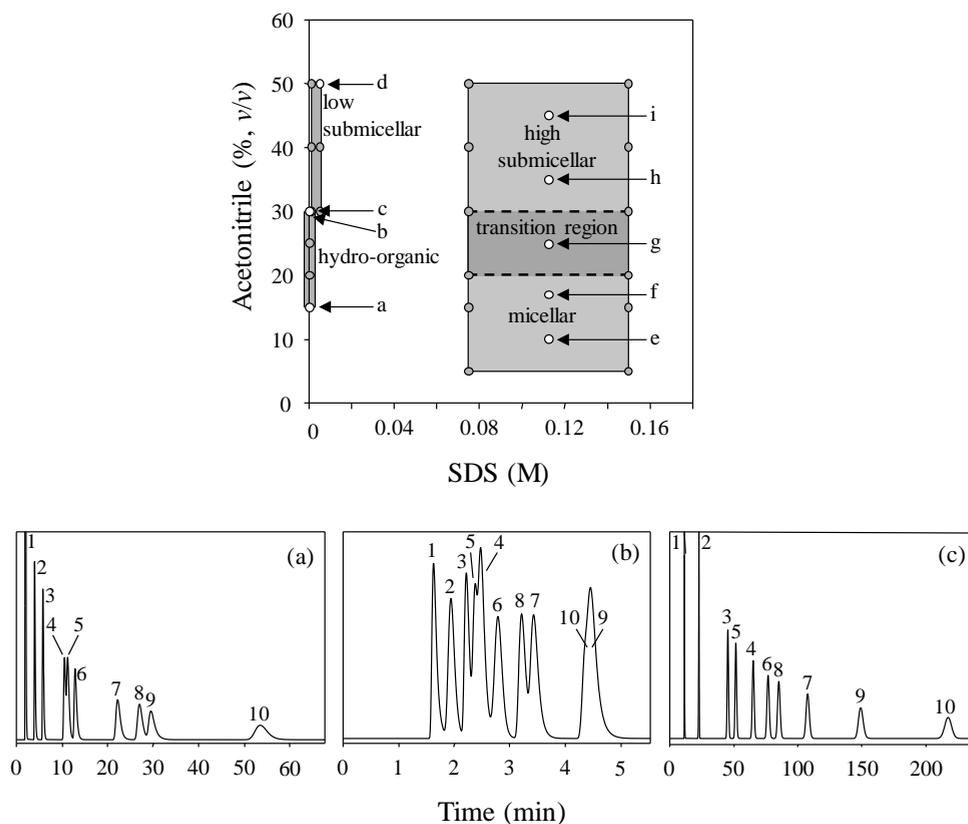
Figure 6.4. Chromatographic performance for mobile phases containing acetonitrile or acetonitrile and SDS, using a C18 Kromasil column. Top: Mobile phase compositions. Bottom: Chromatograms for: (a) 15% acetonitrile, (b) 30% acetonitrile, (c) 30% acetonitrile/0.001 M SDS, (d) 50% acetonitrile/0.005 M SDS, (e) 10% acetonitrile/0.1125 M SDS, (f) 17.5% acetonitrile/0.1125 M SDS, (g) 25% acetonitrile/0.1125 M SDS, (h) 35% acetonitrile/0.1125 M SDS, (i) 45% acetonitrile/0.1125 M SDS. Compounds: (1) atenolol, (2) carteolol, (3) pindolol, (4) timolol, (5) acebutolol, (6) metoprolol, (7) esmolol, (8) celiprolol, (9) oxprenolol, and (10) labetalol.
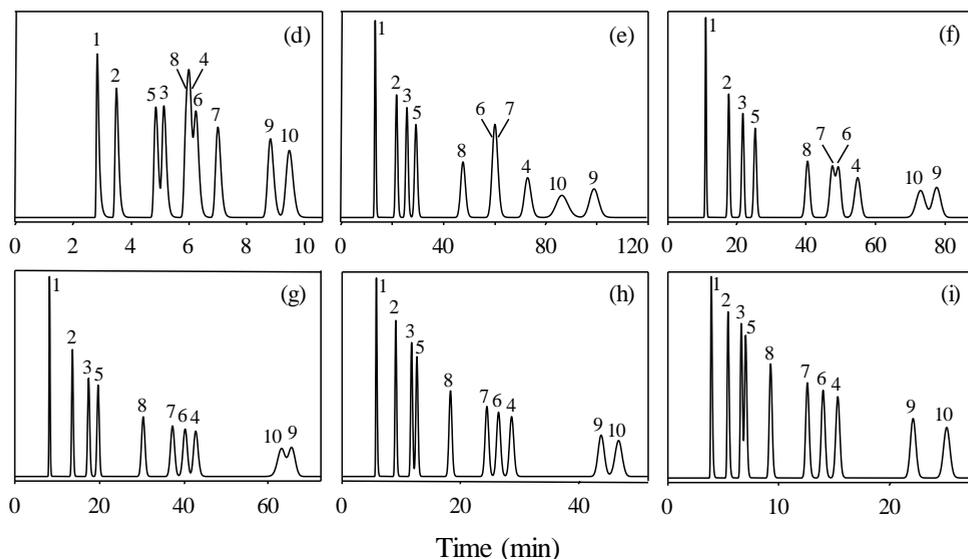
Figure 6.4 (continued).

6.5.2. *Hybrid micellar liquid chromatography*

The idea of using aqueous micellar solutions as mobile phases (i.e., only water and surfactant) is attractive, but suffers two main drawbacks compared to conventional RPLC: excessive retention of apolar compounds and poor efficiencies owing to the increased volume of stationary phase due to the adsorbed surfactant. This reduces the analyte mass transfer rate within the stationary phase. Propanol, butanol, pentanol or acetonitrile (especially propanol) are usually added to decrease the retention to practical values, giving rise to the so-called "hybrid MLC".

Acetonitrile, a common solvent in RPLC, has been scarcely used. Butanol and pentanol are chosen to elute strongly retained compounds. Equally important is that organic solvents reduce the amount of adsorbed surfactant in the stationary phase, enhancing the peak shape, which can be similar or even

improved with respect to conventional RPLC (see Figure 6.4). The highly symmetrical peaks obtained with SDS for basic drugs indicate that the ion-exchange mechanism with the sulphate group of the surfactant is a fast process and prevents the analyte penetration into the bonded alkyl chains to interact with the buried silanols. However, the attraction of the cationic compounds to the negatively charged stationary phase (by adsorption of the SDS anion) may significantly increase the retention. In contrast, the addition of Brij-35 to an organic mobile phase produces poor peak shape for basic drugs. However, the efficiency with Brij-35 has been shown to increase significantly with temperature, being close to that obtained with an acetonitrile-water eluent at 80 °C [33].

As commented, the anionic SDS requires the addition of an organic solvent to decrease the retention times and increase the efficiency, especially for basic drugs. Meanwhile, the non-ionic Brij-35 has the interesting feature of reducing the stationary phase polarity. This decreases the retention times significantly. However, the retention of polar compounds may be too short in the absence of specific interactions with Brij-35. An interesting solution is the preparation of mixed mobile phases of SDS and Brij-35 without organic solvent [34]. This gives rise to successful "green" RPLC procedures, yielding good resolution and adequate analysis times for basic drugs of intermediate polarity.

Although the separation mode with hybrid MLC is still predominantly micellar in nature, micelles are perturbed by the organic solvent, giving rise to changes in the CMC and the surfactant aggregation number. A high percentage of organic solvent is in principle undesirable, because of micelle disruption. The organic solvent concentration still preserving the integrity of micelles is approximately 15% for propanol and acetonitrile, 10% for butanol, and 6% for pentanol (the solubility of the two latter alcohols is significantly increased in

the surfactant medium). Organic-solvent-rich mobile phases can sweep out completely the adsorbed surfactant molecules from the bonded phase surface. However, a "submicellar RPLC" mode (with surfactant monomers in the mobile phase but without micelles), obtained at high concentration of surfactant and organic solvent, can yield good resolution and short analysis times (see Figures 6.4h and i) [35].

The most interesting features offered by MLC are the richness of interactions among solutes, stationary phase, aqueous phase, and micelles; the possibility of separating both charged and neutral solutes in a single run or analytes of different hydrophobicity in retention time windows narrower than in classical RPLC (making gradient elution less necessary); the high solubilisation capability of micelles, which facilitates dissolution of most matrices (saving time in sample preparation and enabling the direct on-column injection of physiological fluids); the low organic solvent concentration (translated in lower cost, toxicity, and environmental impact of wastes with regard to conventional RPLC); the smaller evaporation of organic solvents (making micellar phases stable for a longer time); and the enhanced luminescence detection, among others. The only real limitation is related to the use of ELS and MS detection, as direct on-line coupling is hindered by the presence of high concentrations of surfactant in the mobile phase.

6.5.3. Microemulsion liquid chromatography

Microemulsion liquid chromatography (MELC) is a relatively new chromatographic mode, which utilises oil-in-water microemulsions as the mobile phase [36]. These microemulsions consist of nanometre-sized droplets of a water immiscible liquid (e.g., ethyl acetate, octane, isopropyl ether, and cyclohexane) dispersed throughout an aqueous phase. The addition of a

surfactant (usually SDS) and a co-surfactant (a short chained alcohol, especially *n*-propanol and *n*-butanol) reduces the interfacial tension at the oil/water interface to almost zero, resulting in a stable system.

The high aqueous content of oil-in-water microemulsions makes them compatible with RPLC columns, whereas the hydrophobic oil core offers the ability to dissolve non-polar analytes and sample matrices. As in MLC, the stationary phase in MELC is modified by the adsorption of the surfactant. A secondary mechanism exists for analytes, which partition from both mobile phase and stationary phase into the microemulsion droplets. As the co-surfactant and the oil molecules can also be adsorbed on the stationary phase, analyte-solvent interactions in MELC are more complex with regard to MLC.

6.6. Metal complexation

6.6.1. Determination of metal ions

RPLC is a good alternative to direct spectroscopic methods and ion-exchange chromatography, being capable of determining several metals simultaneously, removing matrix interferences, coupling with different detectors, and enabling high sensitivity. The direct IIC separation of transition-metal ions is, however, difficult because of the similar hydration energies. The required selectivity is achieved using a number of secondary equilibria: complex formation, dynamic ion-exchange (and eventually ion pair or association with a micelle in the mobile phase), in addition to acid-base equilibria [37–39]. Neutral complexes are eluted with hydro-organic mixtures, but most frequently the complexes are anionic, and therefore, alkylammonium or tetraalkylammonium salts of a wide range of lipophilicities are used to retain

them in the IIC mode, with or without a competing anion in the mobile phase. Cationic surfactants such as CTAB or cetylpyridinium chloride can be also used below or above the CMC.

The separation of chelates with metallochromic ligands with highly absorptive chromophores dispenses the need of postcolumn derivatisation, with sub-µg/mL-level detection limits. A higher degree of selectivity and sensitivity can be achieved using fluorimetric reagents, which may reach ng/mL levels.

There are two main approaches: pre-column (off-line) formation of the complexes with subsequent separation, and injection of the metal ions and on-line formation with a ligand added to the mobile phase. The feasibility of these approaches depends on the stability of the complexes. Binary complexes are usually formed, with a few examples of ternary complexes to enhance both selectivity and sensitivity. Many chelating reagents (often previously used in spectrophotometric methods) are used, such as 8-hydroxyquinoline, 4-(2-pyridylazo)resorcinol, 1,10-phenanthroline, and several dithiocarbamates and azo dyes, which form stable neutral or ionic chelates with a number of metal ions, readily detected by spectrophotometry. In some cases, selectivity is improved by adding a second ligand to mask interferences and eliminate the corresponding peak. The integrity of metal chelates is susceptible to pH, as side reactions are expected at low pH with the ligand and at high pH with the metal ions. The narrow pH range of conventional columns may be unsuitable for complex formation.

Complete chelate separation from the excess reagent added at the off-line chelation step allows detection of the chelate in the absence of background contributions. When the complexation reaction is slow at room temperature, heating prior to injection may be needed. A selective and sensitive analysis is possible by combining off-line complexation with solvent extraction, which

also allows the analysis of neutral complexes. Poor water solubility of some chelates requires a mobile phase with a high proportion of organic solvent or a surfactant.

With off-line complexation, only thermodynamically or kinetically stable chelates survive during elution and reach the detector, as each chelate migrates separately from the ligand, resulting in a steep decrease in ligand concentration close to the chelate peaks. In these conditions, weak complexes tend to dissociate in the analytical column, typically through solvolysis or ligand-exchange reactions. This means that the column can work not only as a conventional separation device, but also as a powerful kinetic discriminator to selectively detect the chelates. The approach has been named "kinetic differentiation chromatography". Here, the synergic interactions of four origins of unique selectivity are combined: pre-column chelation, chromatographic separation, dissociation kinetics, and spectral selectivity.

Many chelates used to determine metal ions by spectrophotometry after solvent extraction are not sufficiently strong, and dissociate in the RPLC column. This can be prevented by a combination of the off- and on-line approaches (i.e., the injection of the complexes and the inclusion of the ligand in the mobile phase). Also, a strong chelating reagent can be useful for extraction of the metal ions in a sample, but not at all for an RPLC separation, due to the lack of selectivity or instability of the complexes at the separation conditions, or for detection. The ligand-exchange approach can solve this problem, replacing the first ligand with another added to the mobile phase.

A simpler approach is the direct injection of the metal ion (without previous extraction), which is complexed inside the column (on-line), in an approach called "dynamic chelating (or complexation) chromatography". The separation is based on a combination of ion-exchange and complexation selectivity, which

is provided by the strengths and reaction rates of the metal with the ligand and the IIC counterion in the mobile phase. Kinetic problems may be alleviated by thermostating the chromatographic column or using more suitable ligands.

Hydrophobic metallochromic ligands such as xylenol orange and methyl thymol blue can be used to coat an RPLC stationary phase, producing a chelating capacity to separate metal ions. Two approaches are possible: pre-coating the stationary phase with the ligand and elution with an inorganic salt, and inclusion of the ligand within the mobile phase to dynamically coat the stationary phase. The second approach allows an increased column capacity and stability, improved separation efficiency and selectivity, and the ability to exploit the ligand in the mobile phase for metal detection.

6.6.2. Determination of organic compounds

Metal cations can be used as well to modulate the selectivity in the separation of organic compounds by complex formation. There are two basic approaches: the introduction of the metal ions into the stationary phase or into the mobile phase. When metal ions are added as salts of weak complexing anions such as nitrate or perchlorate, the mobile phase should be acidic to avoid metal hydrolysis. Also, the column performance is often poor in terms of selectivity and peak shape. The addition of charged metal chelates (anionic or cationic depending on the analyte charge) to the mobile phase is a more versatile and simpler approach, which has shown enhanced performance in comparison with conventional IIC reagents, especially in the separation of amino acids (free or derivatised), peptides, and aromatic compounds. A ligand-exchange process may occur between the analyte and the ligands in the complexes. In some cases, the formation of ternary complexes has also been suggested. This process involves hydrophobic selectivity, but steric selectivity

can be quite high, connected with the conformationally rigid structures of the chelates, which act as templates.

The metal choice is a compromise between several factors, such as the ability to form complexes, solubility in the hydro-organic solvent, and detection. The general classification of transition metals according to their tendency to form complexes is as follows: $Pt^{4+} > Pd^{2+} > Hg^{2+} > Cu^{2+} > Ni^{2+} > Co^{2+} > Zn^{2+} > Cd^{2+} > Fe^{2+} > Mn^{2+} > Ag^+$ (inversions can occur depending on the ligands). Metal salts of Cu^{2+}, Ni^{2+}, Zn^{2+}, and Ag^+ are the most common. "Silver ion (or argentation) chromatography" is particularly applied to the analysis of lipids. However, the incorporation of Ag^+ into the solid support is preferred, as the RLPC mode has the disadvantage of using a mobile phase troublesome to handle.

6.7. Use of redox reactions

Finally, redox reactions may also be useful to enhance the separation selectivity of RPLC when the analytes exhibit redox behaviour. The redox reaction may occur on-column or on-line [40]. On-column derivatisation is assisted by the redox activity of the packing material, such as porous graphitic carbon or carbon manipulated using an electrochemically modulated liquid chromatographic technique. The analyte compound migrates in the column as a mixture of oxidised and reduced forms, so that their retention is determined by the relative concentration of the two forms inside the column (similarly to the acid-base species, Equation (6.2)). The on-line system consists of two separation columns with a redox derivatisation unit between them. The redox reaction proceeds rapidly in the derivatisation unit, so that the analyte migrates as its original form in the first column, while as its oxidised or reduced form in

the second column. The retention of the analytes is thus controlled by the lengths of the two separation columns in this system.

6.8. References

[1] M.C. García Álvarez-Coque, J.J. Baeza Baeza, G. Ramis Ramos, Reversed phase liquid chromatography, in *Analytical Separation Science Series*, Vol. 1 (edited by J.L. Anderson, A. Stalcup, A. Berthod, V. Pino), Wiley, New York, 2015, pp. 159–198.

[2] J. Nawrocki, The silanol group and its role in liquid chromatography, *J. Chromatogr. A* 779 (1997) 29–71.

[3] C. Horváth, Enhancement of retention by ion-pair formation in liquid chromatography with nonpolar stationary phases, *Anal. Chem.* 49 (1977) 2295–2305.

[4] C.H. Lochmüller, H.H. Hangac, Mobile phase additives vs. bonded phases for HPLC, *Crit. Rev. Anal. Chem.* 27 (1997) 27–48.

[5] J.P. Foley, W.E. May, Optimization of secondary chemical equilibria in liquid chromatography: Theory and verification, *Anal. Chem.* 59 (1987) 102–109.

[6] P.J. Schoenmakers, R. Tijssen, Modelling retention of ionogenic solutes in liquid chromatography as a function of pH for optimization purposes, *J. Chromatogr. A* 656 (1993) 577–590.

[7] U.D. Neue, C.H. Phoebe, K. Tran, Y.F. Cheng, Z. Lu, Dependence of reversed-phase retention of ionisable analytes on pH, concentration of organic solvent and silanol activity, *J. Chromatogr. A* 925 (2001) 49–67.

[8]   M. Rosés, E. Bosch, Influence of mobile phase acid-base equilibria on the chromatographic behaviour of protolytic compounds, *J. Chromatogr. A* 982 (2002) 1–30.

[9]   Ch. Zisi, S. Fasoula, A. Pappa Louisi, P. Nikitas, Properties of the retention time of ionizable analytes in reversed-phase liquid chromatography under organic modifier gradients in different eluent pHs, *J. Chromatogr. A* 1314 (2013) 138–141.

[10] A. Andrés, M. Rosés, E. Bosch, Prediction of the chromatographic retention of acid-base compounds in pH buffered methanol-water mobile phases in gradient mode by a simplified model, *J. Chromatogr. A* 1385 (2015) 42–48.

[11] M. Rosés, Determination of the pH of binary mobile phases for reversed-phase liquid chromatography, *J. Chromatogr. A* 1037 (2004) 283–298.

[12] A. Suu, L. Jalukse, J. Liigand, A. Kruve, D. Himmel, I. Krossing, M. Rosés, I. Leito, Unified pH values of liquid chromatography mobile phases, *Anal. Chem.* 87 (2015) 2623–2630.

[13] S. Pous Torres, J.R. Torres Lapasió, J.J. Baeza Baeza, M.C. García Álvarez-Coque, Combined effect of solvent content, temperature and pH on the chromatographic behaviour of ionisable compounds, *J. Chromatogr. A* 1163 (2007) 49–62.

[14] L.G. Gagliardi, M. Tascon, C.B. Castells, Effect of temperature on acid-base equilibria in separation techniques, *Anal. Chim. Acta* 889 (2015) 35–57.

[15] D.V. McCalley, Study of retention and peak shape in hydrophilic interaction chromatography over a wide pH range, *J. Chromatogr. A* 1411 (2015) 41–49.

[16] A. Bartha, G. Vigh, Z. Varga Puchony, Basis of the rational selection of the hydrophobicity and concentration of the ion-pairing reagent in reversed-phase ion-pair high-performance liquid chromatography, *J. Chromatogr.* 499 (1990) 423–434.

[17] T. Cecchi, Ion pairing chromatography, *Crit. Rev. Anal. Chem.* 38 (2008) 161–213.

[18] M.C. García Álvarez-Coque, G. Ramis Ramos, M.J. Ruiz Ángel, Liquid chromatography, ion pair, in *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering Series* (edited by J. Reedijk), Elsevier, Waltham, Massachusetts, 2015.

[19] J.G. Chen, S.G. Weber, L.L. Glavina, F.F. Cantwell, Electrical double-layer models of ion-modified (ion-pair) reversed-phase liquid chromatography, *J. Chromatogr. A* 656 (1993) 549–576.

[20] T. Cecchi, Theoretical models of ion pair chromatography: A close up of recent literature production, *J. Liq. Chromatogr. Relat. Technol.* 38 (2015) 404–414.

[21] M.C. Gennaro, S. Angelino, Separation and determination of inorganic anions by reversed-phase high-performance liquid chromatography, *J. Chromatogr. A* 789 (1997) 181–194.

[22] D.V. McCalley, The challenges of the analysis of basic compounds by high performance liquid chromatography: Some possible approaches for improved separations, *J. Chromatogr. A* 1217 (2010) 858–880.

[23] J.S. Kiel, S.L. Morgan, R.K. Abramson, Effects of amine modifiers on retention and peak shape in reversed-phase high-performance liquid chromatography, *J. Chromatogr. A* 320 (1985) 313–323.

[24] S. Calabuig Hernández, M.C. García Álvarez-Coque, M.J. Ruiz Ángel, Performance of amines as silanol suppressors in reversed-phase liquid chromatography, *J. Chromatogr. A* 1465 (2016) 98–106.

[25] J.J. Fernández Navarro, M.C. García Álvarez-Coque, M.J. Ruiz Ángel, The role of the dual nature of ionic liquids in the reversed-phase liquid chromatographic separation of basic drugs, *J. Chromatogr. A* 1218 (2011) 398–407.

[26] M.C. García Álvarez-Coque, M.J. Ruiz Ángel, A. Berthod, S. Carda Broch, On the use of ionic liquids as mobile phase additives in high-performance liquid chromatography, *Anal. Chim. Acta* 883 (2015) 1–21.

[27] V. Pino, A.M. Afonso, Surface-bonded ionic liquid stationary phases in high-performance liquid chromatography, *Anal. Chim. Acta* 714 (2012) 20–37.

[28] J.J. Baeza Baeza, M.J. Ruiz Ángel, S. Carda Broch, M.C. García Álvarez-Coque, Half-width plots, a simple tool to predict peak shape, reveal column kinetics and characterise chromatographic columns in liquid chromatography: State of the art and new results, *J. Chromatogr. A* 1314 (2013) 142–153.

[29] A. Berthod, M.C. García Álvarez-Coque, *Micellar Liquid Chromatography*, Marcel Dekker, New York, 2000.

[30] M.J. Ruiz Ángel, M.C. García Álvarez-Coque, A. Berthod, New insights and recent developments in micellar liquid chromatography, *Sep. Purif. Rev.* 38 (2009) 45–96.

[31] M.C. García Álvarez-Coque, M.J. Ruiz Ángel, S. Carda Broch, Micellar Liquid Chromatography: Fundamentals, in *Analytical Separation Science Series*, Vol. 2 (edited by J.L. Anderson, A. Stalcup, A. Berthod, V. Pino), Wiley-VCH, New York, 2015, pp. 371–406.

[32] L. Escuder Gilabert, S. Sagrado, R.M Villanueva Camañas, M.J. Medina Hernández, Quantitative retention-structure and retention-activity relationship studies of local anesthetics by micellar liquid chromatography, *Anal. Chem.* 70 (1998) 28−34.

[33] J.J. Baeza Baeza, Y. Dávila, J.J. Fernández Navarro, M.C. García Álvarez-Coque, Measurement of the elution strength and peak shape enhancement at increasing modifier concentration and temperature in RPLC, *Anal. Bioanal. Chem.* 404 (2012) 2973–2984.

[34] M.J. Ruiz Ángel, E. Peris García, M.C. García Álvarez-Coque, Reversed-phase liquid chromatography with mixed micellar mobile phases of Brij-35 and sodium dodecyl sulphate: A method for the analysis of basic compounds, *Green Chem.* 17 (2015) 3561–3570.

[35] M.J. Ruiz Ángel, S. Carda Broch, M.C. García Álvarez-Coque, High submicellar liquid chromatography, *Sep. Purif. Rev.* 43 (2014) 124–154.

[36] A. Marsh, B.J. Clark, K.D. Altria, A review of the background, operating parameters and applications of microemulsion liquid chromatography, *J. Sep. Sci.* 28 (2005) 2023–2032.

[37] P. Wang, H.K. Lee, Recent applications of high-performance liquid chromatography to the analysis of metal complexes, *J. Chromatogr. A* 789 (1997) 437–451.

[38] C. Sarzanini, Liquid chromatography: A tool for the analysis of metal species, *J. Chromatogr. A* 850 (1999) 213–228.

[39] B. Paull, P.R. Haddad, Chelation ion chromatography of trace metal ions using metallochromic ligands, *Trends Anal. Chem.* 18 (1999) 107–114.

[40] M. Shibukawa, K. Saitoh, S. Ozaki, H. Nakajima, Redox derivatization chromatography, *Bunseki Kagaku* 62 (2013) 985–1000.

CHAPTER 7

INTERPRETIVE SEARCH OF OPTIMAL ISOCRATIC AND GRADIENT SEPARATIONS IN MICELLAR LIQUID CHROMATOGRAPHY IN EXTENDED ORGANIC SOLVENT DOMAINS

7.1. Abstract

Micellar liquid chromatography (MLC) is a reversed-phase mode with aqueous mobile phases containing an organic solvent and a micellised surfactant. Most procedures developed in MLC are implemented in the isocratic mode, since the general elution problem in chromatography is less troublesome. However, gradient elution may be still useful to analyse mixtures of compounds within a wide range of polarities, in shorter times. MLC using gradients is also attractive to determine moderate to low polar compounds in physiological samples by direct injection. In these analyses, the use of initial micellar conditions (isocratic or gradient) with a fixed amount of surfactant above the critical micellar concentration, keeping the organic solvent content low, will provide better protection to the column against the precipitation of the proteins in the physiological fluid. Once the proteins are swept away, the elution strength can be increased using a positive gradient of organic solvent to reduce the analysis time. This may give rise to the transition from the micellar to the submicellar mode, since micelles are destroyed at sufficiently high concentration of organic solvent. In this work, several retention models covering extended solvent domains in MLC are developed, tested, and applied to investigate the performance in isocratic, linear and multi-linear gradient separations. The study was applied to the screening of β-adrenoceptor antagonists in urine samples, using mobile phases prepared with sodium dodecyl sulphate and 1-propanol. Predicted chromatograms were highly accurate in all situations, although suffered of baseline problems and minor shifts for peaks eluting close to a steep gradient segment. Two columns (C18 and C8) were investigated, the C8 column being preferable owing to the smaller amount of adsorbed surfactant.

7.2. Introduction

Micellar liquid chromatography (MLC) is a reversed-phase liquid chromatographic (RPLC) mode, where the mobile phase contains an ionic or neutral surfactant above the critical micellar concentration (CMC). The low elution strength and poor peak properties of aqueous solutions containing only surfactant force the addition of small amounts of an organic solvent to enhance the chromatographic performance [1–3]. In MLC systems, the stationary phase is covered with a layer of surfactant monomers, whose hydrophobic tail is associated to the alkyl chains bonded to silica. This results in the formation of a stable modified stationary phase [4,5], with a behaviour neatly different from an uncoated alkyl-bonded phase. The excess of free surfactant monomers in the mobile phase is arranged in small clusters or micelles [1,2]. The presence of organised surfactant structures in both phases leads to important changes in the chromatographic properties (retention time, selectivity and efficiency), with regard to classical RPLC, particularly in the analysis of ionisable solutes using mobile phases containing ionic surfactants [6–10]. However, the most important advantage of micellar mobile phases is the possibility of performing the direct injection of physiological samples into the column, without protein precipitation and subsequent column clogging [11].

In hybrid MLC (with surfactant and organic solvent), the separation is based on the existence of different distribution equilibria between the solute and the modified stationary phase, hydro-organic eluent and micelles, and in the case of submicellar conditions, the formation of ionic pairs between the solute and free surfactant monomers [1,12]. Most analytical procedures in MLC have been developed in the isocratic mode. In spite of the benefits of isocratic elution, the use of gradients has shown be useful for reducing the analysis time [13–25]. In most reports, linear gradients of organic solvent (acetonitrile, 1-propanol or

1-butanol) are used, keeping constant the surfactant concentration (mainly sodium dodecyl sulphate, SDS, or polyoxyethylene(23)lauryl ether, known as Brij-35).

Micelles are unstable at high organic solvent concentration. Thus, when high organic solvent contents are reached during the gradient, micelle disruption may occur [12]. In addition, the amount of adsorbed surfactant on the stationary phase is reduced. These conditions have given rise to the so-called high submicellar liquid chromatography (HSLC). In this mode, retention is decreased and peak shape improved with respect to MLC and classical RPLC [12,20,26]. However, a gradient of organic solvent where high contents are reached can be incompatible with the analysis of physiological fluids, due to the precipitation of proteins [11]. To avoid this, the protein front should be swept off the column under pure micellar conditions, or be eluted at low organic solvent in the initial region of the gradient [20,22,23]. After the elution of the proteins, the organic solvent content can be freely increased to get proper elution of the most retained compounds, reaching even submicellar conditions.

Most reported methods involving gradient elution in MLC have been optimised by trial and error. This strategy is only valid for simple samples, and often fails in offering fair separations. Instead, interpretive strategies make an exhaustive inspection of the optimal conditions, and are more reliable and efficient [27–29]. These strategies are based on the use of mathematical equations (i.e., models) that describe the chromatographic behaviour. In a first step, information about retention and peak properties is gathered according to a pre-established experimental design that can involve different factors, from which models are fitted. The predictive capability depends on the quality of the information provided, which may come from isocratic, gradient or mixed experimental data. In a second step, the fitted models are used to conduct a

supervised search of the optimal separation. This implies the prediction of the performance under a high number of experimental conditions, defined usually on a grid basis in either isocratic or gradient modes.

In previous work [22], a commercial software application (Drylab) was used to find the best linear gradient conditions to separate a mixture of β-adrenoceptor antagonists (βAAs), using an interpretive optimisation protocol. These compounds are basic and show high retention in MLC with SDS, owing to the attraction of the cationic species (formed at the usually acidic mobile phase pH) towards the adsorbed layer of surfactant on the column. In order to obtain chromatograms with practical analysis times, gradient elution with an organic solvent reaching relatively high concentrations, where micelles are disrupted, is needed. Drylab was designed to optimise gradients of organic solvent in classical RPLC. It was found that the presence of surfactant in both stationary and mobile phases caused deviations in the predictions. The optimisation of the surfactant concentration was also not possible.

In the current work, we studied the feasibility of gradient separations, in an extended organic solvent range covering micellar and high submicellar conditions, when a physiological sample is directly injected in the chromatographic system. The effect of linear and multi-linear gradients on the chromatogram baseline is evaluated for C18 and C8 columns. The separation conditions (i.e., organic solvent and surfactant contents, gradient complexity, and analysis time) are evaluated using an interpretive methodology and application software designed in our laboratory. Finally, the advantages and disadvantages of using isocratic, and linear or multi-linear gradients, in pure, hybrid micellar and high submicellar conditions, are studied for achieving complete resolution. Before performing the optimisation study, the prediction

performance of several retention models, some of them proposed in this work, was evaluated.

7.3. Theory

A large variety of retention models accounting the presence of surfactant and organic solvent have been proposed to describe the chromatographic behaviour in micellar and submicellar conditions [10,30–32]. This section introduces the main models described in the literature for these chromatographic modes, and proposes some modifications to extend their validity domain. In principle, gradients of both surfactant and organic solvent are possible, but gradients of organic solvent offer better performance in practice. The reliability of gradient optimisation in MLC and HSLC requires isocratic predictions as accurate as possible for the inclusion of the retention models in the fundamental equation of gradient elution. For this reason, part of this work is dedicated to the improvement of predictions in extended organic solvent ranges. Some new models are proposed and the significance of the parameters is statistically evaluated.

For the proposal of new models in MLC and HSLC, it is interesting to revise the proposals carried out in classical RPLC to forecast the retention against variations in the organic solvent content. Over the years, different models have been proposed based on thermodynamic considerations, together with other models with a more or less empirical nature [33]. In the best cases, prediction errors of 1–2% are obtained (exceptionally, below 1%) [34,35]. These errors depend strongly on the type of equation, solute nature, existence of additional equilibria, and magnitude of the variations in organic solvent. Amongst the reported models, the logarithmic-quadratic relationship proposed by Schoenmakers [36], and the Snyder's linear simplification [37], which relate

the logarithm of the retention factor (k) with the volume fraction of organic solvent in the mobile phase (φ), are extensively used in optimisation strategies:

$$\ln k = \ln \frac{t_R - t_0}{t_0 - t_{ext}} = \ln k_w - S\,\varphi + T\,\varphi^2 \approx \ln k_w - S\,\varphi \tag{7.1}$$

where t_R, t_0 and t_{ext} are the retention time, dead time, and extra-column time, respectively, $\ln k_w$ is the logarithm of the retention factor when pure water is used as eluent, S measures the elution strength of the solvent, and T accounts for deviations from linearity of $\ln k$ versus φ. Later, Schoenmakers et al. proposed the inclusion of a square root term to further improve predictions at low modifier concentrations [38]:

$$\ln k = \ln k_w - S\,\varphi + T\,\varphi^2 + U\,\sqrt{\varphi} \tag{7.2}$$

Coefficient U depends on the polarity of the stationary phase. The reciprocal of the retention factor has been also proposed in some RPLC retention models, instead of $\ln k$. An example is a model proposed by Lee et al. [39].

$$k = c_0 + \frac{c_1}{\varphi} + \frac{c_2}{\varphi^2} \approx c_0 + \frac{c_1}{\varphi} \tag{7.3}$$

where c_{0-2} are model parameters.

In MLC, mechanistic models have been proposed, most of them being reciprocal, although logarithmic models have also been reported [30–32]. For instance, the following mechanistic model, which describes the retention with experimental errors usually below 2%, is based on a reciprocal functionality:

$$k = \frac{K_{AS} \dfrac{1 + K_{SD}\,\varphi}{1 + K_{AD}\,\varphi}}{1 + K_{AM} \dfrac{1 + K_{MD}\,\varphi}{1 + K_{AD}\,\varphi}[M]} \tag{7.4}$$

where $[M]$ is the concentration of surfactant involved in micelle formation, K_{AS} and K_{AM} are constants related to the solute-stationary phase and solute-micelle distribution equilibria, respectively, and K_{SD}, K_{AD} and K_{MD} quantify the shifts of the solute distribution equilibria with the addition of organic solvent, towards the stationary phase (K_{SD}), bulk mobile phase (K_{AD}), and micelle (K_{MD}). The K_{SD} coefficient is only significant for compounds of low polarity and can be neglected otherwise, which in practice happens very often [40,41]. Thus, for solutes of low or intermediate hydrophobicity, Equation (7.4) can be reformulated assuming that $K_{SD} \approx 0$, which gives rise to the following simplified model where the reciprocal functionality is evident:

$$
\begin{aligned}
\frac{1}{k} &= \frac{1}{K_{AS}}(1 + K_{AD}\,\varphi) + \frac{K_{AM}}{K_{AS}}(1 + K_{MD}\,\varphi)[M] = \\
&= c_0 + c_1\,\varphi + c_2\,[M] + c_3\,\varphi\,[M]
\end{aligned}
\tag{7.5}
$$

In situations where the organic solvent content is kept fixed, or when the mobile phase is purely micellar, Equation (7.5) can be simplified after rearranging the terms:

$$
\frac{1}{k} = \frac{1 + K_{AM}\,[M]}{K_{AS}} = \frac{1}{K_{AS}} + \frac{K_{AM}}{K_{AS}}[M] = c_0 + c_1\,[M]
\tag{7.6}
$$

In previous work [10], an extension of Equation (7.4) (without K_{SD}) was proposed to account for both MLC and HSLC conditions (i.e., from low to high concentrations of organic solvent):

$$
k = \frac{K_{AS}\,\dfrac{1}{1 + K_{AD}\,\varphi}}{1 + K_{AM}\,\dfrac{1 + K_{MD}\,\varphi}{1 + K_{AD}\,\varphi}[S] + K_{\varphi}\,\varphi^2}
\tag{7.7}
$$

Note that in Equation (7.7), the concentration of total surfactant monomers [*S*] (forming micelles or not) is used instead of the micellised surfactant [*M*]. Accordingly, K_{AM} and K_{MD} are referred to the interaction of solutes with the surfactant monomers. The term φ^2 has been added to account for the larger impact of the organic solvent in the mobile phase when the solvent domain is extended. Rearranging the terms in Equation (7.7), the following is obtained:

$$\frac{1}{k} = \frac{1}{K_{AS}}(1 + K_{AD}\,\varphi) + \frac{K_{\varphi}}{K_{AS}}(1 + K_{AD}\,\varphi)\,\varphi^2 + \frac{K_{AM}}{K_{AS}}(1 + K_{MD}\,\varphi)[S]$$

$$= c_0 + c_1\,\varphi + c_2\,\varphi^2 + c_3\,\varphi^3 + c_4\,[S] + c_5\,\varphi\,[S] \qquad (7.8)$$

A simplification can be made by neglecting the cubic term:

$$\frac{1}{k} = \frac{1}{K_{AS}}(1 + K_{AD}\,\varphi) + \frac{K_{\varphi}}{K_{AS}}\varphi^2 + \frac{K_{AM}}{K_{AS}}(1 + K_{MD}\,\varphi)[S]$$

$$= c_0 + c_1\,\varphi + c_2\,\varphi^2 + c_3\,[S] + c_4\,\varphi\,[S] \qquad (7.9)$$

Another useful simplification of practical interest concerns situations where the surfactant concentration is constant (e.g., gradients of organic solvent at constant surfactant concentration), which requires less experimental effort:

$$\frac{1}{k} = c_0 + c_1\,\varphi + c_2\,\varphi^2 + c_3\,\varphi^3 \qquad (7.10)$$

In this work, an extended model based on Equations (7.2) and (7.8) is also evaluated:

$$\frac{1}{k} = c_0 + c_1\,\varphi + c_2\,\varphi^2 + c_3\,\varphi^3 + c_4\,[S] + c_5\,\varphi[S] + c_6\,\sqrt{\varphi}\,[S] \qquad (7.11)$$

If the surfactant concentration remains constant, the model can be simplified to:

$$\frac{1}{k} = c_0 + c_1 \varphi + c_2 \varphi^2 + c_3 \varphi^3 + c_4 \sqrt{\varphi} \qquad (7.12)$$

which is analogous to Equation (7.2) when the surfactant concentration is constant. Other possible simplifications of Equation (7.11), with five or six coefficients, are:

$$\frac{1}{k} = c_0 + c_1 \varphi + c_2 \varphi^2 + c_3 [S] + c_4 \varphi [S] \qquad (7.13)$$

$$\frac{1}{k} = c_0 + c_1 \varphi + c_2 \varphi^3 + c_3 [S] + c_4 \varphi [S] \qquad (7.14)$$

$$\frac{1}{k} = c_0 + c_1 \varphi + c_2 [S] + c_3 \varphi [S] + c_4 \sqrt{\varphi} [S] \qquad (7.15)$$

$$\frac{1}{k} = c_0 + c_1 \varphi + c_2 \varphi^2 + c_3 [S] + c_4 \varphi [S] + c_5 \sqrt{\varphi} [S] \qquad (7.16)$$

$$\frac{1}{k} = c_0 + c_1 \varphi + c_2 \varphi^3 + c_3 [S] + c_4 \varphi [S] + c_5 \sqrt{\varphi} [S] \qquad (7.17)$$

7.4. Experimental

7.4.1. Reagents

The following set of βAAs was considered: (1) atenolol, (2) carteolol, (3) nadolol, (4) acebutolol, (5) metoprolol, (6) oxprenolol, (7) propranolol, and (8) alprenolol, all of them from Sigma (Saint Louis, MO, USA). Stock solutions containing 100 μg/mL of each drug were prepared in 10% (*v/v*) 1-propanol from Scharlau (Sentmenat, Barcelona, Spain), assisted with an ultrasonic bath Elmasonic (Singen, Germany), and stored at 4ºC. Working

standard solutions of 20 μg/mL were obtained by dilution of the stock solutions in nanopure water, obtained from an Adrona B30 trace purification system (Burladingen, Germany).

Isocratic mobile phases and gradients contained sodium dodecyl sulphate from Merck (99% purity, Darmstadt, Germany) and 1-propanol. The pH was set at 3.0 by addition of 0.01 M anhydrous sodium dihydrogen phosphate from Fluka (Steinheim, Germany), and the appropriate amount of 0.1 M HCl and NaOH from Scharlau. Gradient elution was performed by combining Solvent A containing 5% 1-propanol, and Solvent B containing 35% 1-propanol, both with the same amount of SDS.

All reagents were of analytical grade or better. Before injection in the chromatographic system, the solutions were filtered through a 0.45 μm Nylon membrane from Micron Separation (Westboro, MA, USA).

7.4.2. Apparatus, columns and experimental design

The chromatographic analyses were performed with an HP1200 Agilent instrument (Waldbronn, Germany), composed of quaternary pump, autosampler equipped with 2 mL vials, thermostated column compartment, and UV-Vis detector set at 225 nm, making duplicate injections. The flow rate was kept constant at 1.0 mL/min, and the injection volume was 20 μL. Column temperature was set at 25ºC. The dead time was measured by injection of KBr from Acros Organics (Fair Lawn, NJ, USA), monitored at 210 nm. The system dwell time (1.16 min) was determined by removing the column and using an acetone gradient. The instrumental extra-column contribution (0.12 min) was evaluated after removing the column by elution of metoprolol with a mobile phase containing 0.10 M SDS and 5% 1-propanol.

Chromatographic elution was carried out in the isocratic and gradient modes, using two analytical columns: Zorbax Eclipse XDB-C18 and Zorbax Eclipse XDB-C8 (150×4.6 mm, 5 μm particle size) from Agilent. The retention behaviour of the βAAs and a urine endogenous compound, which yielded a prominent peak, was studied using isocratic experiments, at different levels of SDS and 1-propanol. The assayed concentrations were: 0.05 and 0.15 M SDS containing 0, 5, 15, 25, or 35% (v/v) 1-propanol, and 0.10 M SDS containing 0, 10, 20, or 30% (v/v) 1-propanol. For the C18 column, an additional experimental point in the design (0.10 M SDS and 5% 1-propanol) was carried out.

7.4.3. Software

Experimental data were acquired with an OpenLAB CDS LC ChemStation (Agilent B.04.03). Peak properties (retention times and peak half-widths) were measured with the MICHROM software [42]. Mathematical treatment and data processing were performed using Matlab 2017b (The MathWorks Inc., Natick, MA, USA) and Visual Basic (Microsoft Co., Redmond, WA, USA).

7.5. Results and discussion

One of the aims of this work was exploring the benefits and disadvantages associated to the use of the two most common stationary phases (C18 and C8) in MLC and HSLC, under gradient conditions. Due to their different carbon load, these columns have a differentiated surfactant adsorption capability, and therefore, they can be expected to suffer in a different extent the effects of an organic solvent gradient, with regard to surfactant desorption along the gradient. A parallel presentation of the results will be done for both columns. It

should be noted, however, that the C18 column was first investigated, and some decisions are explained by the order in which the assays were performed.

The optimisation of the screening conditions for the eight βAAs in MLC was carried out based on the prediction of the retention times and peak profiles (widths and asymmetries) at several mobile phase compositions. For this purpose, peak properties were modelled using data obtained from the injection of aqueous solutions of standards. It should be, however, noted that the purpose of this work was to study the separation of drugs in urine samples, and the chromatograms of urine show besides protein bands, signals corresponding to endogenous compounds, from which one is especially prominent [43]. Therefore, in order to optimise the separation conditions for the βAAs, the retention of this compound was taken into account. Since its identity was unknown, its chromatographic behaviour could be only modelled from injections of blank urine samples in those mobile phases where the direct injection of urine was feasible. Once both the analytes and endogenous compound were modelled, chromatograms at several mobile phase compositions could be predicted, and from these the optimal separation was found. Finally, 1:25 diluted urine samples were fortified up to reach 5 µg/mL of each drug and injected, in order to check the accuracy of the predictions, and the feasibility of the direct injection of urine. It should be indicated that the samples were obtained from healthy human volunteers, who consented to their use in this study.

This work has two main sections dedicated to: (i) the modelling step, with an extensive discussion of the most suitable organic solvent (Section 7.5.1) and the accuracy of several retention models (some of them proposed for this work) (Section 7.5.2), and (ii) the optimisation of the separation of basic drugs in

urine samples, using different gradients, compared to isocratic elution (Section 7.5.3).

7.5.1. Solvent selection

Hybrid eluents used in MLC are usually prepared with surfactant solutions containing a pH buffer and a short-chain alcohol with one to five carbon atoms, selected according to the hydrophobicity of analytes. More recently, acetonitrile has been found to offer excellent performance in MLC, although the elution strength is weaker. For the selection of the best organic solvent to prepare the hybrid micellar mobile phases for the screening of the analytes, information on the chromatographic behaviour of eight βAAs, taken from our laboratory database, was first considered. Four of these compounds agreed with drugs analysed in this work (acebutolol, atenolol, metoprolol and oxprenolol), to which celiprolol, esmolol, pindolol and timolol were added. The drugs were analysed with a conventional C18 column and hybrid mobile phases prepared with SDS and either acetonitrile, ethanol or 1-propanol. The experimental conditions used to obtain the retention data can be consulted in Ref. [44].

Figure 7.1 shows the results predicted for the isocratic elution of the set of basic drugs using hybrid micellar mobile phases at several concentrations of SDS in the 0.04–0.16 M range, where each SDS level in the plots considers variable organic solvent contents as follows: 5–50% for acetonitrile, 5–40% for ethanol, and 5–35% for 1-propanol. In this study, it should be considered that micelle disruption symptoms are observed above 30–35% acetonitrile, 30% ethanol, and 22% 1-propanol [45,46].
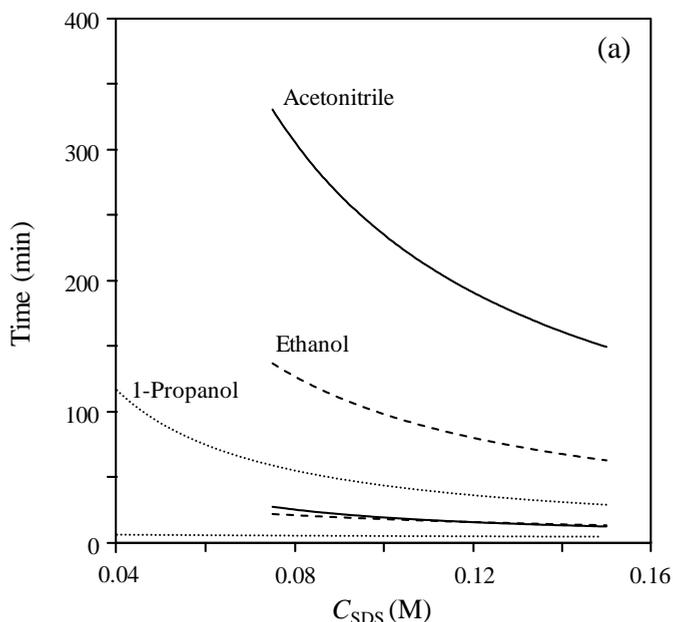
Figure 7.1. Effects of the nature of the organic solvent on the retention times (a), and resolution measured as peak purity (b) for a mixture of βAAs: acetonitrile (solid line), ethanol (dashed line), and 1-propanol (dotted line). The lower curves in (a) correspond to the retention time of the first peak using the mobile phase with the highest organic solvent content for each SDS level. The upper lines provide the analysis times (retention time of the most retained compound), using the mobile phase with the smallest organic solvent content for each SDS level. The study explores the results of the optimisation of the concentration for each organic solvent sorted by SDS levels, based on a grid search (organic solvent × surfactant contents, see text for details). The inspected ranges of organic solvent were: 5–50% for acetonitrile, 5–40% for ethanol, and 5–35% for 1-propanol.
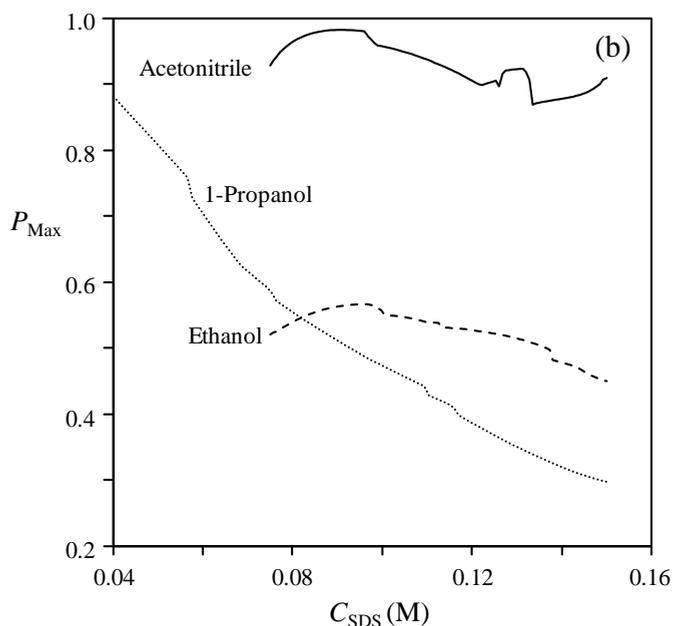
Figure 7.1 (continued).

The plots in Figure 7.1 summarise the results of grid searches (organic solvent × surfactant contents) for the three organic solvents. Each plotted point gives the requested property (maximal and minimal analysis times, and maximal resolution), considering for each solvent all experiments carried out with the same amount of surfactant and variable organic solvent concentration.

In Figure 7.1a, the curves represent the extreme retention values in the solute set, considering all possible organic solvent concentrations for each surfactant level: the upper curves show the dependence of the analysis time (retention time of the last eluted peak) upon changes in the SDS concentration in the mobile phase, and the lower curves depict the changes in the retention time of the first eluted peak. Figure 7.1b shows the maximal resolution found at each SDS level, for the three organic solvents. The resolution curves are less

regular than those in Figure 7.1a, as a result of peak crossings and change in the components of the critical pairs.

It can be seen that the analysis time is too high for acetonitrile, which may even exceed 300 min, although reaching complete resolution. More acetonitrile should be added to reduce the analysis time to practical values, but this would involve micelle disaggregation and abrupt changes of the organic solvent content during the gradient, resulting in undesirable baselines in the chromatograms. This behaviour will be commented in Sections 7.5.3.2 and 7.5.3.3 for 1-propanol.

Ethanol and 1-propanol provided more reasonable analysis times (Figure 7.1a), smaller for 1-propanol. With regard to ethanol, 1-propanol reached higher resolution at lower SDS concentration (Figure 7.1b), without implying excessively long analysis times. It should be also noted that the reduction in analysis time with gradients of organic solvent in the presence of surfactant is less important than for gradients in conventional RPLC. Although the eluted compounds correspond only partially with the set of analytes selected for this work, the general trend of progressive deterioration of the resolution when SDS is increased is transferable. Therefore, 1-propanol was selected as the most appropriate organic solvent for the following studies.

7.5.2. Accuracy of the retention models

This work inspects the separation performance with eluents containing surfactant above the CMC using isocratic elution and gradients of organic solvent covering wide ranges. As will be seen, the predictions in the isocratic mode can be expected to be accurate and reliable. However, in gradient elution, collateral phenomena associated to large solvent changes can ruin the separation expectancies. Gradient predictions should be also less favourable,

since the system will respond with some delay to the changes in organic solvent, and the transition from MLC to HSLC can give rise to deviations between predicted and experimental chromatograms.

Next, the methodology followed for the prediction of retention and the obtained accuracy is discussed, for both the analytes and main endogenous compound. For the fitting of retention models, and the calculation of confidence intervals and other statistics, a software application developed in Visual Basic, written in our laboratory, was used. This application implements the Powell's method with a graphical interface that allows a friendly operation. The main features of the Powell's method are described in Chapter 2.

7.5.2.1. β-Adrenoceptor antagonists

The models used for the prediction of chromatographic peaks were obtained from isocratic experiments. The training sets for the eight βAAs analysed in this work, used to model the chromatographic behaviour, consisted of the retention times and half-widths for the peaks in the chromatograms obtained with 12 and 11 mobile phases containing SDS and 1-propanol, following (3×4) and (3×4 − 1) experimental designs, for the C18 and C8 columns, respectively. In these designs, all organic solvent concentrations at the central SDS level were shifted downwards. This shift was deliberate in order to break the symmetry of the experimental design, and provided more freedom for assaying the models. The specific compositions of the mobile phases can be consulted in Section 7.4.2, where the 0% 1-propanol mobile phases were excluded, due to excessive retention for most compounds. The assayed ranges were 0.05–0.15 M for SDS, and 5–35% for 1-propanol.

To evaluate the modelling quality, the following statistics were calculated [47,48]:

(i) Adjusted correlation coefficient

$$R_{adj} = \sqrt{1 - (1 - R^2) \times \frac{ne - 1}{ne - np - 1}} \qquad (7.18)$$

where R is the Pearson's correlation coefficient:

$$R = \sqrt{1 - \frac{\sum\limits_{i=1}^{ne} (\hat{k}_i - k_{exp,i})^2}{\sum\limits_{i=1}^{ne} (k_{exp,i} - \bar{k}_{exp})^2}} \qquad (7.19)$$

In the above expressions, *ne* corresponds to the number of mobile phases in the experimental design, *np* is the number of model parameters, $\hat{k}_i$ and $k_{exp,i}$ are the predicted and experimental retention factors, respectively, for each mobile phase i in the design, and $\bar{k}_{exp}$ is the mean experimental retention factor.

(ii) Mean relative error

$$RE = \frac{\sum\limits_{i=1}^{ne} \left| \hat{k}_i - k_{exp,i} \right|}{\sum\limits_{i=1}^{ne} k_{exp,i}} \times 100 \qquad (7.20)$$

(iii) Snedecor's F

$$F = \frac{\sum\limits_{i=1}^{ne} (\hat{k}_i - \hat{k}_{mean})^2}{\dfrac{\sum\limits_{i=1}^{ne} (k_{exp,i} - \hat{k}_i)^2}{ne - np}} \qquad (7.21)$$

$\hat{k}_{\text{mean}}$ being the mean predicted retention factor.

(iv) Standard error in prediction

$$s_{\text{pred}} = \sqrt{\dfrac{\sum_{i=1}^{ne} (\hat{k}_i - k_{\exp,i})^2}{ne - np}} \tag{7.22}$$

The values of the determination coefficients measuring the performance of the data fitting, for some selected retention models from those in Section 7.3, are indicated in Tables 7.1 and 7.2. These coefficients were calculated according to the conventional definition (R^2), and considering the differences in the degrees of freedom (R^2_{adj}). Several retention models derived from Equation (7.8) have been considered. Those models containing the same number of parameters (e.g., Equations (7.12)–(7.15)) can be analysed considering both determination coefficients, but when the models to be compared include a different number of parameters, only R^2_{adj} (or any other statistic insensitive to this difference, as the Snedecor's F), should be used. In the tables, the results for the eight βAAs studied in this work are ordered by increasing solute hydrophobicity. The endogenous compound was not included due to the insufficient number of data (modelling of this compound will be commented in Section 7.5.2.2).

Table 7.1. Evaluation of the modelling performance of different retention models for each individual solute eluted with the C18 column.

np^a	Retention models[b]	Statistics[c]	Atenolol	Carteolol	Nadolol	Acebutolol	Metoprolol	Oxprenolol	Propranolol	Alprenolol
4	Eq. (7.5) (MLC)	R^2	0.99997	0.99998	0.99995	0.99996	0.99991	0.99989	0.99992	0.99989
		R^2_{adj}	0.99996	0.99997	0.99993	0.99994	0.99988	0.99984	0.99989	0.99985
	Eq. (7.5) (MLC + HSLC)	R^2	0.98805	0.97970	0.98133	0.96065	0.97008	0.97198	0.97752	0.97765
		R^2_{adj}	0.98357	0.97209	0.97432	0.94590	0.95885	0.96147	0.96909	0.96926
5	Eq. (7.13)	R^2	0.99993	0.99975	0.99955	0.99933	0.99916	0.99874	0.99847	0.99858
		R^2_{adj}	0.99990	0.99966	0.99939	0.99908	0.99885	0.99827	0.99790	0.99805
	Eq. (7.14)	R^2	0.99997	0.99994	0.99979	0.99970	0.99957	0.99925	0.99905	0.99907
		R^2_{adj}	0.99995	0.99991	0.99971	0.99959	0.99940	0.99896	0.99869	0.99872
	Eq. (7.15)	R^2	0.99791	0.99660	0.99627	0.99489	0.99485	0.99472	0.99544	0.99568
		R^2_{adj}	0.99712	0.99533	0.99487	0.99298	0.99292	0.99274	0.99373	0.99406
6	Eq. (7.8)	R^2	0.99998	0.99996	0.99986	0.99981	0.99973	0.99960	0.99967	0.99954
		R^2_{adj}	0.99998	0.99994	0.99981	0.99974	0.99963	0.99944	0.99954	0.99937
	Eq. (7.16)	R^2	0.99996	0.99985	0.99977	0.99952	0.99946	0.99918	0.99896	0.99899
		R^2_{adj}	0.99994	0.99980	0.99968	0.99934	0.99926	0.99986	0.99888	0.99861
	Eq. (7.17)	R^2	0.99998	0.99995	0.99986	0.99980	0.99968	0.82128	0.99929	0.97741
		R^2_{adj}	0.99998	0.99994	0.99980	0.99973	0.99956	0.75426	0.99902	0.96894
7	Eq. (7.11)	R^2	0.99999	0.99998	0.99991	0.99997	0.99986	0.99978	0.99984	0.99978
		R^2_{adj}	0.99998	0.99997	0.99987	0.99995	0.99980	0.99969	0.99978	0.99969

[a] np = number of model parameters. [b] The models are described in Section 7.3. [c] R^2 = Pearson's determination coefficient; R^2_{adj} = adjusted determination coefficient.

Table 7.2. Evaluation of the modelling performance of different retention models for each individual solute eluted with the C8 column.

np[a]	Retention models[b]	Statistics[c]	Atenolol	Carteolol	Nadolol	Acebutolol	Metoprolol	Oxprenolol	Propranolol	Alprenolol
4	Eq. (7.5) (MLC)	R^2	0.99994	0.99995	1.00000	0.99991	0.99993	0.99995	0.99997	0.99995
		R^2_{adj}	0.99992	0.99993	1.00000	0.99987	0.99990	0.99993	0.99995	0.99993
	Eq. (7.5) (MLC + HSLC)	R^2	0.98866	0.97747	0.97784	0.95345	0.97414	0.97666	0.98287	0.98365
		R^2_{adj}	0.98380	0.96782	0.96676	0.93350	0.96305	0.96666	0.97553	0.97663
5	Eq. (7.13)	R^2	0.99968	0.99969	0.99974	0.99952	0.99957	0.99946	0.99922	0.99939
		R^2_{adj}	0.99954	0.99956	0.99962	0.99931	0.99939	0.99922	0.99889	0.99912
	Eq. (7.14)	R^2	0.99956	0.99973	0.99975	0.99962	0.99971	0.99969	0.99951	0.99955
		R^2_{adj}	0.99937	0.99962	0.99963	0.99946	0.99959	0.99955	0.99930	0.99936
	Eq. (7.15)	R^2	0.99757	0.99604	0.99921	0.99407	0.99589	0.99625	0.99702	0.99749
		R^2_{adj}	0.99653	0.99434	0.99882	0.99152	0.99413	0.99465	0.99574	0.99642
6	Eq. (7.8)	R^2	0.99982	0.99986	0.99985	0.99985	0.99985	0.99986	0.99985	0.99986
		R^2_{adj}	0.99974	0.99980	0.99977	0.99978	0.99979	0.99980	0.99979	0.99980
	Eq. (7.16)	R^2	0.99991	0.99988	0.99989	0.99976	0.99982	0.99976	0.99972	0.99975
		R^2_{adj}	0.99987	0.99983	0.99984	0.99965	0.99974	0.99966	0.99960	0.99964
	Eq. (7.17)	R^2	0.99971	0.99987	0.99978	0.99991	0.99980	0.99977	0.99964	0.99968
		R^2_{adj}	0.99927	0.99954	0.99962	0.99945	0.99949	0.99942	0.99909	0.99920
7	Eq. (7.11)	R^2	0.99988	0.99990	0.99996	0.99987	0.99987	0,99987	0.99986	0.99986
		R^2_{adj}	0.99970	0.99975	0.99988	0.99968	0.99969	0.99968	0.99965	0.99965

[a] np = number of model parameters. [b] The models are described in Section 7.3. [c] R^2 = Pearson's determination coefficient; R^2_{adj} = adjusted determination coefficient.

The consequences of applying the micellar equation (Equation (7.4) with $K_{SD} = 0$, i.e., Equation (7.5)), out of the micellar domain, can be observed in the table. As can be seen, Equation (7.5) offers excellent performance in micellar conditions, with R^2 values usually above 0.9999, although the fitting quality tends to deteriorate as the hydrophobicity of solutes increases, due to the elimination of the K_{SD} term. However, when the data from hybrid micellar and high submicellar conditions (MLC + HSLC) were modelled altogether, the decrease in the fitting performance was significant. This highlights the changes that are taking place in both mobile and stationary phases. The deterioration of the fittings, owing to the extension of the organic solvent domain, was usually larger for the C18 column, as a consequence of the higher amount of adsorbed surfactant.

Other alternative models were checked, trying to improve the prediction capability when the transition region (between the micellar and high submicellar conditions) is included. Such models (Equations (7.13)–(7.15), which contained five parameters) were obtained by adding one more parameter to Equation (7.5), which yielded a significant improvement in the prediction performance (see R^2_{adj} values). The results obtained for the C18 column were less accurate, especially for the most hydrophobic solutes. Note that the experimental domain was extended up to 35% 1-propanol. For the C8 column, R^2 was above 0.999 for all solutes.

Amongst the proposed models containing six parameters (Equations (7.8), (7.16) and (7.17)), Equation (7.8) offered the best predictions for all solutes with the C18 column, and for the most retained solutes for the C8 column. Meanwhile, Equation (7.16) yielded better performance for the less retained solutes with the C8 column. In practice, Equation (7.8) was usually satisfactory for the two assayed columns. Finally, Equation (7.11) (with seven parameters)

offered the highest R^2_{adj} with the C18 column for all solutes, with a prediction quality comparable with that offered by Equation (7.5) in the micellar domain. For the C8 column, Equation (7.8) offers better prediction performance, but the difference is less important for faster solutes.

We have checked with the eight βAAs included in our database that Equation (7.11) offered neatly better results with regard to Equation (7.8), when the organic solvent domain was extended up to 50% 1-propanol. Therefore, Equation (7.11) was selected to optimise the separation of the βAAs. However, in spite that Equation (7.11) is statistically the most appropriate model for the data set, in practice, the optimisation of chromatographic resolution could be done with Equation (7.8), since the incidental improvements with Equation (7.11) are of small magnitude.

Table 7.3 shows the regression statistics for the fitting of the retention models, using the two chromatographic columns. Table 7.4 gathers the model parameters and the uncertainties associated to the estimation of each parameter for both columns. Typical relative errors (*ER* in Table 7.3) are between 0.3 and 1.7%, and often these are close to 1%, except for the most hydrophobic solutes, which suffer in a larger extent the modifications in the chromatographic system (micelle disruption and surfactant desorption from the column). In absolute terms, a prediction error of 0.02–1.01 and 0.07–0.99 retention factor units can be expected for the C18 and C8 columns, respectively. These uncertainties can be considered excellent in the field of isocratic MLC. In order to appraise properly the quality of these uncertainties, the reader should note that the domain of organic solvent is, in the case of study, more than two-fold the maximal domain of hybrid MLC, but keeps the same accuracy level, in spite of the drastic changes in the separation system.

MLC optimisation in extended solvent domains

Table 7.3. Statistics accounting the fitting quality of the individual retention data obtained with Equation (7.11) for the two assayed columns.

Solute	Statistics[a]	C18 column	C8 column
Atenolol	ne	12	11
	R^2_{adj}	0.99998	0.99970
	RE	0.33	1.44
	F	135878	8386
	s_{pred}	0.02	0.07
Carteolol	ne	12	11
	R^2_{adj}	0.99997	0.99975
	RE	0.60	1.18
	F	61335	10300
	s_{pred}	0.04	0.10
Nadolol	ne	12	11
	R^2_{adj}	0.99987	0.99988
	RE	0.89	1.15
	F	17620	10276
	s_{pred}	0.11	0.10
Acebutolol	ne	12	11
	R^2_{adj}	0.99995	0.99968
	RE	0.76	1.32
	F	41422	7458
	s_{pred}	0.07	0.19

Solute	Statistics[a]	C18 column	C8 column
Metoprolol	ne	12	11
	R^2_{adj}	0.99980	0.99969
	RE	1.07	1.39
	F	11657	9458
	s_{pred}	0.88	0.32
Oxprenolol	ne	12	11
	R^2_{adj}	0.99969	0.99968
	RE	1.16	1.38
	F	7777	9940
	s_{pred}	0.95	0.53
Propranolol	ne	12	11
	R^2_{adj}	0.99978	0.99965
	RE	1.19	1.59
	F	10954	8891
	s_{pred}	0.97	0.76
Alprenolol	ne	12	11
	R^2_{adj}	0.99969	0.99965
	RE	1.33	1.67
	F	8085	9154
	s_{pred}	1.01	0.99

[a] R_{adj} = adjusted correlation coefficient (Equation (7.18)); RE = mean relative error (Equation (7.20)); F = Snedecor's F (Equation (7.21)); s_{pred} = standard error in prediction (Equation (7.22)).

Table 7.4. Fitted model parameters and their corresponding uncertainties for each individual solute, using Equation (7.11). The urine endogenous compound was fitted using Equation (7.5), since only measurements in the micellar domain were possible.

Solute	Coefficients	C18 column	C8 column	Solute	Coefficients	C18 column	C8 column
Atenolol	c_0	0.000 ± 0.006	0.11 ± 0.02	Oxprenolol	c_0	0.03 ± 0.01	0.01 ± 0.01
	c_1	2.00 ± 0.03	-0.64 ± 0.07		c_1	0.80 ± 0.01	0.54 ± 0.04
	c_2	-0.57 ± 0.14	-3.20 ± 0.36		c_2	0.75 ± 0.10	0.25 ± 0.15
	c_3	28.67 ± 1.67	6.90 ± 4.22		c_3	11.89 ± 1.31	11.53 ± 1.01
	c_4	2.15 ± 0.88	18.84 ± 1.49		c_4	-7.23 ± 0.75	-3.92 ± 1.33
	c_5	19.08 ± 1.94	-12.32 ± 0.98		c_5	20.11 ± 1.55	13.66 ± 2.8
	c_6	-2.20 ± 0.51	11.36 ± 0.99		c_6	-4.31 ± 0.78	-3.58 ± 0.24
Carteolol	c_0	-0.01 ± 0.01	6.69 ± 0.07	Propranolol	c_0	-0.02 ± 0.01	-1.15 ± 0.01
	c_1	1.67 ± 0.02	0.10 ± 0.26		c_1	0.59 ± 0.22	0.51 ± 0.13
	c_2	-0.07 ± 0.03	-1.80 ± 1.03		c_2	0.58 ± 0.27	0.36 ± 0.03
	c_3	24.51 ± 0.87	11.06 ± 4.88		c_3	9.11 ± 2.40	10.20 ± 1.70
	c_4	-4.11 ± 0.28	6.86 ± 6.78		c_4	-5.36 ± 2.01	-3.45 ± 0.26
	c_5	26.3 ± 1.1	7.58 ± 14.90		c_5	13.29 ± 4.09	10.61 ± 1.03
	c_6	-3.89 ± 0.16	4.57 ± 3.08		c_6	-3.27 ± 1.49	-3.45 ± 0.83
Nadolol	c_0	0.00 ± 0.01	0.12 ± 0.15	Alprenolol	c_0	0.020 ± 0.001	0.013 ± 0.001
	c_1	0.63 ± 0.18	-3.99 ± 7.77		c_1	0.59 ± 0.05	0.67 ± 0.13
	c_2	-0.11 ± 0.29	-2.02 ± 1.99		c_2	0.56 ± 0.05	0.42 ± 0.03
	c_3	14.60 ± 0.46	-14.22 ± 48.77		c_3	8.98 ± 0.95	11.64 ± 1.96
	c_4	-3.31 ± 3.41	5.33 ± 9.39		c_4	-5.04 ± 0.47	-4.16 ± 0.36
	c_5	24.35 ± 8.91	12.94 ± 16.97		c_5	12.46 ± 1.05	10.56 ± 1.78
	c_6	0.65 ± 0.74	22.90 ± 39.26		c_6	-3.53 ± 0.42	-4.78 ± 1.00

Table 7.4 (continued).

Solute	Coefficients	C18 column	C8 column
Acebutolol	c_0	-3.96 ± 2.06	1.41 ± 6.04
	c_1	2.6 ± 3.8	1.36 ± 0.13
	c_2	0.93 ± 3.15	-0.33 ± 0.11
	c_3	28.05 ± 0.52	18.14 ± 4.58
	c_4	-9.97 ± 0.28	-1.99 ± 0.77
	c_5	31.26 ± 1.03	17.49 ± 2.67
	c_6	-11.32 ± 0.15	-5.55 ± 1.79
Metoprolol	c_0	-0.03 ± 0.02	0.01 ± 0.01
	c_1	1.22 ± 0.05	0.50 ± 0.02
	c_2	0.81 ± 0.43	-0.23 ± 0.02
	c_3	17.2 ± 1.1	12.56 ± 1.78
	c_4	-8.6 ± 3.2	-1.68 ± 1.89
	c_5	27.08 ± 7.20	13.68 ± 6.16
	c_6	-6.17 ± 0.32	-2.70 ± 0.38

Solute	Coefficients	C18 column	C8 column
Urine endogenous	c_0	7.06 ± 3.49	8.55 ± 6.21
	c_1	-18.37 ± 26.40	-23.51 ± 45.70
	c_2	-22.86 ± 25.75	-29.29 ± 43.61
	c_3	27.87 ± 190.78	34.80 ± 317.75

The parallelism between the magnitude of the model coefficients for both columns (Table 7.4) is noteworthy, especially for the most hydrophobic solutes. This indicates that the retention surfaces have comparable features and shape.

7.5.2.2. Endogenous compound

Modelling of the most prominent endogenous compound in urine presents new challenges. For this compound, the retention was modelled from the information obtained by injection of urine blanks, at several mobile phase compositions (SDS/1-propanol). To avoid the precipitation of the protein matrix at high concentrations of 1-propanol, the endogenous compound was only measured in those mobile phases where the organic solvent content was $\leq 15\%$ (i.e., the MLC domain). Therefore, to model the retention, data from only five mobile phases were available, following a $2\times2 + 1$ experimental design (at 0.05 and 0.15 M SDS, each at 5% and 15% 1-propanol, together with 0.10 M SDS and 10% 1-propanol).

Due to the smaller number of available measurements for the endogenous compound, the equation describing the retention in the micellar media (Equation (7.5)) was selected. Observe in the provided experimental chromatograms of this work that the endogenous compound elutes always in micellar conditions (at low solvent content with gradient elution), for which Equation (7.5) is perfectly valid. Even using this equation and data strictly in the micellar domain, the model parameters were still rather uncertain (see Table 7.4). Thus, the deviations in the prediction of retention were expected to be larger for the endogenous compound, compared to the βAAs.

7.5.2.3. Use of models with only organic solvent as a factor

From a practical point of view, it is interesting to model the behaviour using as few experiments as possible. With the selected two-factor retention models and one degree of freedom, a minimum of 7 and 8 experiments are required for Equations (7.8) and (7.11), respectively, and probably, the number should be higher in practice. Since practical situations in gradient elution involve changes in the concentration of organic solvent with constant SDS, simplified equations can be used instead, such as Equation (7.10) (with 4 parameters, derived from Equation (7.8)), and Equation (7.12) (with 5 parameters, derived from Equation (7.11)). The validation of these two models (Equations (7.10) and (7.12)) could not be adequately done with the available data obtained for this work, since there were no enough experiments at constant surfactant concentration. For this reason, the set of experiments from the laboratory database was used again. This contained enough information for 12 βAAs, using 1-propanol as organic solvent (the eight drugs studied in this work, except nadolol, together with celiprolol, labetalol, pindolol, timolol and esmolol).

Figure 7.2 shows box-and-whiskers diagrams drawn for R^2_{adj} using Equations (7.10) and (7.12) to fit the retention data obtained at each of two SDS levels (0.075 M and 0.15 M). It can be seen that the fittings are more accurate at higher SDS, and Equation (7.12) offers higher prediction quality than Equation (7.10). This result supports again the conclusion that Equation (7.11) (with the $\sqrt{\varphi}\,[S]$ term) is the most accurate model to predict the retention in extended organic solvent domains.
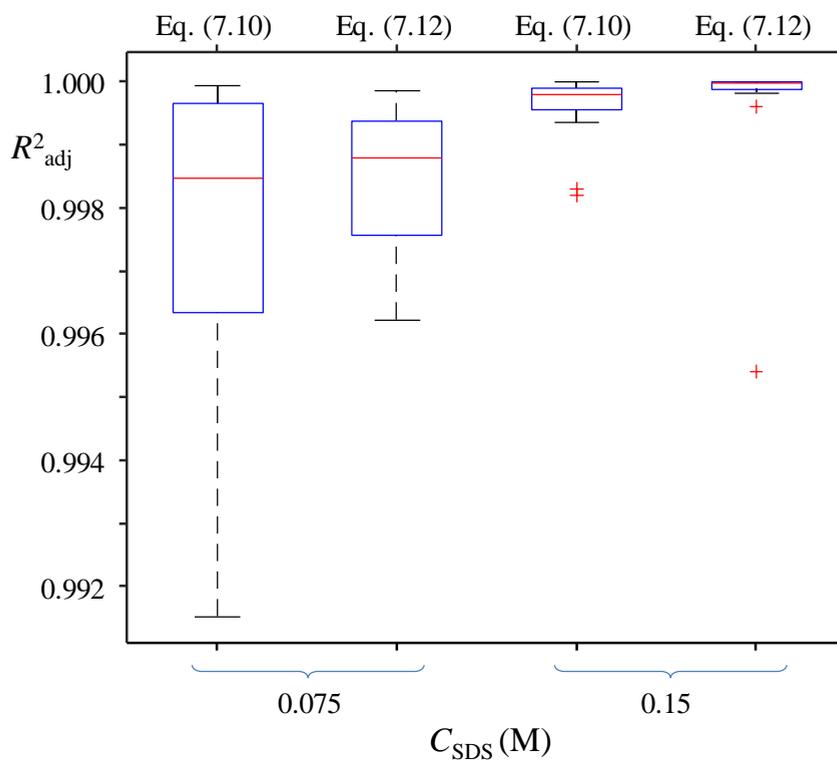
Figure 7.2. Box-and-whiskers diagram showing the range of adjusted determination coefficients for Equations (7.10) and (7.12), at two SDS concentration levels.

7.5.3. Separation performance

This section studies the separation obtained for the C18 and C8 columns under similar conditions, by applying linear and multi-linear gradients comparing the results with isocratic elution. The problems arising from the use of eluents containing surfactant in gradient elution with organic solvent, and the features of the optimal separations, are considered. It should be noted that the main objective of this work is the study of the different elution modes, rather than the analysis of the mixture of βAAs, which should be taken just as an example. The presence of the matrix of a physiological fluid (urine) has also been included in the study, so that the samples contained, besides the eight analytes, the peaks associated to the matrix, corresponding to proteins and several endogenous compounds, one of them giving rise to a very prominent peak.

As commented, in previous work, commercial optimisation software (Drylab) was used for a similar separation. However, it suffered of severe limitations when applied to MLC and HSLC, the most important being the lack of a retention model truly valid. In this work, proper equations are used for predicting the retention. Also, the optimisation of the separation conditions was carried out using a methodology developed for RPLC, based on the modelling of retention and peak shape. The tools applied in this work for the prediction of gradient retention times and peak widths, simulation of chromatograms, measurement of resolution and search of the optimal separation conditions, are described in Chapters 3 and 4. More details are given elsewhere [29,49,50].

Before starting the optimisation study, the operation limits that allowed the direct injection of urine were fixed by examining the effects of increasingly disturbing elution conditions (use of higher organic solvent content, smaller amount of micelles, and steeper gradients).

370

Figure 7.3 shows chromatograms of a filtered urine sample, obtained under different elution conditions (isocratic or gradient). Figures 7.3a to c show the results of isocratic runs, which lead to chromatograms with clean baselines (once the proteins have been eluted). Concentrations of at least 15% 1-propanol can be used without protein precipitation risk. Figures 7.3d to f show the results of the use of gradients, which is potentially more troublesome, owing to the strong baseline fluctuations: the steeper the gradient, the more severe the fluctuations. The protein band elutes at the beginning of the chromatogram. Following the proteins, some peaks corresponding to minor endogenous compounds can be observed, and at intermediate times, the prominent peak of the main endogenous compound, whose behaviour was modelled in Section 7.5.2.2. Provided the concentration of organic solvent is kept below 15% up to the elution of this endogenous compound, gradient elution will be possible. It should be, however, noted that in the conditions of Figure 7.3, the elution of the endogenous compound was detrimental for the detection of atenolol, whose peak co-elutes in a wide range of conditions.

7.5.3.1. Isocratic separation

For comparison purposes, the optimisation of isocratic elution was first carried out, considering the concentrations of SDS and organic solvent as experimental factors. In Figure 7.4, resolution contour maps are depicted for the two columns under study (C18 and C8). Each point in the plots indicates the expected resolution, measured as global peak purity (P) (see Chapter 5 for its definition), for the studied experimental ranges of surfactant and organic solvent. Only the regions of high resolution ($P > 0.90$) have been plotted.
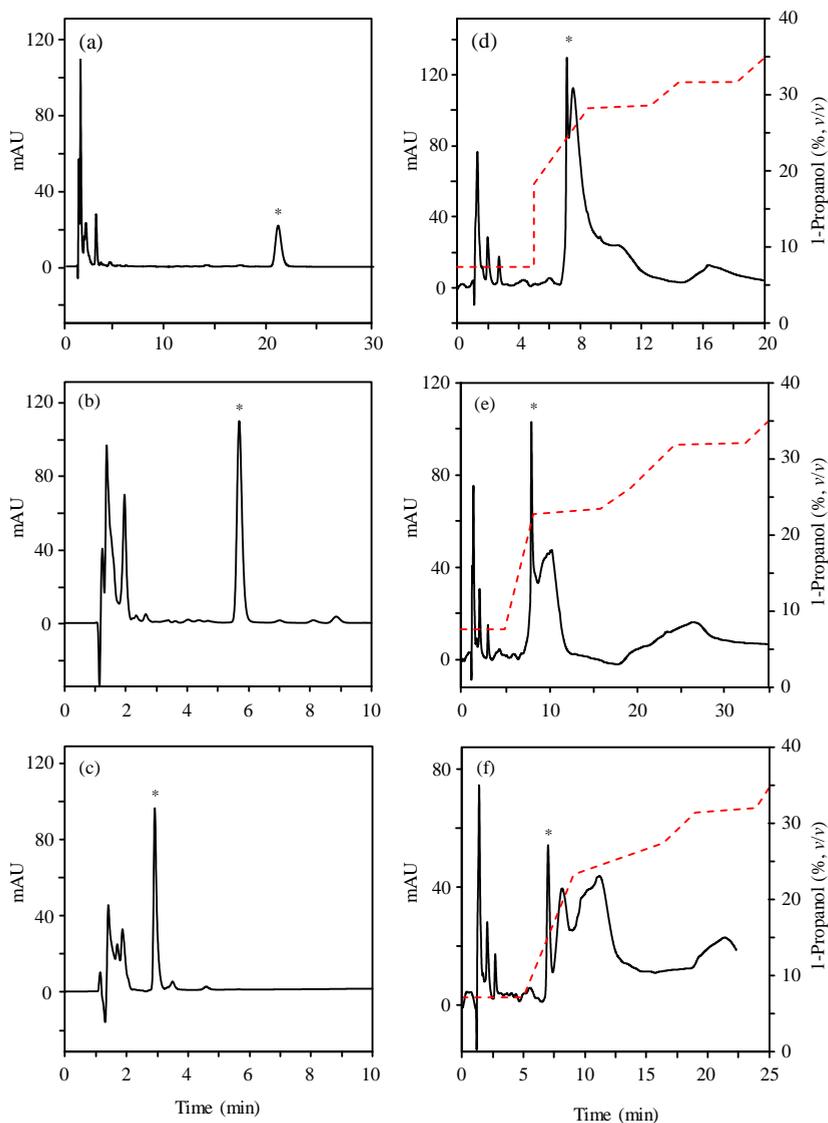
Figure 7.3. Chromatograms of filtered urine samples using the C18 column under different conditions: (a) isocratic elution with 0.05 M SDS (pure micellar medium), (b) 0.10 M SDS/5% 1-propanol, (c) 0.15 M SDS/15% 1-propanol, and (d-f) multi-linear gradients of 1-propanol in the presence of 0.05 M SDS. The endogenous compound is marked with an asterisk. Experimental signals are given in milli absorbance units.
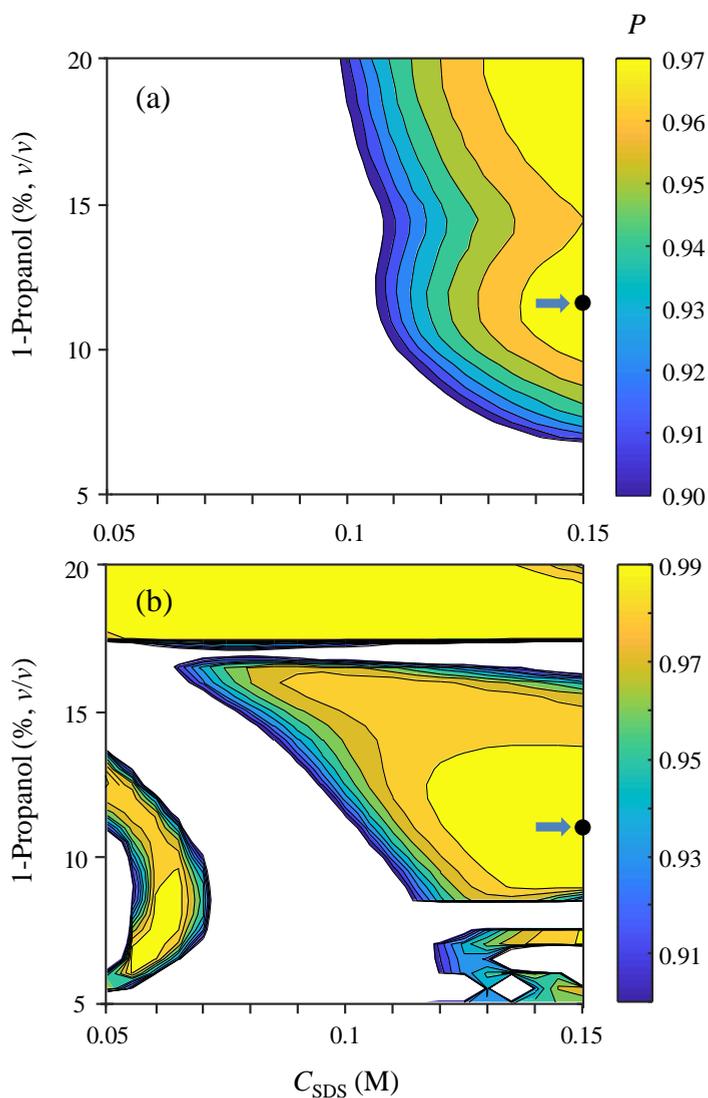
Figure 7.4. Contour map of global resolution expressed as peak purity (P) under isocratic conditions, using the C18 (a), and C8 (b) columns, for the separation of the eight βAAs and the main endogenous compound. The conditions of maximal resolution that were selected for experimental validation (see Figure 7.5) are marked with an arrow: (a) 11.71% and (b) 11.14% 1-propanol, both containing 0.15 M SDS.
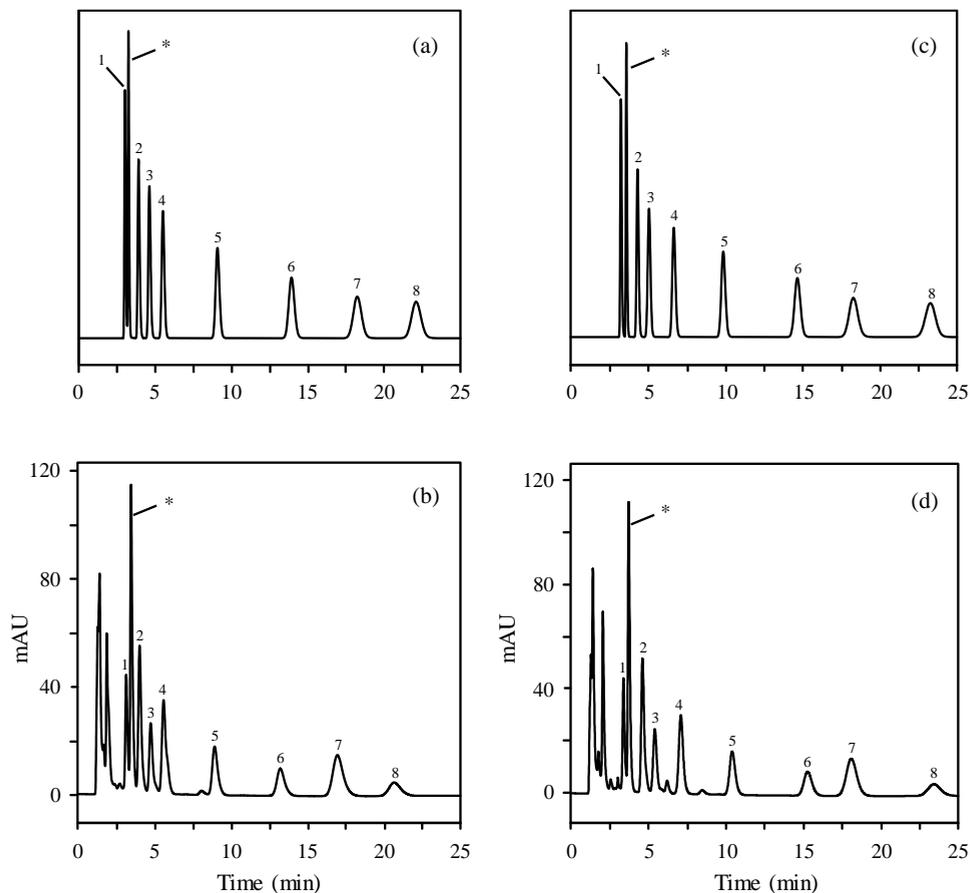
Figure 7.5. Predicted (a,c) and experimental (b,d) optimal chromatograms for the isocratic conditions indicated in Figure 7.4, using the C18 (a,b) and C8 (c,d) columns. The experimental chromatograms correspond to the direct injection of a urine sample fortified with the eight βAAs. Codes for the compound identities are given in Section 7.4.1. Other details are given in Figure 7.3. The simulations were carried out with normalised peak areas.

Figure 7.5 shows predicted and experimental chromatograms for the optimal separation conditions marked with an arrow in Figure 7.4. These conditions were chosen because maximal resolution ($P > 0.95$) is reached, and at the same time the existence of micelles in the mobile phase is guaranteed (note that the concentration of 1-propanol is below 15%). Thus, urine proteins are kept in solution. Despite the differences in the contour maps, it should be noted that both C18 and C8 columns show favourable resolution in similar regions, but this similarity should be considered circumstantial.

Experimental conditions able to reduce the analysis time even more would imply too high concentration of organic solvent, which would not allow the formation of micelles. Micelles are needed to avoid the precipitation of proteins in urine, which would damage the column. With the selected mobile phase, the analysis time would be around 20–22 min. As observed, the experimental chromatograms (Figures 7.5b and d) agree with those predicted (Figures 7.5a and c).

7.5.3.2. Linear gradients of organic solvent

In order to reduce the analysis time, keeping good separation after sweeping the proteins in urine off the column, different types of gradient were investigated. The elution of the proteins in isocratic conditions of short duration and minimal organic solvent content was first checked. After this step, the concentration of organic solvent was increased by applying a linear gradient to accelerate the elution of the analytes, whose slope was optimised at constant SDS concentration. This study was carried out with the C18 column.

The optimisation was initially performed at only three SDS levels (0.05, 0.10 and 0.15 M), forcing an initial 5 min long isocratic step. The concentration of organic solvent in this isocratic step was set to 5%, from which the linear

gradient started. In all cases, solute retention was modelled with Equation (7.12), at each of the three SDS levels. It was found that the separation was not feasible, since the prominent endogenous compound and atenolol partially co-eluted, even in the best conditions, and the peaks of acebutolol and nadolol were almost totally overlapped. As an example, Figure 7.6 shows the best expected chromatogram using 0.05 M SDS. As can be seen, the separation is deficient.
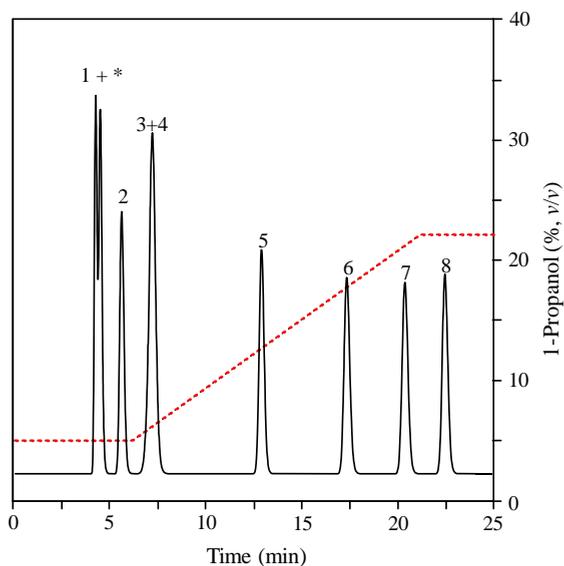
Figure 7.6. Best linear gradient and corresponding predicted optimal chromatogram for the separation of the eight βAAs and the endogenous compound (marked with an asterisk), using the C18 column when a 5 min isocratic pre-elution step in micellar conditions was programmed to elute the proteins. The SDS concentration was 0.05 M. Other details are given in Figure 7.5.

These unsatisfactory results suggested the need of developing a more comprehensive search with a wider scope of conditions, for both columns, by including more factors:

(i)   SDS concentration levels in the 0.05–0.15 M range, with 0.01 M steps,

(ii) initial concentration of 1-propanol in the linear gradient in the 5–15% range, with 1% steps,

(iii) final concentration of 1-propanol between each initial value explored in (ii) and 35%, with 1% steps, and

(iv) gradient times between 5 and 30 min, with 5 min steps.

The search was carried out in the presence and absence of an initial 5 min isocratic pre-elution step, using Equation (7.11) for predicting the retention of the βAAs. Since the results indicated that the inclusion of an initial isocratic pre-elution was detrimental to achieve good resolution, only the comprehensive study in the absence of such pre-elution will be shown. The increase of organic solvent along the gradient was allowed to begin just at the start of the injection, but the maximal concentration of organic solvent at the start of the gradient was limited to 15%. Also, those gradients involving 1-propanol above 15% at the time of the elution of the endogenous compound were discarded. In this way, proteins in urine always eluted in the presence of micelles.

Figure 7.7 shows the systematic exploration of linear gradients for the C18 (Figure 7.7a), and C8 (Figure 7.7b) columns, without an isocratic pre-elution step. The figure depicts the maximal resolution (right axis, dashed line), measured as global peak purity reached with the linear gradients run at each SDS level. The maximal analysis time (left axis, solid line), obtained at each SDS level is overlaid.
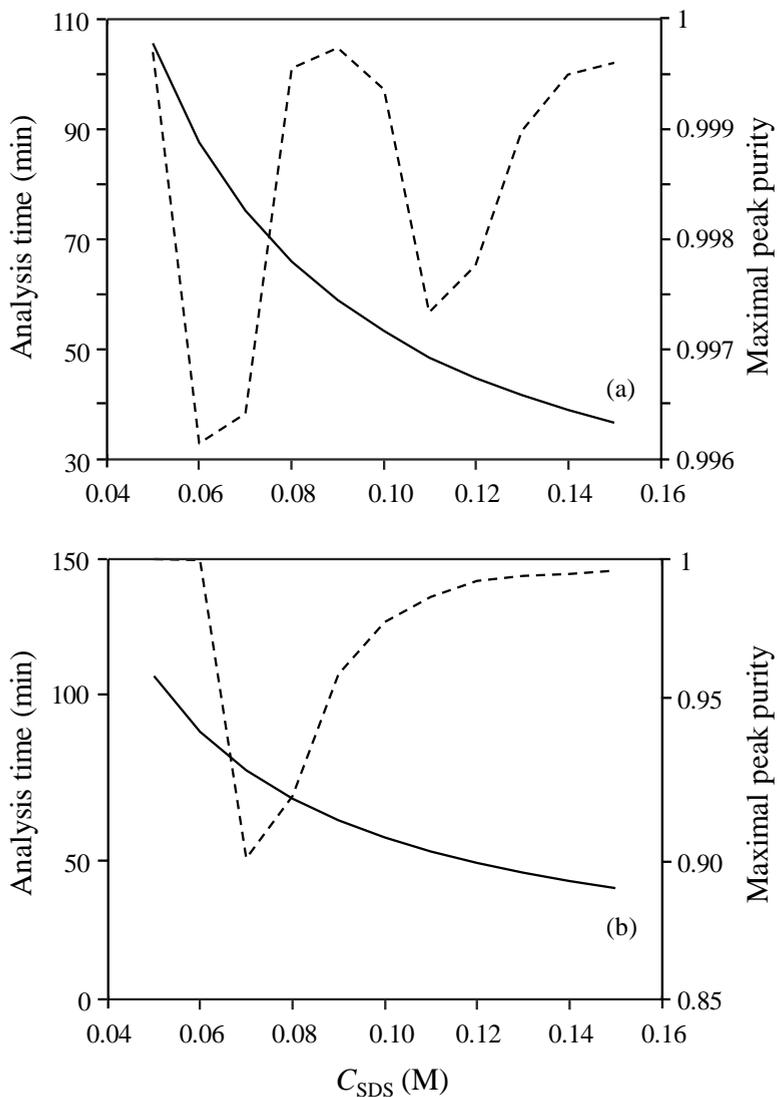
Figure 7.7. Analysis time (solid line) and maximal resolution measured as peak purity (dashed line), for the systematic exploration of linear gradients for the C18 (a), and C8 (b) columns, without isocratic pre-elution step (see text for details).

As observed, the C18 column is able to resolve the sample in the entire domain ($P > 0.995$), at least at one linear gradient for each examined SDS level. The C8 column resolves the sample mainly at the extreme SDS levels. Since at low SDS concentration the analysis time was too high (> 60 min), we concentrated the effort on gradients with an SDS concentration above 0.12 M.

The optimisation of the separation conditions is usually carried out attending only to resolution, without ranking the solutions (i.e., gradients) according to the analysis time. In this work, the results of the optimisation are given as Pareto optimality plots [51] (Figure 7.8), where each point in the plot corresponds to a gradient characterised by its resolution performance and analysis time. In these plots, there is no unique optimal gradient, but a set of optimal gradients, giving rise to the so-called Pareto front, which gathers the gradients where one objective (resolution or analysis time) cannot be improved without worsening the other.

The chromatograms shown in Figure 7.9 belong to the Pareto front (the marked solutions in Figure 7.8), for the C18 and C8 columns, respectively. The separation with the C8 column was very satisfactory, with an excellent agreement between predicted and experimental chromatograms (Figures 7.9c and d). Meanwhile, the C18 column (with a higher amount of SDS adsorbed on the stationary phase at the beginning of the gradient) yielded a problematic baseline and larger deviations in the predictions (Figures 7.9a and b). This is not surprising, considering that the column chemistry is undergoing a larger modification during the elution. These results indicate that the C8 column is more advisable when a gradient separation is aimed.
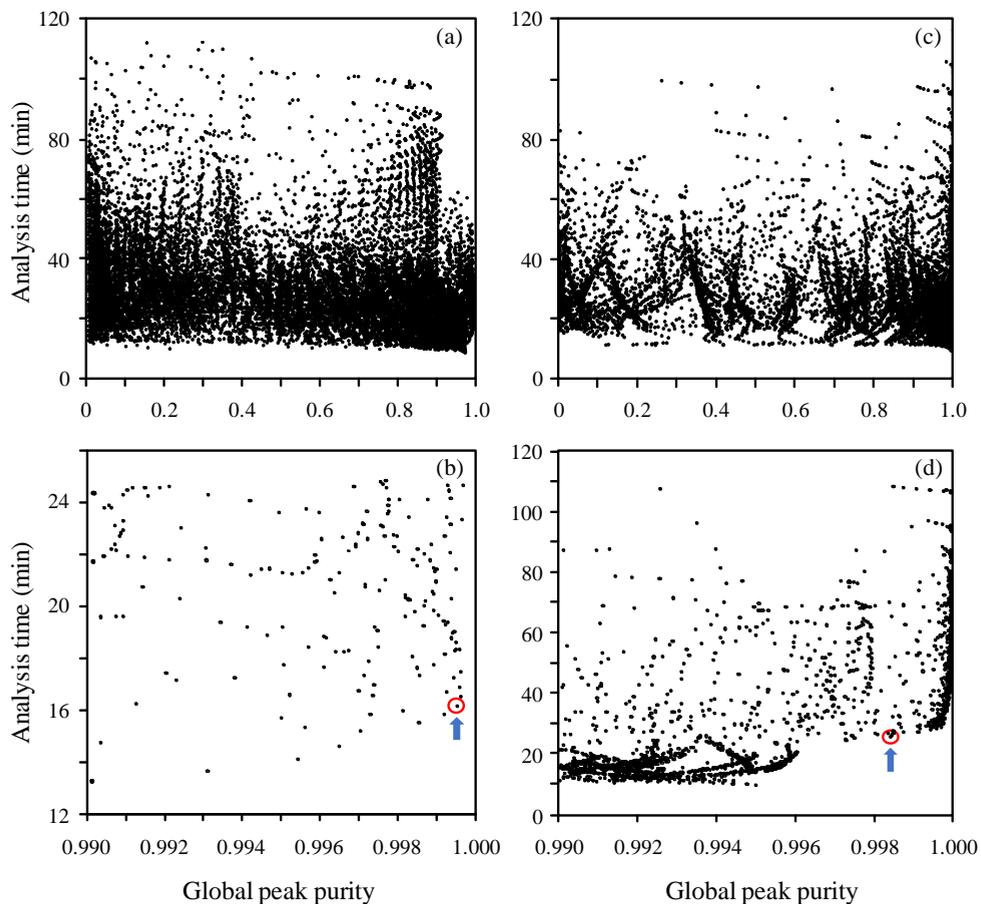
Figure 7.8. Pareto optimality plots corresponding to the optimisation of linear gradients performing a systematic search, for the C18 (a,b), and C8 (c,d) columns. The plots below expand the regions of highest resolution. The two selected gradients for experimental validation (Figure 7.9) are marked with an arrow in (b) and (d).
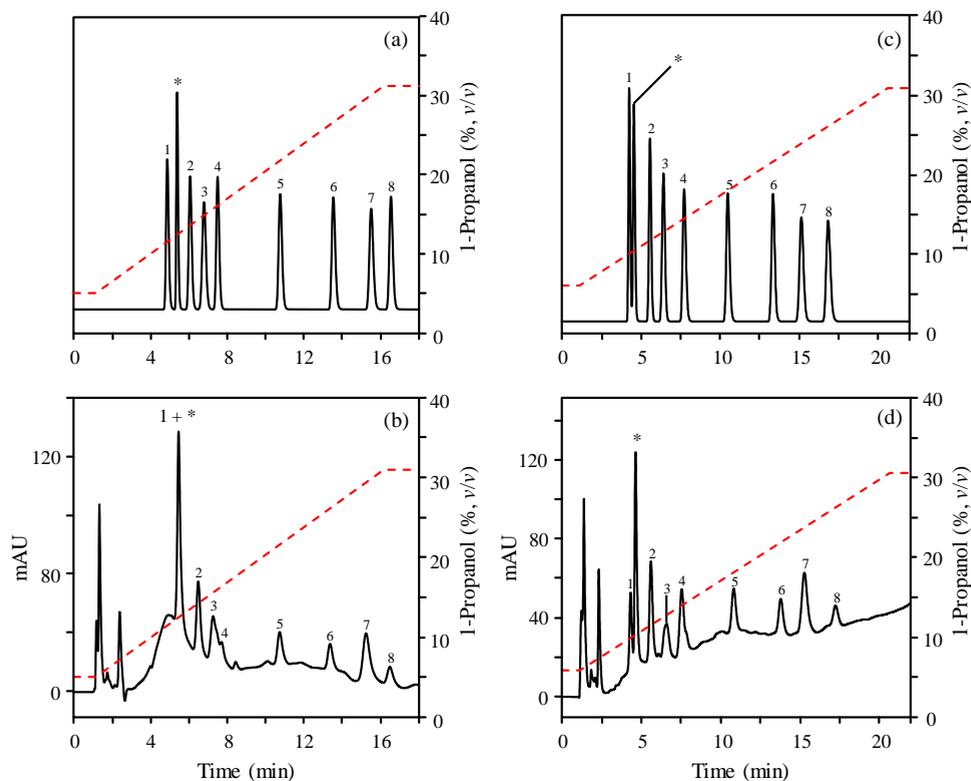
Figure 7.9. Best linear gradients and corresponding optimal chromatograms for the separation of the eight βAAs and the main endogenous compound, using the C18 (a,b) and C8 (c,d) columns: (a,c) predicted and (b,d) experimental chromatograms. Note that all gradient programs include the delay associated to the dwell time. The concentration of SDS was 0.15 M. Other details are given in Figure 7.5.

7.5.3.3. Multi-linear gradients of organic solvent

Finally, the consequences of applying multi-linear gradients were evaluated. These gradients are built by setting a certain number of intermediate nodes inside a linear gradient, whose initial (t_{dwell}, φ_0) and final (t_G, φ_F) nodes delimit the search space. The multi-linear gradients are then built by dividing the preliminary linear gradient into a series of consecutive linear segments at constant or increasing concentration of organic solvent. For the multi-linear search, the 0.15 M SDS level was again selected, since this concentration offered the best performance when the concentration of organic solvent was optimised using linear gradients (see Figure 7.7). The search of the optimal location of each node was operated by Genetic Algorithms (GAs). Excessively sudden transitions between segments were discarded as valid solutions, since the baseline usually presents stability problems in MLC gradient elution (see Section 7.5.3.2).

For both the C18 and C8 columns, the same search settings in the GAs were used to obtain the best multi-linear gradients. The initial population was randomly generated and consisted of 50 gradients. Other parameters in the configuration of the algorithm were the probability of mutation (3%), reintroduction of the best solution (5%), and cross-linking (100%). The successive populations generated during the evolution were stored and represented in Figure 7.10 as Pareto optimality plots. The observation of the Pareto plots indicates that the C8 column should offer better separation performance, giving rise to a larger number of multi-linear gradients with resolution exceeding $P = 0.9$. This column will also allow a larger reduction in the analysis time, up to about 10 min, keeping the resolution at values close to $P = 0.9$.
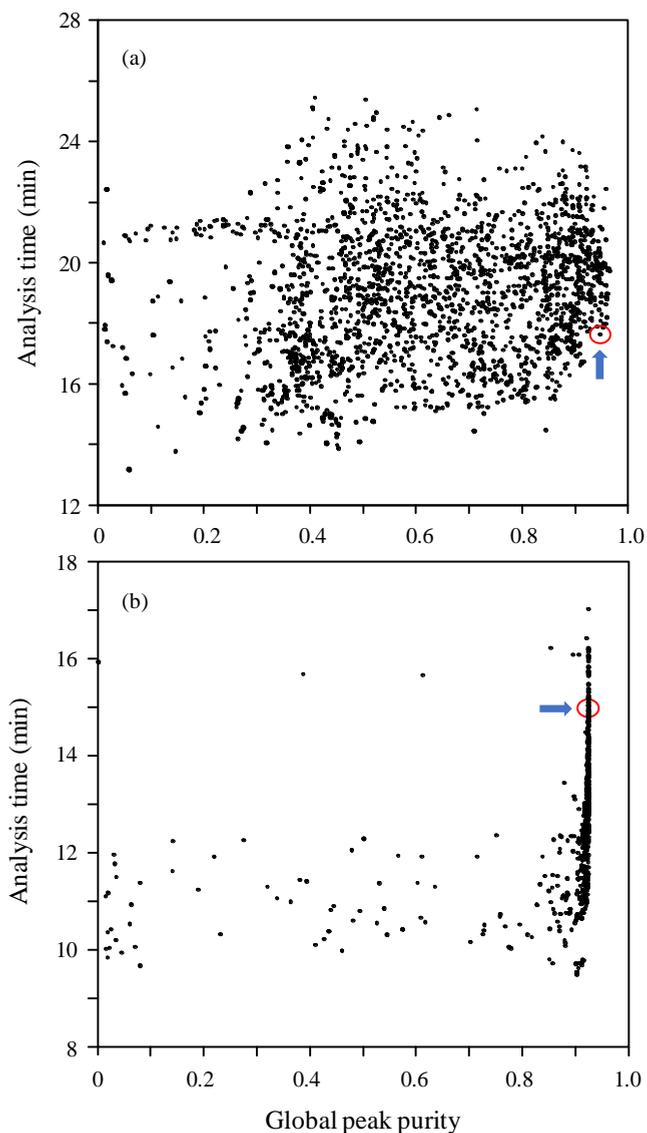
Figure 7.10. Pareto optimality plots corresponding to the optimisation of multilinear gradients carried out by GAs, for the C18 (a), and C8 (b) columns. The selected gradients for experimental validation (Figure 7.11) are marked with an arrow.
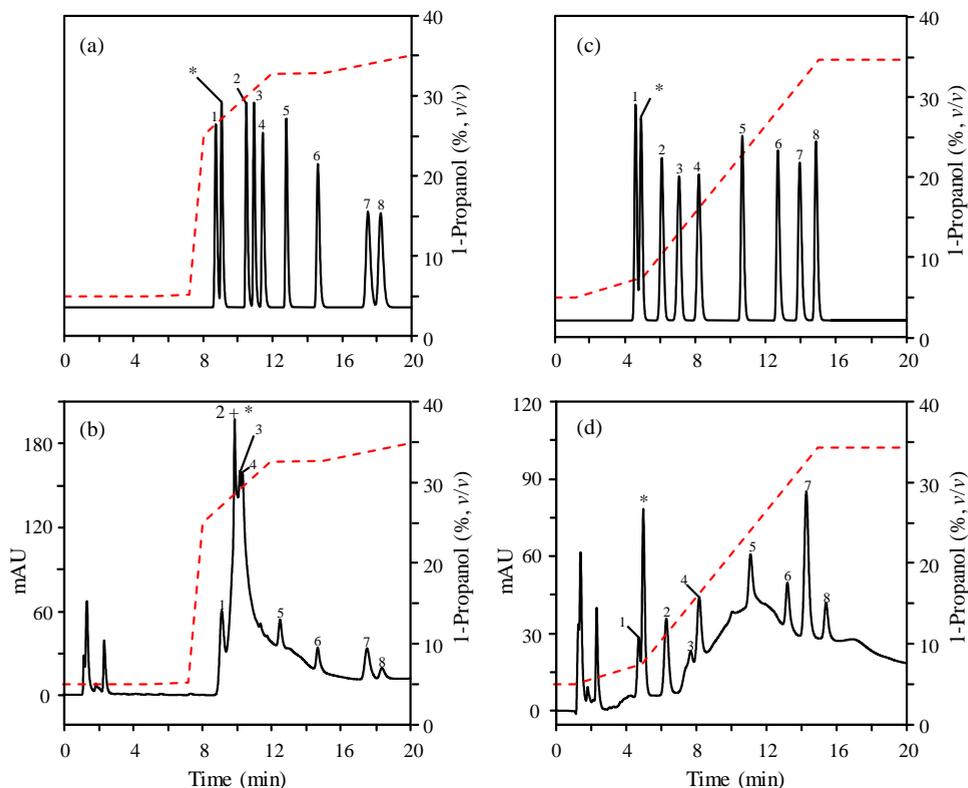
Figure 7.11. Best multi-linear gradients and corresponding optimal chromatograms for the separation of the eight βAAs and the main endogenous compound, using the C18 (a,b) and C8 (c,d) columns: (a,c) predicted and (b,d) experimental chromatograms. Other details are given in Figure 7.9.

An optimal gradient with similar analysis time (around 15–19 min) was selected for each column from those belonging to the Pareto front, for their experimental validation. Figure 7.11 depicts the corresponding predicted and experimental chromatograms for these multi-linear gradients. The experimental chromatograms were obtained with samples of urine fortified with the eight βAAs. The high magnitude of the baseline disturbance for the C18 column should be highlighted, which is induced by the changes in the concentration of organic solvent along the gradient. This disturbance occurs due to surfactant desorption from the column and the strong variation in the concentration of organic solvent during the first steps of the gradient (of almost 20% in about 1 min). The C8 column adsorbs a smaller amount of surfactant, which results in disturbances of smaller magnitude (note also that the gradient slope is less steep for the C8 column).

Attending to these results, it can be concluded that multi-linear gradients in MLC/HSLC are only acceptable when the changes in organic solvent are mild. These gradients, which are excellent in conventional RPLC, are strongly limited by detection problems when mobile phases with surfactant above the CMC are used as eluents. Therefore, simple linear gradients or multi-linear gradients with smooth transitions are the best option for the analysis of samples using eluents containing surfactant in extended organic solvent domains. The high flexibility of multi-linear gradients is unfavourable, not only because of detection problems, but also because surfactant desorption shifts the solutes towards smaller retention. In spite of this, the good agreement between predicted and experimental chromatograms (see peaks 2 to 4 in Figure 7.11), even in these conditions, is remarkable.

7.6. Conclusions

The interpretive study carried out in this work has allowed a comprehensive evaluation of the performance of MLC in extended organic solvent domains, using isocratic and gradient elution to analyse drugs in physiological fluids. Despite the good expectations of isocratic elution in MLC, the main objective is to investigate the separations in eluents at fixed surfactant concentration by applying organic solvent gradients in very wide intervals, which would allow reducing the analysis time. The accuracy of the retention model developed for extended organic solvent domains, fitted using isocratic experiments, indicates that the predictions in the isocratic mode can be expected to be correct and reliable. The situation is less favourable for gradient predictions, since the system will respond with some delay to changes in eluent composition, and these transient states towards stabilisation will result in deviations between predicted and experimental chromatograms, together with severe baseline problems.

For the sample under study, 1-propanol offered better performance than ethanol and acetonitrile, since high resolution was reached within reasonable analysis times. In these conditions, several equations were studied to check the description capability of retention for the eight βAAs. Equation (7.11) allowed accurate predictions for both hybrid micellar and high submicellar regions. This equation improves the prediction performance of Equation (7.5), which is widely used in hybrid MLC. Also, Equation (7.11) results in a simplified equation (Equation (7.12)), valid for organic solvent gradients at fixed surfactant concentration, which provides highly accurate results. In practice, when optimising a new sample, the optimisation can be first attempted with this equation, developing five or six experiments at the highest level of SDS (faster analysis times). If gradient optimisation is successful in these conditions,

considerable time will be saved. Otherwise, the design should be expanded to include the influence of other SDS concentrations.

For the endogenous compound showing the prominent peak in urine, which elutes at rather short retention times, insufficient experiments were available for fitting Equation (7.11), since the measurements above 15% 1-propanol were not feasible in the presence of proteins. Since the elution of the endogenous compound in urine should occur in the presence of micelles, a model appropriate for hybrid micellar media was used instead (Equation (7.5)). However, occasional deviations were observed for the peak of the endogenous compound in the experimental chromatograms, with regard to the predictions. Acquiring more experimental data in the micellar region would improve the prediction accuracy for this compound.

The optimisation in the isocratic mode for the eight βAAs resulted in good resolution in reasonable analysis time (ca. 25 min). Although this time is somewhat longer than that required with a linear gradient, the lack of re-equilibration makes the isocratic elution an excellent alternative to analyse these samples using surfactant-mediated eluents. In principle, gradient separations always offer the advantage of a smaller analysis time, but when a column is coated with surfactant, collateral problems appear, associated with the large solvent alteration, which can ruin the separation expectations. Thus, when a gradient implies a rapid change in the concentration of organic solvent, it is no longer feasible, because baseline disturbances deform the signals and prevent obtaining proper quantifications. Linear gradients result in a less problematic baseline than more complex gradients, while allow shorter analysis times compared to isocratic separations. Therefore, linear gradients can be an acceptable alternative in MLC, even when extended ranges of organic solvent are used. The only disadvantage is that if a second injection is desired, a

re-equilibration step is needed. With the sample analysed in this work, the protein matrix could be swept off in the first minutes, before the elution of the analytes, which reached good resolution.

This work shows that very complex multi-linear gradients are non-advisable in MLC/HSLC. Only in case the different consecutive segments in the gradient have moderate and similar slopes, acceptable baselines could be obtained. Also, the C8 column has been shown as an appropriate option (compared to a C18 column) if the separation is carried out using a linear gradient. In the isocratic mode, both columns are equally acceptable. In any case, this work repeatedly confirmed the low magnitude of the deviations between predicted and experimental peaks for the studied analytes in all examined situations. In future work, the applicability of the proposed methodology that uses linear gradients of organic solvent in extended organic solvent domains should be further examined to check the performance of the separation of drugs in different physiological fluids. The research may involve other organic solvents as well.

7.7. References

[1] A. Berthod, M.C. García Álvarez-Coque, *Micellar Liquid Chromatography* (edited by J. Cazes), Marcel Dekker, New York, 2000.

[2] M.J. Ruiz Ángel, M.C. García Álvarez-Coque, A. Berthod, New insights and recent developments in micellar liquid chromatography, *Sep. Purif. Rev.* 38 (2009) 45–96.

[3] M.C. García Álvarez-Coque, M.J. Ruiz Ángel, S. Carda Broch, Micellar Liquid Chromatography: Fundamentals, in *Analytical Separation Science Series*, Vol. 2 (edited by J.L. Anderson, A. Stalcup, A. Berthod, V. Pino), Wiley-VCH, New York, 2015, pp. 371–406.

[4] A. Berthod, I. Girard, C. Gonnet, Micellar liquid chromatography, adsorption isotherms of two ionic surfactants on five stationary phases, *Anal. Chem.* 58 (1986) 1356–1358.

[5] B.K. Lavine, S. Hendayana, Band broadening in micellar liquid chromatography, *J. Liq. Chromatogr. Rel. Technol.* 19 (1996) 101–123.

[6] I. Rapado Martínez, R.M. Villanueva Camañas, M.C. García Álvarez-Coque, Micellar liquid chromatography: A worthy technique for the determination of beta-antagonists in urine samples, *Anal. Chem.* 71 (1999) 319–326.

[7] R.D. Caballero, S. Carda Broch, M.C. García Álvarez-Coque, Hydro-organic and micellar-organic reversed-phase liquid chromatographic procedures for the evaluation of sulphonamides in pharmaceuticals, *Anal. Lett.* 34 (2001) 1189–1203.

[8] M.J. Ruiz Ángel, S. Carda Broch, J.R. Torres Lapasió, E.F. Simó Alfonso, M.C. García Álvarez-Coque, Micellar-organic versus aqueous-organic mobile phases for the screening of β-blockers, *Anal. Chim. Acta* 454 (2002) 109–123.

[9] M.J. Ruiz Ángel, S. Carda Broch, E.F. Simó Alfonso, M.C. García Álvarez-Coque, Optimised procedures for the reversed-phase liquid chromatographic analysis of formulations containing tricyclic antidepressants, *J. Pharm. Biomed. Anal.* 32 (2003) 71–84.

[10] M.J. Ruiz Ángel, J.R. Torres Lapasió, M.C. García Álvarez-Coque, S. Carda Broch, Submicellar and micellar reversed-phase liquid chromatographic modes applied to the separation of beta-blockers, *J. Chromatogr. A* 1216 (2009) 3199–3209.

[11] M.C. García Álvarez-Coque, S. Carda Broch, Direct injection of physiological fluids in micellar liquid chromatography, *J. Chromatogr. B* 736 (1999) 1–18.

[12] M.J. Ruiz Ángel, S. Carda Broch, M.C. García Álvarez-Coque, High submicellar liquid chromatography, *Sep. Purif. Rev.* 43 (2014) 124–154.

[13] J.S. Landy, J.G. Dorsey, Rapid gradient capabilities of micellar liquid chromatography, *J. Chromatogr. Sci.* 22 (1984) 68–70.

[14] J.G. Dorsey, M.G. Khaledi, J.S. Landy, J.L. Lin, Gradient elution micellar liquid chromatography, *J. Chromatogr.* 316 (1984) 183–191.

[15] L.S. Madamba-Tan, J.K. Strasters, M.G. Khaledi, Gradient elution in micellar liquid chromatography. I. Micelle concentration gradient, *J. Chromatogr. A* 683 (1994) 321–334.

[16] L.S. Madamba-Tan, J.K. Strasters, M.G. Khaledi, Gradient elution in micellar liquid chromatography. II. Organic modifier gradients, *J. Chromatogr. A* 683 (1994) 335–345.

[17] A.R. Ghorbani, F. Momenbeik, J.H. Khorasani, M.K. Amini, Simultaneous micellar liquid chromatographic analysis of seven water-soluble vitamins: Optimization using super-modified simplex, *Anal. Bioanal. Chem.* 379 (2004) 439–444.

[18] S.M. Bryant, K.D. Altria, An initial assessment of the use of gradient elution in microemulsion and micellar liquid chromatography, *J. Sep. Sci.* 27 (2004) 1498–1502.

[19] J. Cao, H. Qu, Y. Cheng, Micellar and aqueous-organic liquid chromatography using sub-2 mm packings for fast separation of natural phenolic compounds, *J. Sep. Sci.* 33 (2010) 1946–1953.

[20] R. Nakao, M. Schou, C. Halldin, Direct plasma metabolite analysis of positron emission tomography radioligands by micellar liquid chromatography with radiometric detection, *Anal. Chem.* 84 (2012) 3222–3230.

[21] R. Nakao, M. Schou, C. Halldin, Rapid metabolite analysis of positron emission tomography radioligands by direct plasma injection combining micellar cleanup with high submicellar liquid chromatography with radiometric detection, *J. Chromatogr. A* 1266 (2012) 76–83.

[22] J. Rodenas Montano, C. Ortiz Bolsico, M.J. Ruiz Ángel, M.C. García Álvarez-Coque, Implementation of gradients of organic solvent in micellar liquid chromatography using Drylab®: Separation of basic compounds in urine samples, *J. Cromatogr. A* 1344 (2014) 31–41.

[23] E. Peris García, M.T. Ubeda Torres, M.J. Ruiz Ángel, M.C. García Álvarez-Coque, Effect of sodium dodecyl sulphate and Brij-35 on the analysis of sulphonamides in physiological samples using direct injection and acetonitrile gradients, *Anal. Methods* 8 (2016) 3941–3952.

[24] J. Ke, Y. Dong, T. Luo, Y. Xie, Development of a gradient micellar liquid chromatographic method eluting from micellar mode to high submicellar mode for the rapid separation of free amino acids, *Anal. Methods* 9 (2017) 1762–1770.

[25] Y. Xie, T. Luo, J. Yang, Y. Dong, Rapid determination of amino acids in beer, red wine, and donkey-hide gelatin by gradient elution of HPLC: From micellar liquid chromatography to high submicellar liquid chromatography, *J. AOAC Int.* 101 (2018) 249–255.

[26] X. Li, S. Fritz, Novel additives for the separation of organic compounds by high performance liquid chromatography, *J. Chromatogr. A* 728 (1996) 235–247.

[27] M.C. García Álvarez-Coque, J.R. Torres Lapasió, J.J. Baeza Baeza, Models and objective functions for the optimisation of selectivity in reversed-phase liquid chromatography, *Anal. Chim. Acta* 579 (2006) 125–145.

[28] R. Cela, E.Y. Ordoñez, J.B. Quintana, R. Rodil, Chemometric-assisted method development in RPLC, *J. Chromatogr. A* 1287 (2013) 2–22.

[29] J.R. Torres Lapasió, M.C. García Álvarez-Coque, Levels in the interpretive optimisation of selectivity in high-performance liquid chromatography: A magical mystery tour, *J. Chromatogr. A* 1120 (2006) 308–321.

[30] M.C. García Álvarez-Coque, J.R. Torres Lapasió, J.J. Baeza Baeza, Modelling of retention behaviour of solutes in micellar liquid chromatography, *J. Chromatogr. A* 780 (1997) 129–148.

[31] O. Jiménez, M.L. Marina, Retention modeling in micellar liquid chromatography, *J. Chromatogr. A* 780 (1997) 149–163.

[32] A.P. Boichenko, A.L. Iwashchenko, L.P. Loginova, A.U. Kulikov, Heteroscedasticity of retention factor and adequate modeling in micellar liquid chromatography, *Anal. Chim. Acta* 576 (2006) 229–238.

[33] P. Nikitas, A. Pappa-Louisi, Retention models for isocratic and gradient elution in reversed-phase liquid chromatography, *J. Chromatogr. A* 1216 (2009) 1737–1755.

[34] J.R. Torres Lapasió, M. Rosés, E. Bosch, M.C. García Álvarez-Coque, Interpretive optimisation strategy applied to the isocratic separation of phenols by reversed-phase liquid chromatography with acetonitrile-water and methanol-water mobile phases, *J. Chromatogr. A* 886 (2000) 31–46.

[35] T. Álvarez Segura, C. Camacho Molinero, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Analysis of amino acids using serially coupled columns, *J. Sep. Sci.* 40 (2017) 2741–2751.

[36] P.J. Schoenmakers, H.A.H. Billiet, R. Tijssen, L. de Galan, Gradient selection in reversed-phase liquid chromatography, *J. Chromatogr. A* 149 (1978) 519–537.

[37] L.R. Snyder, J.J. Kirkland, J.L. Glajch, *Practical HPLC Method Development*, 2nd ed., John Wiley & Sons, New York, 1997.

[38] P.J. Schoenmakers, *Optimisation of Chromatographic Selectivity: A Guide to Method Development*, Elsevier, Amsterdam, 1986.

[39] Y.W. Lee, M.S. So, J.W. Lee, S.T. Chung, K.H. Row, Retention models of capacity factor with different compositions of organic modifier in RP-HPLC, *Korean J. Chem. Eng.* 13 (1996) 578–584.

[40] M.C. García Álvarez-Coque, J.R. Torres Lapasió, J.J. Baeza Baeza, Description of the partitioning behaviour of solutes and data treatment in micellar liquid chromatography with modifiers, *Anal. Chim. Acta* 324 (1996) 163–173.

[41] S. López Grío, J.J. Baeza Baeza, M.C. García Álvarez-Coque, Modelling of the elution behaviour in hybrid micellar eluents with different organic modifiers, *Anal. Chim. Acta* 381 (1999) 275–285.

[42] J.R. Torres Lapasió, *MICHROM Software*, Marcel Dekker, New York, 2000.

[43] E. Bonet Domingo, M.J. Medina Hernández, M.C. García Álvarez-Coque, On the direct injection of urine samples in micellar liquid chromatography, *Quim. Anal.* 12 (1993) 167–172.

[44] M.J. Ruiz Ángel, J.R. Torres Lapasió, S. Carda Broch, M.C. García Álvarez-Coque, Performance of short-chain alcohols versus acetonitrile in the surfactant-mediated reversed-phase liquid chromatographic separation of β-blockers, *J. Chromatogr. A* 1217 (2010) 7090–7099.

[45] J. Steflova, M. Stefl, S. Walz, M. Knop, O. Trapp, Comprehensive study on critical micellar concentrations of SDS in acetonitrile-water solvents, *Electrophoresis* 37 (2016) 1287–1295.

[46] M.J. Ruiz Ángel, S. Carda Broch, M.C. García Álvarez-Coque, Effect of short-chain alcohols on surfactant-mediated reversed-phase liquid chromatographic systems, *J. Chromatogr. A* 1217 (2010) 7082–7089.

[47] I.E. Frank, R. Todeschini, *The Data Analysis Handbook: Data Handling in Science and Technology*, Vol. 14, Elsevier Science, Amsterdam, 1994.

[48] S.N. Deming, Y. Michotte, D.L. Massart, L. Kaufman, B.G.M. Vandeginste, *Handbook of Chemometrics and Qualimetrics*, Part B, Elsevier Science, Amsterdam, 1998.

[49] J.A. Navarro Huerta, A. Gisbert Alonso, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Benefits of solvent concentration pulses in retention time modelling of liquid chromatography, *J. Chromatogr. A* 1597 (2019) 76–88.

[50] J.R. Torres Lapasió, J.J. Baeza Baeza, M.C. García Álvarez-Coque, Modeling of peak shape and asymmetry, in *Chemometrics in Chromatography* (edited by Ł. Komsta, Y. Vander Heyden, J. Sherma), CRC Press, Taylor and Francis Group, Boca Raton, FL, 2018, pp. 217–238.

[51] A.K. Smilde, A. Knevelman, P.M.J. Coenegracht, Introduction of multi-criteria decision making in optimization procedures for high-performance liquid chromatographic separations, *J. Chromatogr. A* 369 (1986) 1–10.

Part 2

IMPROVING THE SEPARATION PERFORMANCE FOR CHROMATOGRAPHIC FINGERPRINTS

CHAPTER 8

**ASSISTED BASELINE SUBTRACTION
IN COMPLEX CHROMATOGRAMS
USING THE BEADS ALGORITHM**

8.1. Abstract

The data processing step of complex signals in high-performance liquid chromatography may constitute a bottleneck to obtain significant information from chromatograms. Data pre-processing should be preferably done with little (or no) user supervision, for a maximal benefit and highest speed. In this work, a tool for the configuration of a state-of-the-art baseline subtraction algorithm, called BEADS (Baseline Estimation And Denoising using Sparsity) is developed and verified. A quality criterion based on the measurement of the autocorrelation level was designed to select the most suitable working parameters to obtain the best baseline. The use of a log transformation of the signal attenuated artifacts associated to a large disparity in signal size between sample constituents. Conventional BEADS makes use of trial and error strategies to set up the working parameters, which makes the process slow and inconsistent. This constitutes a major drawback in its successful application. In contrast, the assisted BEADS simplifies the setup, shortens the processing time and makes the baseline subtraction more reliable. The assisted algorithm was tested on several complex chromatograms corresponding to extracts of medicinal herbs analysed with acetonitrile-water gradients, and a mixture of sulphonamides eluted with acetonitrile gradients in the presence of the non-ionic surfactant Brij-35 under micellar conditions.

8.2. Introduction

Modern high-performance liquid chromatography (HPLC) instruments are able to provide highly complex signals in routine analysis, from which the relevant information should be extracted [1]. In these analyses, the data processing step constitutes a bottleneck, constraining sample throughput [2,3]. Problems such as noisy signals, co-eluting peaks (sometimes highly overlapped), peak shifts and the presence of irregular baselines should be addressed. The operations to handle these problems should be done preferably with little (or no) user supervision for a maximal benefit and highest speed.

The aim of this work is to improve the baseline subtraction in chromatograms of high complexity, with complete suppression of problematic drifts and little analyst supervision. Very recently, a new algorithm called "Baseline Estimation And Denoising using Sparsity" (BEADS) was proposed [4,5], which presents as novelty the capability of performing a full decomposition of chromatograms in net signal (i.e., the pure signals of the analytes and their accompanying compounds), baseline and noise. The baseline is modelled as a low frequency signal and the noise as a high frequency contribution, while the peaks of analytes are described as sparse signals, whose first and second derivatives are also sparse (a vector signal is classified as "sparse" when most of its elements are zero). For this purpose, BEADS requires that the user specify several parameters to ensure that the recovered signals have chemical meaning (e.g., positive signals for all analytes). It should be noted that most baseline subtraction algorithms also require some user inputs. This is the case of the mixture models based on splines proposed by Rooi and Eilers [6], the adaptive iteratively reweighted penalised least squares (airPLS) [7], and the backcor algorithm [8].

The authors of BEADS validated it in comparison with the airPLS and backcor algorithms [4]. The three methods yielded reasonable estimates of the baselines, but BEADS offered the best performance. Indeed, in our trials with a variety of chromatograms, BEADS was verified to provide excellent results in complex situations. However, we found some issues that make its routine application to real samples difficult, which should be addressed.

The triple decomposition of chromatograms in BEADS is done essentially by using highly efficient frequency filters, which makes the outline easier and the calculation faster. Moreover, the algorithmic framework is based on majorisation-minimisation [4], which converges quickly regardless of the set of values used in its initialisation. The result of the combination of these techniques is a highly efficient algorithm that saves memory. Another advantage is that, in contrast to other baseline algorithms [8], the set of baselines obtained by BEADS is not described as a parametric family of functions. This feature confers BEADS an extreme flexibility to accommodate any baseline, whatever its complexity.

The limitations of BEADS can be classified in two categories. First, it requires a careful adjustment of the working parameters to properly process real signals of different origin. This operation may be difficult for highly complex signals, owing to the instability of the adjustment process (i.e., small changes in the parameters may lead to very different baselines). Secondly, chromatograms must fulfil some conditions (described in detail in Section 8.4.1), mandatory for the application of BEADS, but hardly fulfilled in practice with real chromatograms.

In this work, we analyse comprehensively the limitations of BEADS, and propose some solutions, which improve the results and reliability of this algorithm and contribute to make it more robust, faster and easier to apply to

chromatograms of real highly complex samples of different origin, with little supervision.

8.3. Experimental

8.3.1. Reagents

In order to explore the correct subtraction of the baseline, several fingerprints of medicinal herbs were processed, corresponding to extracts in hot water of horsetail and decaffeinated teas obtained in our laboratory. For the chromatographic analysis, hydro-organic gradients were prepared with acetonitrile (Scharlab, HPLC grade, Barcelona, Spain) and water. This was buffered at pH 3 with 0.01 M sodium dihydrogen phosphate (Sigma, Roedermark, Germany) and a suitable amount of 0.01 M HCl (Scharlab). The chromatographic signals of extracts of red peony root, taken from Ref. [7], were also processed.

The influence of negative peaks associated with refractometric void volume signals was studied using chromatograms for a mixture of 15 sulphonamides: sulphaguanidine, sulphanilamide, sulphacetamide, sulphadiazine, sulpha-thiazole, sulphapyridine, sulphamerazine, sulphamethazine, sulphamethizole, sulphamonomethoxine, sulphachloropyridazine, sulphamethoxazole, sulpha-soxazole, sulphadimethoxine and sulphaquinoxaline, eluted with an acetonitrile gradient in the presence of Brij-35 (Sigma, St. Louis, MO, USA), buffered at pH 3 with 0.01 M sodium dihydrogen phosphate. All solutions were filtered through 0.45 μm Nylon membranes from Micron Separations (Westboro, MA, USA), before their injection into the chromatographic system.

8.3.2. Preparation of extracts of medicinal herbs

The extracts of horsetail and decaffeinated teas were obtained following the recommendations of Dumarey et al. [9]. For this purpose, 20 mL of nanopure water was added to 0.2 g of ground sample, and boiled in the absence of light. The extracts were filtered through 0.2 μm membrane filters from Pall Gelman Laboratory (Karlstein/Main, Germany), to finally fill 2 ml vials for chromatographic analysis.

8.3.3. Apparatus, columns and software

An Agilent modular instrument (HP 1100, Waldbronn, Germany) was used, consisting of quaternary pump, automatic injector, temperature controller, and variable wavelength UV-visible detector. The chromatograms of the medicinal herbs and mixtures of sulphonamides were detected at 210 and 254 nm, respectively. The column temperature was fixed at 25 ºC. The injection volume was 10 μL, and the flow rate was kept constant at 1 mL/min, in all instances.

An OpenLAB CDS LC ChemStation (Agilent, B.04.03 revision) was used for the acquisition of chromatographic signals. Raw chromatograms were processed without any correction by the ChemStation software, unless those associated to the default working parameters, such as autobalance in the pre-run, 5% zero offset, or attenuation to 1000 mAU. Matlab 2016b (The MathWorks Inc., Natick, MA, USA) was applied for data treatment. The Matlab function [5] (which is included in the Supplementary material of Ref. [4]) was used for the conventional application of BEADS.

8.4. Results and discussion

8.4.1. Limitations of BEADS

As indicated, BEADS makes the simultaneous decomposition of a signal $\mathbf{y}$ in three contributions:

$$\mathbf{y} = [y_1, y_2, ..., y_n] = \mathbf{c} + \mathbf{b} + \mathbf{e} \qquad (8.1)$$

where $\mathbf{c}$, $\mathbf{b}$ and $\mathbf{e}$ make reference to the sparse chromatogram, baseline and noise vectors computed by BEADS, which depend on a set of working parameters $\mathbf{p}$. The working parameters are the cutoff frequency (f_c, which constitutes the boundary between the baseline and the rest of contributions), asymmetry (r, which penalises the negative values) and regularisation parameters (λ_0, λ_1 and λ_2, which control the sparsity of vector $\mathbf{c}$). An additional parameter is the amplitude (A), which multiplies the regularisation parameters; thus, the regularisation parameters are actually $A \times \lambda_i$, which makes the ratios among the λ_i parameters independent of their magnitude.

The adaptability of BEADS to real baselines is noteworthy, but its application has the following limitations, especially severe for complex chromatograms:

  (i) Requirement of the same signal intensity for the first and last points in the chromatogram (i.e., periodicity of the signal).

 (ii) Abnormal risings of the baseline under major signals in chromatograms where the analytes exhibit extreme variations in signal size. The overall appearance of the computed baseline is wavy (see figures discussed in Section 8.4.2), instead of having a smooth trend at large scale.

(iii) Problematic processing of chromatograms containing sporadic negative peaks, such as those corresponding to refractometric signals, or those

observed in chromatograms obtained using indirect UV-visible detection. This forces a careful adjustment of the working parameters for each sample.

(iv) Dependence among the working parameters. The baseline is particularly susceptible to the selected cutoff frequency at low frequencies, which results in an unstable adjustment process. This situation is worsened by the wide range of values to be explored, which in some cases comprises several orders of magnitude (a typical chromatogram composed of 10,000 points can involve exploring cutoff frequencies over 4 orders of magnitude).

(v) Need for each chromatogram of a particular adaptation of the working parameters (i.e., each set of parameters is translated in a different baseline). Fortunately, related samples may share similar parameter values.

For the development of the assisted BEADS, we used a set of 65 multi-analyte chromatograms, all of them with severe problems in their respective baselines. Three of these chromatograms are shown in Figures 8.1 and 8.2. Those in Figure 8.1 were obtained in our laboratory, and correspond to extracts of horsetail and decaffeinated teas (Figures 8.1a and b, respectively). The separation was carried out with a Zorbax Eclipse XDB C18 column (150 mm × 4.6 mm I.D., 5 μm particles, Merck, Darmstadt, Germany), using gradient elution where the acetonitrile content was increased from 20 to 60% (v/v), in a gradient time of 10 min, while the pH was kept at a nominal value of 3. A chromatogram taken from Ref. [7] (Figure 8.2), corresponding to an extract of red peony root, was also analysed. This chromatogram belongs to a set of 10 chromatograms originally processed by the authors to subtract baselines, using airPLS [7], the so-called "faster algorithm for betweenness centrality"

(FABC) proposed by Cobas et al. [10], and the alternating least squares (ALS) algorithm [11].

The chromatogram of the horsetail tea sample (Figure 8.1a) is used to illustrate the performance of the solutions proposed in this work to analyse complex signals, particularly the selection of the best working parameter values to be used by BEADS.

8.4.2. *Monitoring the autocorrelation to explore the BEADS working parameters*

The quality of the results offered by BEADS depends critically on the correct selection of the working parameters, especially the cutoff frequency, which has a major influence in the returned baseline. This relies on the fact that the main principle of BEADS is a decomposition based on the frequency. The other working parameters exhibit milder variations. BEADS parameters are conventionally adjusted by trial and error, and when one parameter is modified, others are collaterally misadjusted. This makes the process slow and unpredictable when a chromatogram with unknown characteristics (without any information about the correct frequency) is processed.

To facilitate the selection of the working parameters when the original BEADS algorithm is used, auxiliary plots were designed (see Section 8.4.4). The auxiliary plots assist in the fast and reliable selection of the working parameters, which makes the application of the original BEADS troublesome. The plots are based on the measurement of the autocorrelation, which can be defined as the correlation of a signal with a delayed copy of itself [12]. Therefore, this property measures the similarity between consecutive data points in a series, such as the measurements taken at regular time intervals, as is the case of a chromatogram.
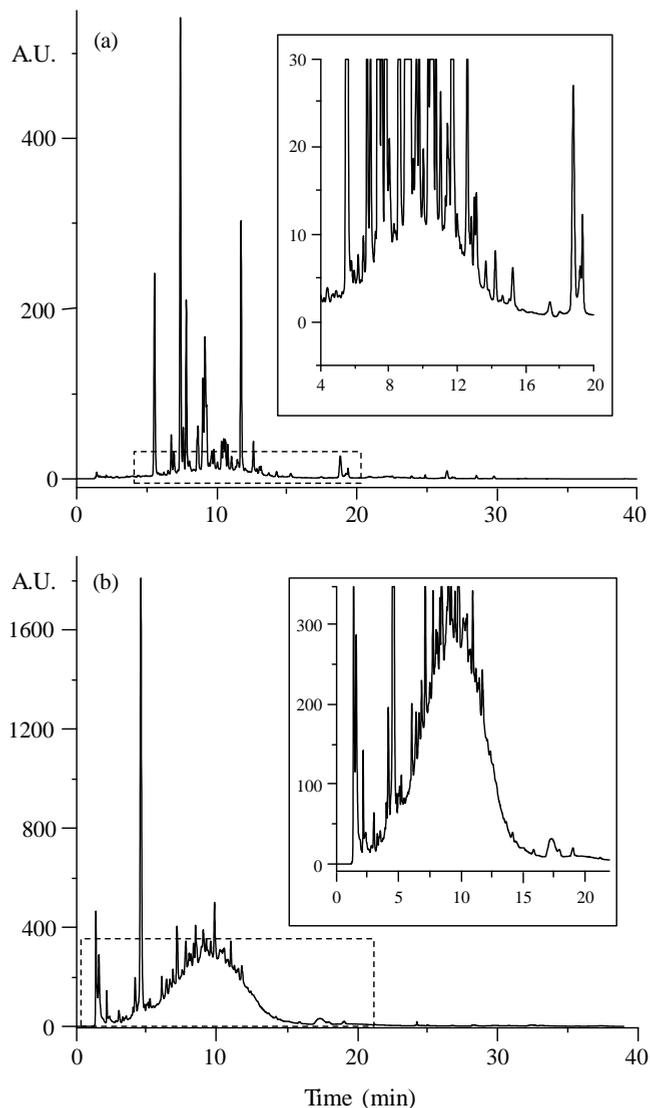
Figure 8.1. Chromatographic fingerprints for extracts of: (a) horsetail tea, and (b) decaffeinated tea, before being processed. The chromatograms were obtained with a 20–60% (*v/v*) acetonitrile gradient reaching the upper concentration in 10 min. The upper inserts magnify the central regions of the chromatograms to highlight the complexity of the baseline associated with the matrix and gradient program.
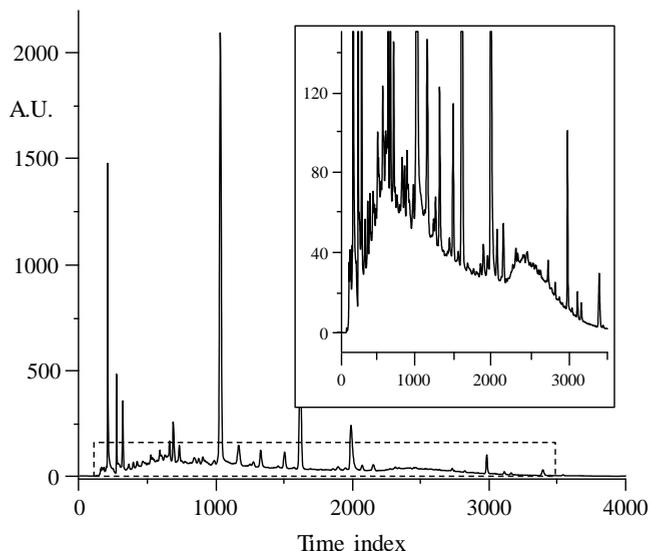
Figure 8.2. Chromatographic fingerprint corresponding to an extract of red peony root, used in Refs. [4,7], before being processed. The upper insert magnifies the central region of the chromatograms to highlight the complexity of the baseline associated with the matrix and gradient program.

The auxiliary plot described in this section will be particularised to the case of the cutoff frequency. We will assume that the other BEADS parameters are more or less correctly set, although this is not so necessary in practice. Our hypothesis is that the removal of a certain feature from the chromatogram, such as the trend in the baseline, even being imperfect, would produce an alteration in the autocorrelation level. If the selected cutoff frequency were correct, by subtracting from the total chromatographic signal the contributions of the sparse chromatogram and baseline estimated by BEADS, only noise would remain. Ideally, this noise should not show any autocorrelation. At any other relatively close cutoff frequency, the decomposition in sparse chromatogram and noise will not be perfect, and some autocorrelation will persist. Therefore,

the observation of the changes in the autocorrelation level will help to select the features to be removed. It also will reveal that a certain feature has been removed in a particular parameter domain.

In this work, we have measured the autocorrelation based on the Durbin-Watson (*DW*) statistic [13]:

$$DW = \frac{\sum\limits_{i=2}^{n}(d_i - d_{i-1})^2}{\sum\limits_{i=1}^{n}d_i^{\,2}} \tag{8.2}$$

Conventionally, this statistic is applied to regression analysis and smoothing, d_i being the difference between the raw signal for point i minus the fitted (or smoothed) signal for that point. Therefore, d_i estimates the lack of fit in the case of fitting, and the noise in the case of smoothing. In the case of applying BEADS, an estimation of the noise can be obtained from Equation (8.1):

$$\mathbf{e} = \mathbf{y} - \mathbf{c} - \mathbf{b} \tag{8.3}$$

This operation is equivalent to subtracting a real signal from the adjusted or smoothed signal, and allows the direct application of Equation (8.2), making the difference vector $\mathbf{d} = \mathbf{e}$

The *DW* statistic can be developed as follows:

$$DW = \frac{\sum\limits_{i=2}^{n}d_i^{\,2} - 2\sum\limits_{i=2}^{n}d_i\,d_{i-1} + \sum\limits_{i=2}^{n}d_{i-1}^{\,2}}{\sum\limits_{i=1}^{n}d_i^{\,2}} \approx \frac{2\sum\limits_{i=2}^{n}d_i^{\,2} - 2\sum\limits_{i=2}^{n}d_i\,d_{i-1}}{\sum\limits_{i=1}^{n}d_i^{\,2}} =$$

$$= \frac{2\sum\limits_{i=2}^{n}d_i^{\,2}}{\sum\limits_{i=1}^{n}d_i^{\,2}} - \frac{2\sum\limits_{i=2}^{n}d_i\,d_{i-1}}{\sum\limits_{i=1}^{n}d_i^{\,2}} = 2 - 2r \tag{8.4}$$

where r measures the autocorrelation level for vector **d**. DW tends to 0 for a perfect positive correlation ($r = 1$), and to 4 for a perfect negative correlation ($r = -1$). As indicated, when the contributions of the analytes (i.e., sparse chromatogram) and baseline are perfectly subtracted from the raw chromatogram, only noise will remain. If this is white noise, it should not exhibit any autocorrelation and DW will tend to 2, since $r = 0$. In practice, the autocorrelation experiences a drop around the optimal cutoff frequency, without necessarily reaching a null value.

Monitoring the DW statistic, which ranges between 0 and 4, there is low probability of reaching the ideal $DW = 2$ (denoting null autocorrelation). In order to measure the autocorrelation, the following expression:

$$r^2 \approx \frac{(2 - DW)^2}{4} \tag{8.5}$$

was found more convenient in practice. If the signal pre-treatment assures that the first and last points in the chromatogram match ($d_1 = d_n$), Equation (8.5) will be exact (i.e., not an approximation).

Figure 8.3a plots the autocorrelation measured as r^2 (estimated from the noise **e**) versus the cutoff frequency in a logarithmic scale, for the analysis of the extracts of the three samples described above (Figures 8.1 and 8.2). This figure should be analysed together with Figure 8.4, which shows the baselines obtained by BEADS corresponding to the cutoff frequencies marked in Figure 8.3a, where several regions can be observed.
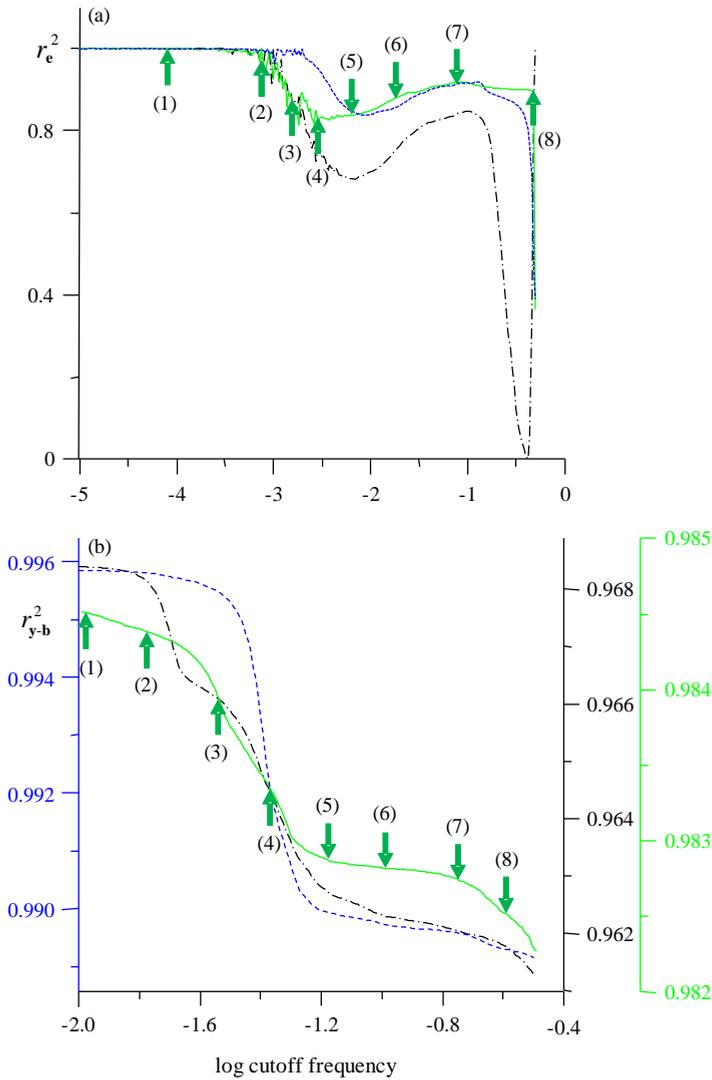
Figure 8.3. Autocorrelation plots expressed as r^2 and the logarithm of the cutoff frequency used for the subtraction of the baseline: (a) original scale where r_e^2 was calculated from the noise, and (b) logarithmic scale where r_{y-b}^2 was calculated from the baseline corrected signal. Extracts: horsetail tea (continuous line and far right y-axis in (b)), decaffeinated tea (short dashed line and left y-axis), and red peony root (dotted dashed line and near right y-axis).
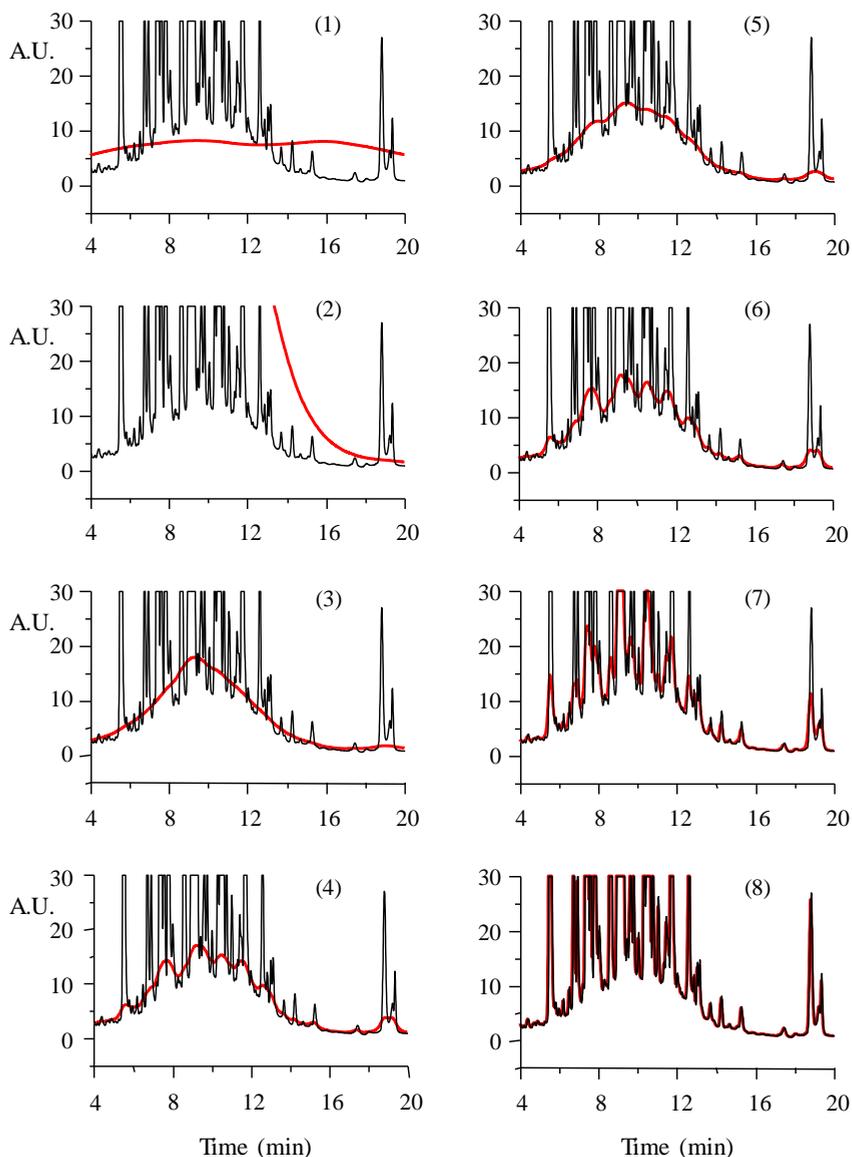
Figure 8.4. Exploration of the cutoff frequency used to subtract the baseline from the chromatogram in Figure 8.1a (horsetail tea), using the original scale. The frequency values correspond to the points marked in Figure 8.3a. The estimated baseline has been overlapped on an enlarged section of the chromatogram.

When the selected cutoff frequency is too low (point (1)), nearly flat baselines are subtracted that scarcely affect the autocorrelation (only the vertical shift of the whole chromatogram is corrected, which is insufficient in most situations). Intermediate cutoff frequencies tend to eliminate several contributions of the baseline at large scale (points (4) and (5)). In this region, the cutoff frequency will be the ideal. Beyond these frequencies, BEADS tends to eliminate gradually the contribution of the analytes and baseline signals, attenuating the peaks (point (7)), until only noise remains (point (8)). At even higher cutoff frequencies, all contributions including the noise would be eliminated, leaving a null vector (in this case, r^2 cannot be calculated, because Equation (8.2) becomes undefined owing to the division by zero). In other words, the baseline to be subtracted would be equal to the raw signal. The observation of the results in Figures 8.3a and 8.4 leads to the conclusion that the observed minimum in the autocorrelation plot points out the optimal cutoff frequency (that one giving rise to the best baseline subtraction), which is specific for each sample.

When the raw signal is processed with BEADS, small variations in the selected cutoff frequency may be translated into large variations in the baseline (see Figure 8.5). This behaviour is observed especially at intermediate frequencies, below the optimal one. In addition, cutoff frequencies above the optimal (frequencies (6) to (8)) make the baseline undesirably sensitive to the peak magnitude. The described disturbances can be cancelled, at least to some extent, by a careful adjustment of the asymmetry and regularisation parameters.
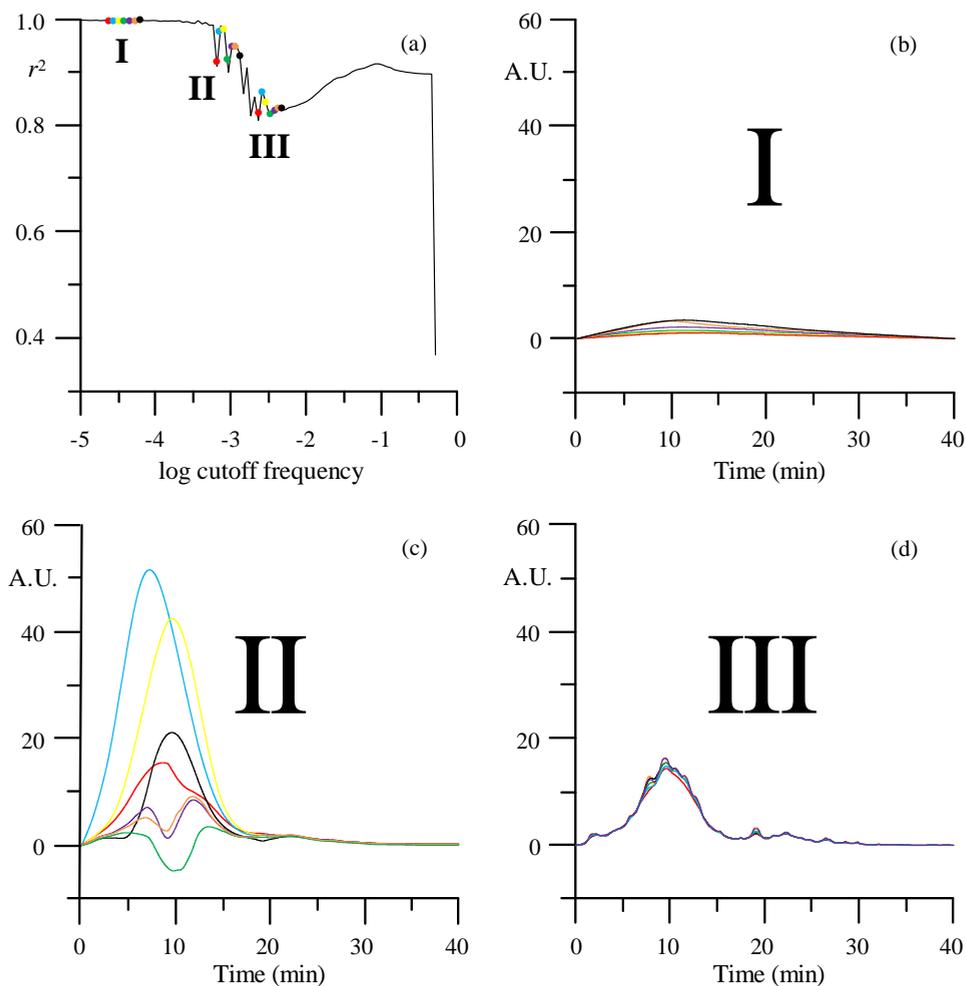
Figure 8.5. Instability of the baselines obtained for the chromatogram of horsetail tea extract (Figure 8.1a), using neighbour cutoff frequencies in different regions of the autocorrelation plot for the original signal (Figure 8.3a): (a) autocorrelation plot, (b) low frequencies, (c) intermediate frequencies, and (d) frequencies in the optimal region.

To sum up, each feature eliminated from the chromatogram at a certain cutoff frequency results in a domain of characteristic r^2 values, starting by a value close to one when no contribution has been removed yet, up to values close to zero when even the noise has been removed. However, as observed for points (4) to (6) in Figure 8.4, even at the best cutoff frequencies, some irregularities (ripples under the main peaks) remain in the baseline. This problem is addressed in Section 8.4.3.2.

8.4.3. Enhancements in the application of BEADS

As commented, particular working parameters for each type of sample and signal are needed for the routine application of BEADS. The quality of the results depends critically on the experience and skill of the analyst. In addition, the process may become excessively slow and prone to subjectivity. On the contrary, using the proposed auxiliary autocorrelation plots, described in Sections 8.4.2 and 8.4.4, BEADS can be quite easily adapted to any kind of sample, reducing the subjectivity in the selection of the working parameters, and providing always reliable results. In addition to the hard selection of the optimal parameters, other limitations of BEADS have been described, which make its practical use for baseline subtraction in complex chromatograms troublesome. Some proposals to overcome each limitation are indicated below.

8.4.3.1. Periodicity of the chromatogram

The correct application of BEADS requires periodic signals: if the signal values at the extremes of the chromatogram differ, artefacts will appear. In a first step, we considered solving the requirement of periodicity at the extremes of the chromatogram, through the subtraction of the straight-line that connects

the first and last points. However, some problems may appear when the slopes at both extremes differ. Only a careful trial and error adjustment of the regularisation parameters can mitigate this. We found that a more practical solution was the subtraction of a parabola, since it is able to fully cancel incidental differences in the slopes at the start and end of the chromatogram (see Figure 8.6).

BEADS is based on the use of high pass filters, which allow all features above a critical selected frequency survive (sparse chromatogram and noise), whereas the lower frequency features are cancelled (baseline, and any added feature to correct the periodicity problem, such as the parabola). The process of correction implies the treatment of a distorted signal ($\mathbf{y'}$) with BEADS, where a parabola has been subtracted to the raw signal in order to make the slopes at the extremes identical. In these conditions and as a result of the high-pass filter, BEADS will give a correct estimation of the sparse chromatogram ($\mathbf{c}$) and noise ($\mathbf{e}$), but a biased baseline ($\mathbf{b'}$):

$$\mathbf{y'} = \mathbf{c} + \mathbf{b'} + \mathbf{e} \tag{8.6}$$

The correct baseline ($\mathbf{b}_{corr}$) can be easily recovered by adding the parabola ($\mathbf{p}$) previously subtracted:

$$\mathbf{b}_{corr} = \mathbf{b'} + \mathbf{p} = \mathbf{b'} + (\mathbf{y} - \mathbf{y'}) \tag{8.7}$$
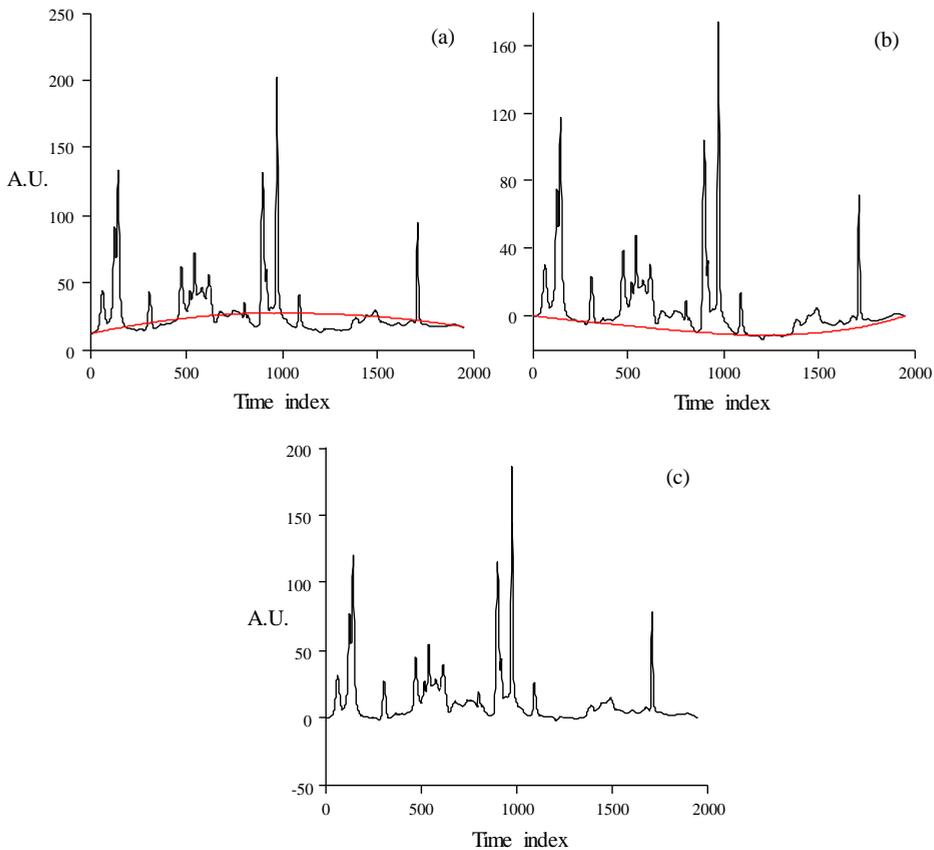
Figure 8.6. Parabolic correction of the signal to fulfil the periodicity condition: (a) Original signal and parabola obtained by fitting the first and last two points in the chromatogram (red line); (b) chromatogram after the subtraction of the parabola, so that the slopes are zero at both extremes, and baseline obtained when such chromatogram is processed by BEADS (red line); (c) final chromatogram when the baseline is subtracted to the distorted chromatogram. The use of a high pass filter allows removing completely the distortion produced by the parabola, since it is a low frequency feature. Sparse chromatogram and noise are unaffected. The processed data correspond to a chromatogram of horsetail extract between 18 and 30 min, monitoring the absorbance at 320 nm.

8.4.3.2. Chromatograms involving peaks with extremely different magnitude

As we showed in Figure 8.4, chromatograms with extreme differences in peak size give rise to ripples when processed by BEADS. To eliminate the influence of the highest peaks on the baseline in such chromatograms, there are at least two solutions. The most straightforward treatment is clipping the highest peaks, so that the signal cannot exceed a selected height. We have explored this strategy with chromatograms of diverse complexity. Clipping works fine with relatively simple chromatograms, but for complex chromatograms with bulky baselines, the ripples remain (see Figure 8.7).

The second solution is using a log transformation of the signal, which is compatible with the operations of the original BEADS algorithm. The signal is transformed to the logarithmic scale after subtracting its minimal value, slightly increased with an arbitrary positive offset, ε:

$$\mathbf{z} = \log\left(\mathbf{y} - \min\left(\mathbf{y}\right) + \varepsilon\right) \tag{8.8}$$

The larger the offset, the less aggressive the pre-treatment. We decided to use an offset $\varepsilon = 1$. This value is appropriate regarding the magnitude of the signals being processed, which reach maxima around 500–10,000. Another reason for selecting $\varepsilon = 1$ is because if $y_i = \min\left(\mathbf{y}\right)$, then $\log\left(y_i - \min(\mathbf{y}) + 1\right) = \log 1 = 0$.
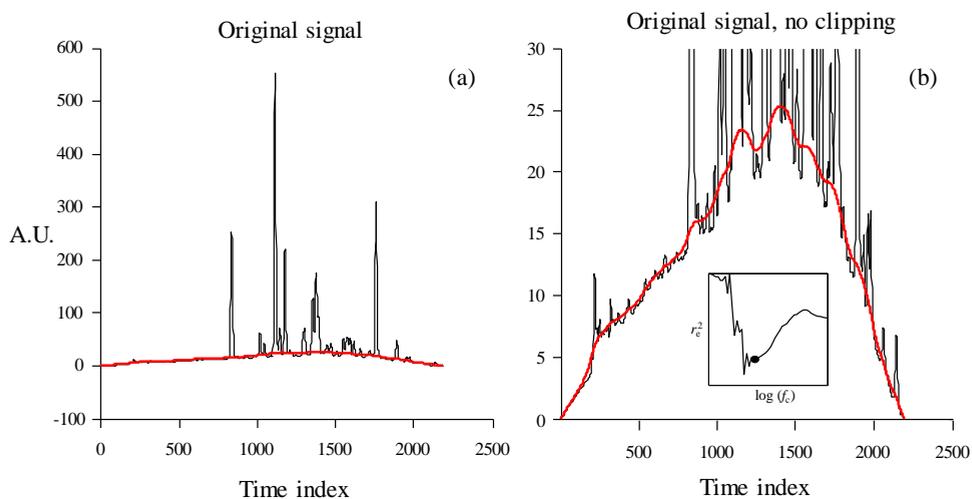
Figure 8.7. Effect of the application of signal clipping on the baselines obtained by BEADS when signals involving peaks of highly dissimilar size are processed: (a) Full signal showing the complete variation in peak size. (b,c,d) Magnified chromatogram and corresponding baselines obtained: (b) without clipping. The inserts show the autocorrelation plots for: (b and c) the noise (r_e^2, where the best cutoff frequency was determined finding the minima in the autocorrelation plots, indicated as black dots), and (d) the chromatogram minus the baseline (or sparse chromatogram plus noise, where the best cutoff frequency was indicated by the last horizontal step). The processed data correspond to the first 15 min of a chromatogram of horsetail extract, monitoring the absorbance at 210 nm.
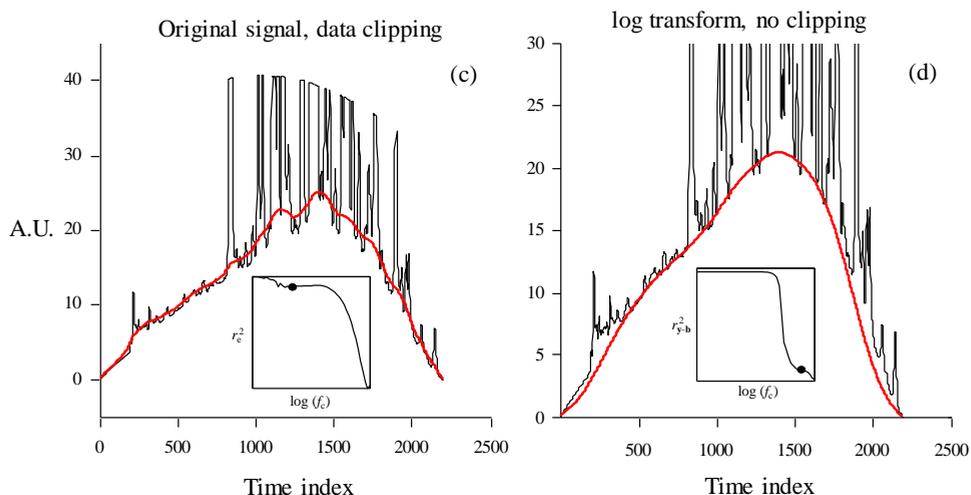
Figure 8.7 (continued). Effect of the application of signal clipping on the baselines obtained by BEADS when signals involving peaks of highly dissimilar size are processed: (c) after clipping those signals larger than 60 A.U., and (d) without clipping and applying the log transformation of the signals.

The log transformation reduces the weight of the largest peaks along the BEADS operation, and as a result, the ripples of the baseline that appear under the main peaks totally disappear. Since the magnitude of the ripples in the baseline under the peaks is correlated to the size of the signals (the higher the peak, the higher the ripple), by operating in the logarithmic scale, the ripples will only be perceptible at very high frequencies. Naturally, after applying BEADS to the log transformation, the results should be back-transformed to the original scale.

The decomposition of the raw signal ($\mathbf{y}$) and the log transformed signal ($\mathbf{z}$), using BEADS, can be denoted as:

$$\mathbf{y} = \mathbf{c_y} + \mathbf{b_y} + \mathbf{e_y} \qquad (8.9)$$

$$\mathbf{z} = \mathbf{c_z} + \mathbf{b_z} + \mathbf{e_z} \qquad (8.10)$$

Considering Equation (8.8), the back transformation of the BEADS results obtained in the logarithmic scale can be carried out as follows:

$$\mathbf{b}_{\text{corr}\mathbf{y}} = 10^{\mathbf{b_z}} + \min(\mathbf{y}) - \varepsilon \qquad (8.11)$$

$$(\mathbf{c} + \mathbf{e})_{\text{corr}\mathbf{y}} = \mathbf{y} - 10^{\mathbf{b_z}} - \min(\mathbf{y}) + \varepsilon \qquad (8.12)$$

Even if the correct BEADS working parameters are used, it will be not possible to differentiate the contributions of the sparse chromatogram and noise, once the log transformation has been applied, because this affects the sparsity of the signal and its derivative, and prevents the persistence of linearity (Equation (8.9)), once returned to the original scale. This has another consequence: the best cutoff frequency cannot be selected from the noise. As will be shown in Section 8.4.4, it can be still obtained from $\mathbf{y} - \mathbf{b}_{\text{corr},\mathbf{y}}$.

It should not be forgotten here that the objective is to correct the raw signal by removing the baseline. This is obtained by subtracting from the raw signal the back transformed baseline.

8.4.3.3. Sporadic negative signals

BEADS can be applied in different ways for the processing of asymmetrical signals (chromatograms with sporadic negative peaks). A first possibility is taking advantage of the asymmetry parameter to set the level of tolerance to negative signals in the sparse chromatogram, using trial and error. This

treatment is slow and has no guarantee of success. We propose an alternative, which consists in running BEADS repeatedly, in iterations, using a fixed value of the asymmetry parameter suitable for positive signals. Along the iterations, and after each BEADS evaluation, those points below a certain threshold under the baseline are replaced by the corresponding values of the baseline found in the current iteration. This iterative replacement process is repeated until convergence, or up to a given maximal number of iterations is fulfilled. Proceeding in this way, not only the problems associated with the negative signals were eliminated, but also spurious contributions (which break the general trends) disappeared. This process does not affect the peak heights.

8.4.4. *Autocorrelation plot using the baseline-corrected signal*

As a consequence of the log transformation, the noise returned by BEADS cannot be used to estimate the autocorrelation. Instead, the signal corrected by subtracting the returned baseline can be used to monitor the changes in the baseline:

$$\mathbf{y}_{bcorr} = \mathbf{c} + \mathbf{e} = \mathbf{y} - \mathbf{b} \tag{8.13}$$

Even though it is not possible to obtain an unbiased estimation of the noise in the original scale, Equation (8.2) can still be applied to estimate the autocorrelation, by taking d_i as the difference between the $\mathbf{y}_{bcorr}$ signal for points i and $i - 1$. Therefore, the consistency of the variations around point i in a window of three points ($i - 1$, i and $i + 1$) is monitored. This means that there is no proper residual for making the comparison, and $\mathbf{y}_{bcorr}$ includes a correlated contribution (the sparse chromatogram, $\mathbf{c}$). However, as will be shown below, monitoring the autocorrelation of the baseline corrected signal can still be useful to set the best working parameters in BEADS.

Indeed, we have found that a plot of $r_{\mathbf{y-b}}^{2}$ (Equation (8.5) applied to the log transform), as a function of the cutoff frequency and considering the full chromatogram vector (Figure 8.3b), is very useful to detect the most appropriate cutoff frequency. This plot should be compared with the plot in Figure 8.3a, where the autocorrelation corresponds to the noise ($r_{\mathbf{e}}^{2}$), without applying any transformation to the data. Both plots show different patterns that depend on the use of the original scale (Figure 8.3a), or the log transformation (Figure 8.3b), and on the kind of data from which the autocorrelation is measured: the noise (Figure 8.3a) or the baseline corrected chromatogram (Figure 8.3b). When the noise is processed, the plot exhibits a minimum at intermediate cutoff frequencies (see Section 8.4.2), whereas the use of the baseline corrected chromatogram leads to a stepped plot. The value of $r_{\mathbf{y-b}}^{2}$ decreases as the diverse baseline contributions to the chromatogram are removed. Each horizontal region in the plot corresponds to a consistent baseline returned by BEADS in a given frequency interval. When the contributions of the peaks of analytes, baseline and noise disappear completely, the autocorrelation of the residuals should be ideally $r_{\mathbf{y-b}}^{2} = 0$. We have observed from a collection of chromatograms that the optimal cutoff frequency is close to the centre of the last step at higher frequencies, that is, around the last inflection point (point (6) in Figure 8.3b). In practice, it is convenient to select slightly lower cutoff frequencies (i.e., a point between the beginning and the centre of the last horizontal region in the autocorrelation plot), to attenuate somehow the flexibility of the baseline.
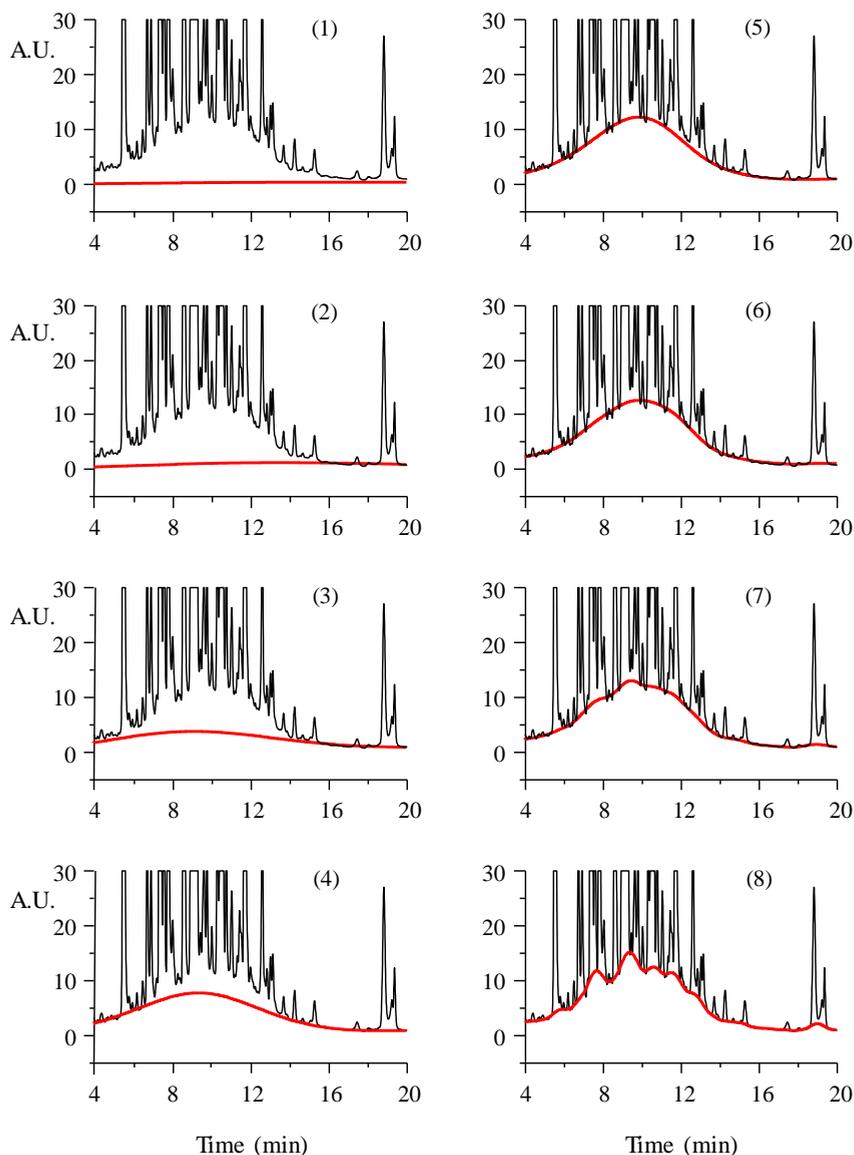
Figure 8.8. Exploration of the cutoff frequency used to subtract the baseline from the chromatogram in Figure 8.1a (horsetail tea), using the log transformation of the signal. The frequency values correspond to the points marked in Figure 8.3b. The estimated baseline has been overlapped on an enlarged section of the chromatogram.

Figure 8.8 illustrates the impact of the cutoff frequency on the baseline subtraction, for the chromatogram of the horsetail tea extract (Figure 8.4 corresponds to the results obtained with BEADS with the original scale, see also Figure 8.3a). The baselines found by BEADS using different cutoff frequencies are overlapped on the chromatograms. For the same cutoff frequency, the baselines calculated from the direct signal (Figure 8.4), and its log transformation (Figure 8.8), do not match. The baselines found using the log transformation of the signal were much more satisfactory, for all assayed chromatograms. Also, an important aspect to remark is that the selection of the cutoff frequency becomes less critical when the signal is translated to the logarithmic scale. The cutoff frequencies (1) to (4) marked in Figure 8.3b are too low, while frequencies (7) and (8) overfit the baseline (i.e., unreal ripples appear under the peaks). For frequencies (5) and (6), the baseline can be considered highly satisfactory.

The fine adjustments of the other working parameters (asymmetry, r, and regularisation parameters, λ_0, λ_1 and λ_2) used in BEADS are shown in Figures 8.9 to 8.13. As observed, in all instances, stepped plots are obtained and the optimal parameter value is close to an inflection point. However, by adjusting only the cutoff frequency after setting approximate values for the remaining parameters, highly satisfactory results were found in all assayed cases by operating with the logarithm of the signal.
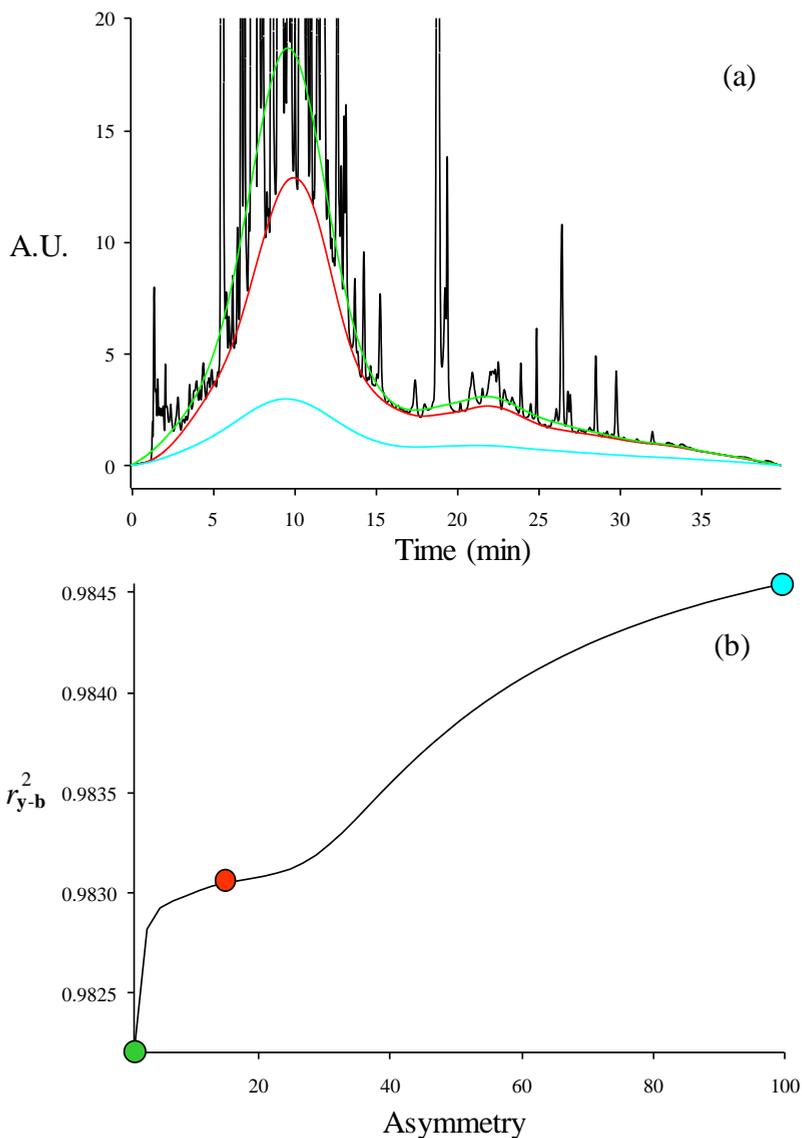
Figure 8.9. Chromatogram (a) and autocorrelation plot (b) for the asymmetry parameter. The red dot and red baseline correspond to the optimal value of the parameter. Two more baselines obtained for the extreme lowest (green) and highest (cyan) parameter values are given for comparison purposes. As observed, the optimal value is obtained in the central inflection point.
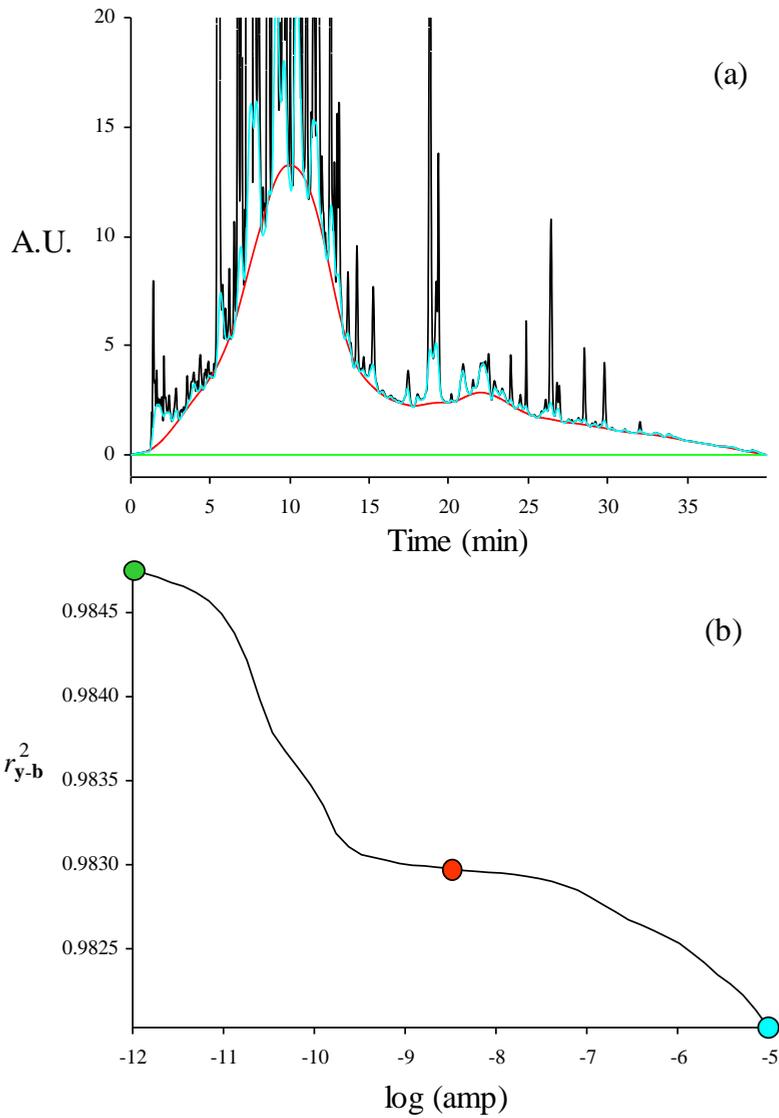
Figure 8.10. Chromatogram (a) and autocorrelation plot (b) for the amplitude
parameter. See Figure 8.9 for other details.
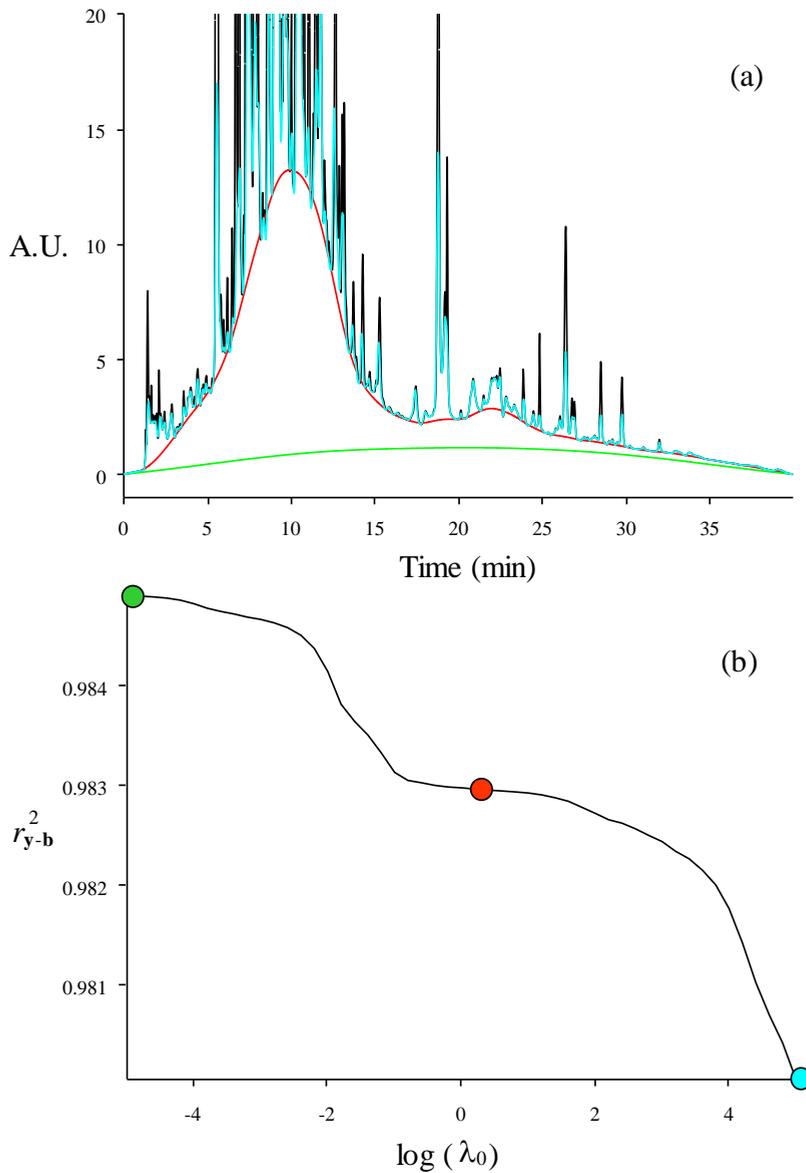
Figure 8.11. Chromatogram (a) and autocorrelation plot (b) for the λ_0 parameter. See Figure 8.9 for other details.
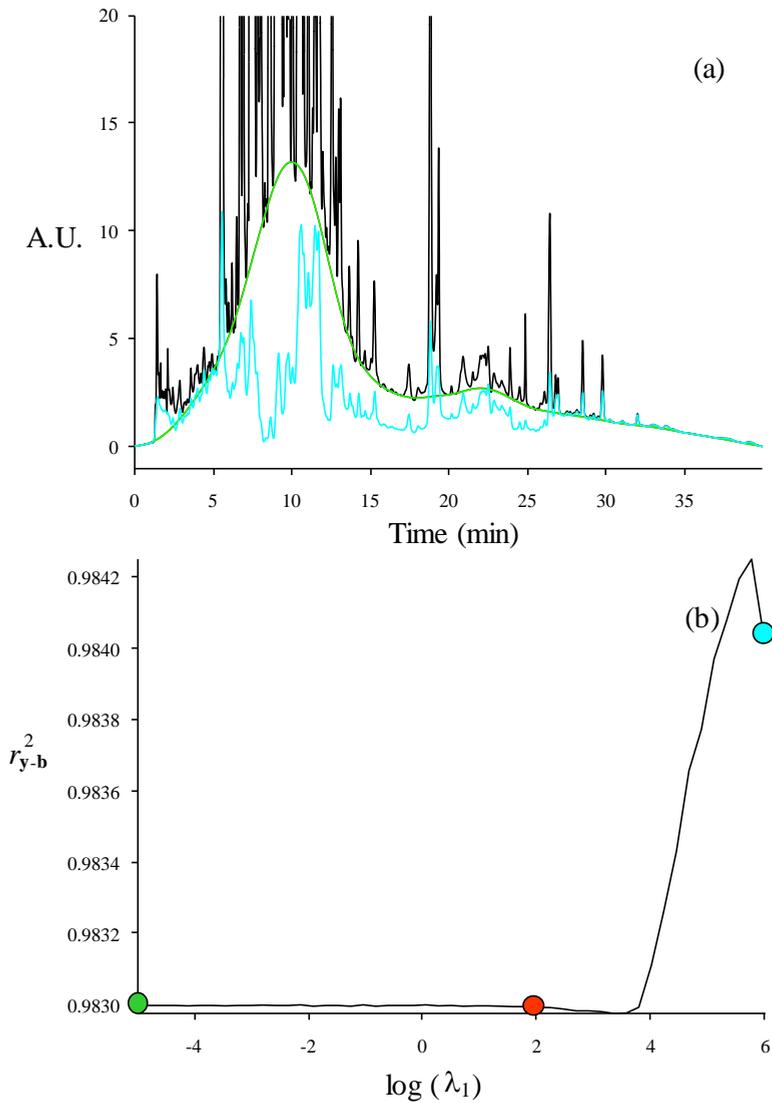
Figure 8.12. Chromatogram (a) and autocorrelation plot (b) for the λ_1 parameter. Low to intermediate values of λ_1 have scarce effect on r_{y-b}^2. See Figure 8.9 for other details.
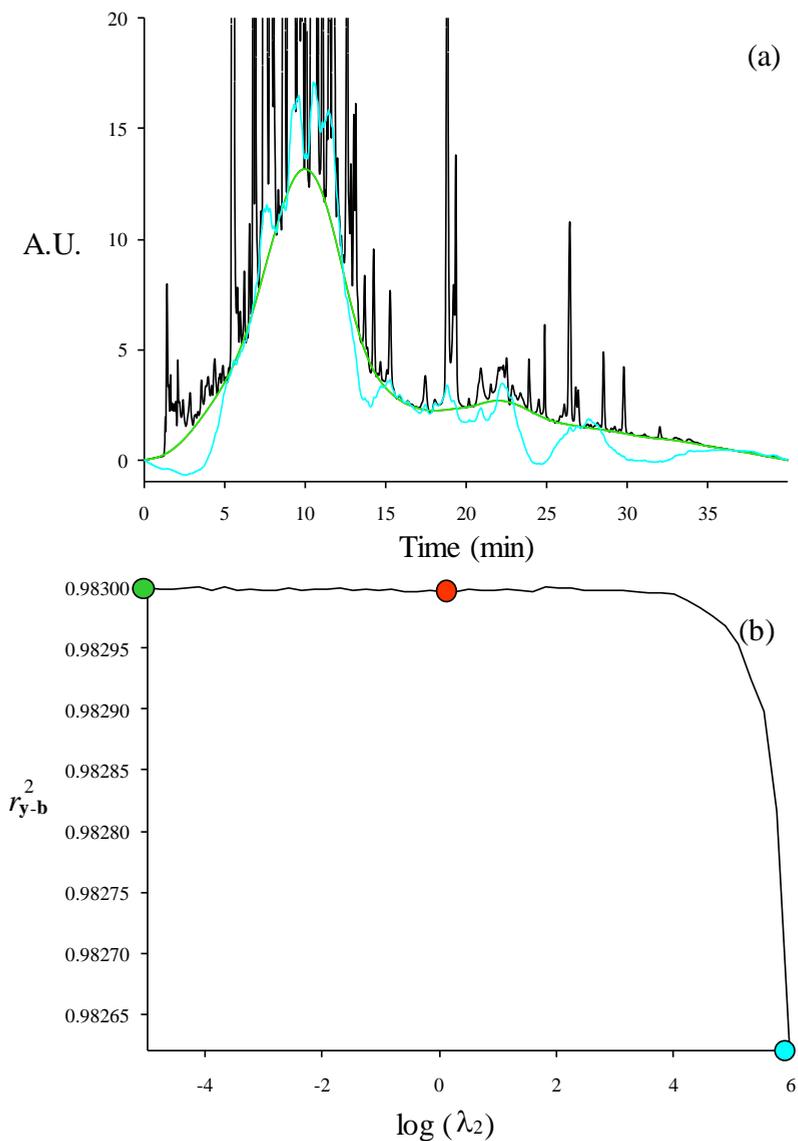
Figure 8.13. Chromatogram (a) and autocorrelation plot (b) for the λ_2 parameter. Low to intermediate values of λ_2 have scarce effect on $r_{\text{y-b}}^2$. The selection of this parameter is not critical: any small value is valid. See Figure 8.9 for other details.

8.4.5. Application of the assisted BEADS

Figure 8.14 illustrates the baseline found after selecting the optimal cutoff frequency, using the log transformation of the signal, for full chromatograms of the three samples of medicinal herbs (Figures 8.1 and 8.2). Figure 8.15 shows the corresponding final baseline-corrected chromatograms. The result is highly satisfactory in all instances.

Figure 8.16 shows the chromatogram of a mixture of sulphonamides (see Section 8.4.3.3), eluted with a gradient of acetonitrile from 0 to 20% (*v/v*), reaching the upper concentration in 30 min in the presence of 0.01 M Brij-35. The separation was carried out using a Chromabond C18 column (150 mm × 4.6 mm I.D., 5 μm particle diameter, Scharlab). In the chromatogram, a refractometric perturbation associated with the mixing of the sample and mobile phase appears close to the void volume. Figure 8.16a shows the baseline in successive iterations, where the points below the negative threshold are replaced by the respective predicted baseline points. In Figure 8.16b, the baselines to be subtracted according to the original BEADS and applying the proposed approach are overlapped. The original BEADS required a modification of all working parameters by trial and error, and the compensation of the negative signal was less perfect. Figure 8.16c shows the baseline-corrected chromatogram according to the proposed approach.
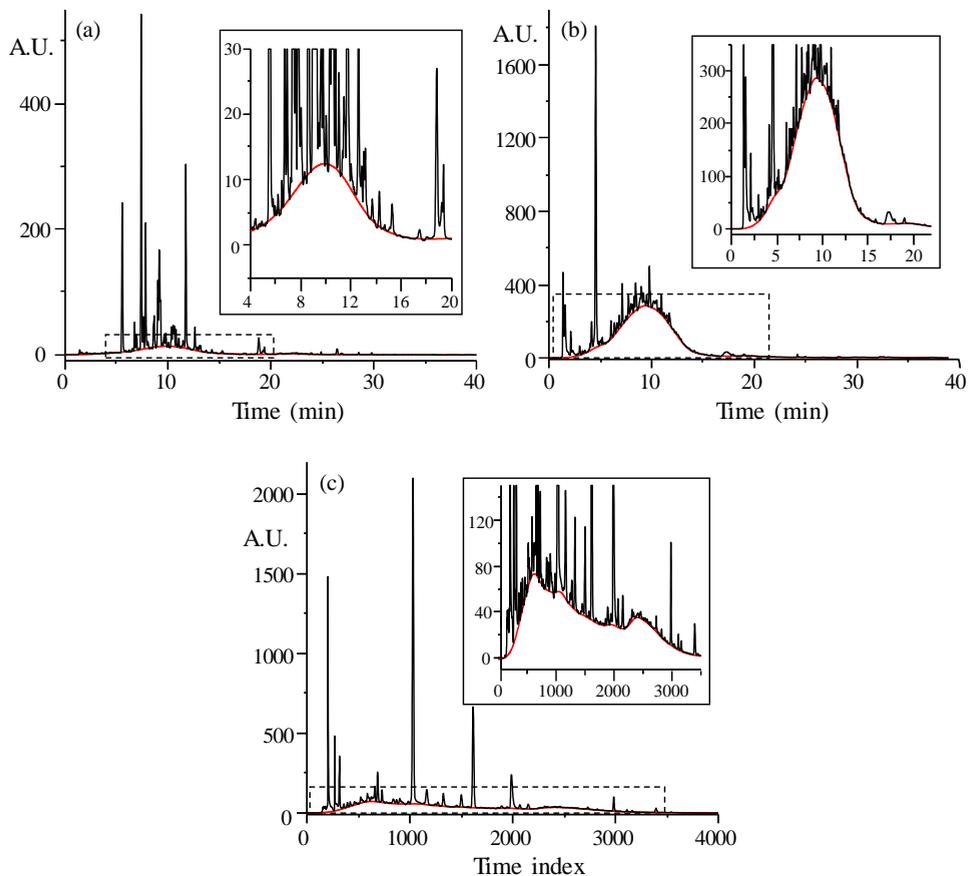
Figure 8.14. Chromatographic fingerprints of medicinal herbs: (a) horsetail tea, (b) decaffeinated tea, and (c) extract of red peony root taken from Ref. [7], with the optimal baseline overlapped, using the assisted BEADS algorithm. Cutoff frequency: (a) 0.105 (see also Figure 8.3b), (b) 0.132 and (c) 0.130. The upper inserts magnify the central regions of the chromatograms to allow a better inspection.
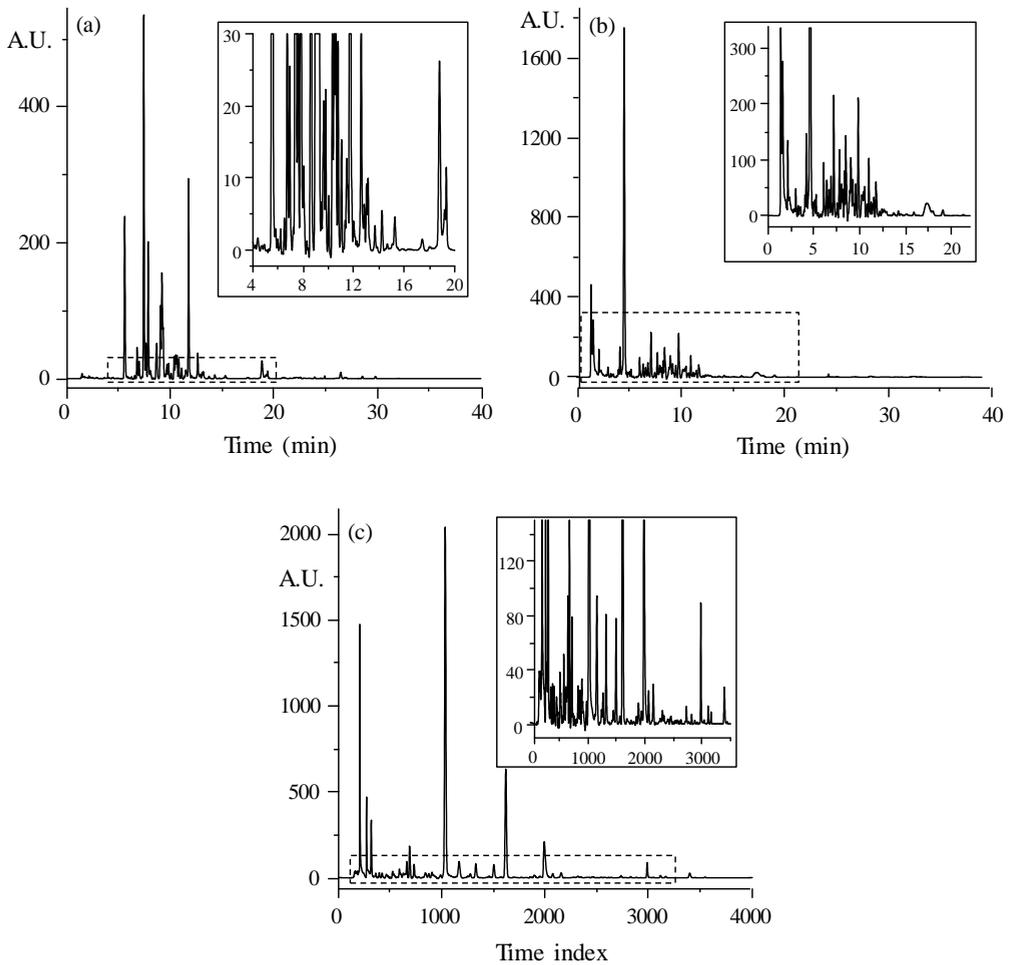
Figure 8.15. Baseline corrected chromatograms for (a) horsetail tea, (b) decaffeinated tea, and (c) extract of red peony root (see Figure 8.14 for the unprocessed signals and the found baselines). The upper inserts magnify the central regions of the chromatograms to allow a better inspection.
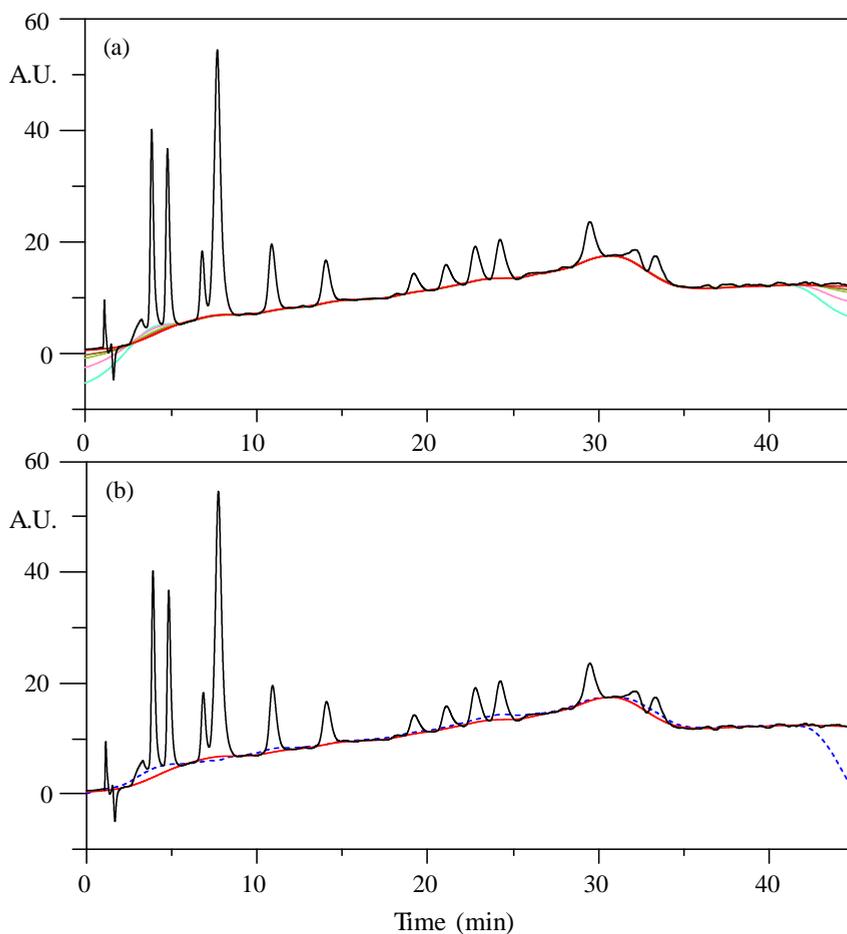
Figure 8.16. Chromatogram showing refractometric negative peaks, corresponding to the elution of a mixture of 15 sulphonamides, using gradient elution with acetonitrile in the presence of Brij-35: (a) progress of the iterations showing the successive baselines up to reach convergence, (b) baseline obtained using the iterative substitution (continuous line) versus that one obtained with the original BEADS (dashed line).
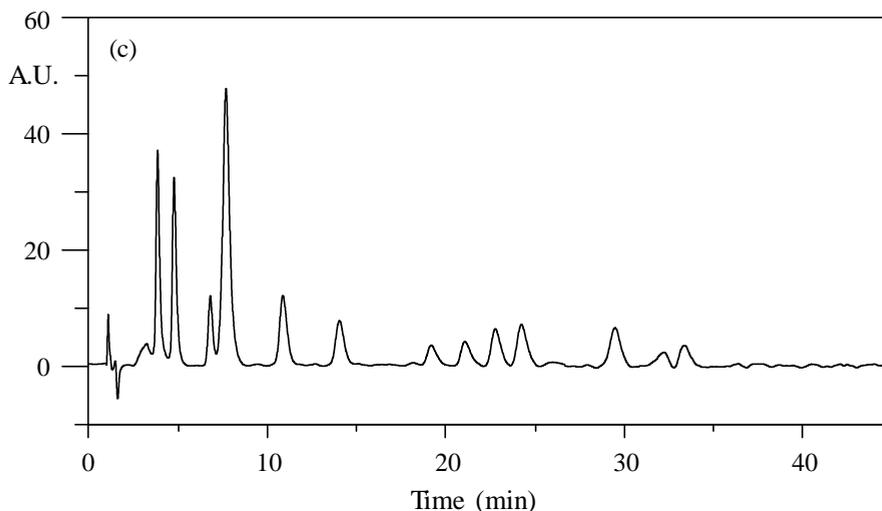
Figure 8.16 (continued). Chromatogram showing refractometric negative peaks, corresponding to the elution of a mixture of 15 sulphonamides, using gradient elution with acetonitrile in the presence of Brij-35: (c) final chromatogram after baseline subtraction using the iterative substitution. In spite of the presence of a negative signal, the same value of asymmetry parameter was used as in Figures 8.14 and 8.15 (which showed only positive peaks).

8.4.6. Quantification

Appraising properly the consequences of a global baseline correction on peak quantification is not easy, since they depend on a number of factors difficult to parametrise. For instance, the results depend on the mutual magnitude of the peaks and the size and frequencies of the baseline fluctuations. Thus, when a poor baseline is subtracted, the recovery error in a large signal can be much smaller than the corresponding error for a small signal with a good baseline correction. Other factors affecting the results are the peak

location in the chromatogram (e.g., in an empty region or at the extremes of the chromatogram), the surroundings of the peak to be quantified (e.g., an isolated peak, a peak in a cluster or a peak in the neighbourhood of a major constituent), the presence of noise or negative signals, among others.

Figure 8.17 gives an idea of the errors that could be expected after BEADS baseline subtraction. The figure shows the chromatogram of 7 sulphonamides, eluted with Brij-35, overlaying the corresponding found baseline. Three artificial peaks (marked as A, B and C) were added in independent in silico experiments, for calculating the errors. The three peaks had the same area (35.00 units), asymmetry factor (1.23) and plate count (8700). For each peak, the chromatogram corresponding to the 7 sulphonamides plus the added peak were processed by BEADS to recover the baseline. The recovered area was then calculated after subtracting the overall signal and the baseline for the respective peak (global baseline correction). In addition, the area obtained by fitting the local baseline around each peak in the respective global chromatogram was also calculated (local baseline correction).

The relative errors for the three peaks (global and local corrections) were: Peak A (0.11%, 0.014%), peak B (2.5%, 3.7%), and peak C (4.9%, 4.0%). As expected, the magnitude of the errors is correlated with the retention time, since the peaks become wider and the weight of the area of the residual signals in the original chromatogram under the peak of interest is increased. Also, the relative errors obtained with a local baseline, which only considers the surroundings of the peak of interest, are usually smaller, since fitting a global baseline implies losing details in particular regions of the chromatogram. In spite of this, the magnitude of the errors is comparable for the global and local baselines.
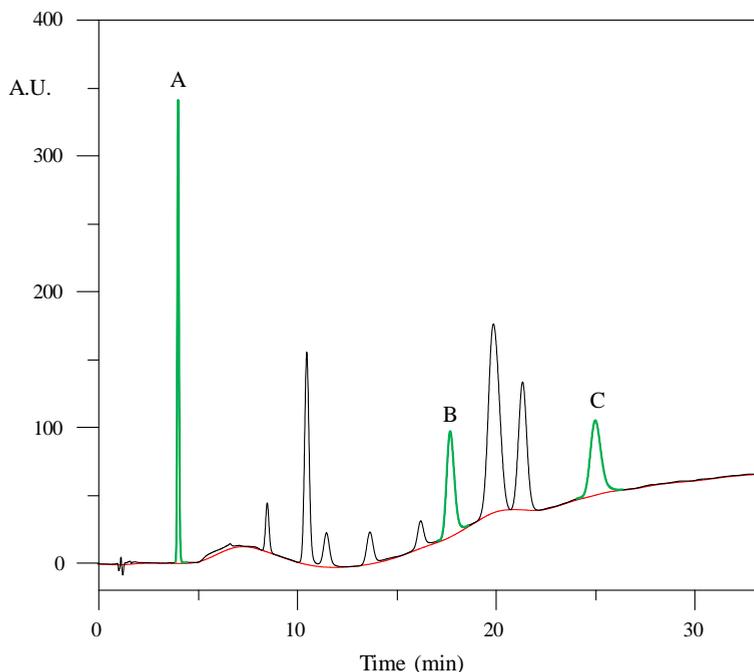
Figure 8.17. Chromatogram of a sample containing 7 sulphonamides, eluted with Brij-35. The original signal is drawn in black, and the respective baseline corrected using the assisted BEADS is overlaid in red. Peaks A, B and C (marked in green) were added in independent artificial experiments (see text for more details).

8.5. Conclusions

The main problem of applying BEADS is the need to set properly the parameters for each specific signal. A correct setup of the BEADS working parameters in its original formulation is difficult to establish, particularly the cutoff frequency, which is by far the most critical working parameter. This work proposes an auxiliary autocorrelation plot to assist in the selection of the

optimal cutoff frequency, which is also valid for adjusting the other working parameters. The irregularities in the baseline associated to large differences in scale between major and trace components (i.e., baseline ripples appearing under the main peaks) are solved by replacing the raw signal by its log transformation.

With the assisted BEADS, the selection of the optimal frequency is less critical. The subtraction of the baseline using straightforwardly BEADS requires some experience and a selection of the working parameters by trial and error, owing to the mutual dependence and sensitivity among them. In contrast, the use of autocorrelation plots and the log transformation allows a fast, simple and reliable selection of the cutoff frequency and other working parameters. The third improvement is an iterative algorithm that discards sporadic negative signals breaking the general trend of the baseline, such as refractometric peaks or transitions associated to gradients.

Our long-term aim is the optimisation of the separation conditions for complex samples, such as chromatographic fingerprints, whose baselines are notably irregular. The origin of these problematic baselines is the complexity of the matrix, together with the use of gradients to expedite the analyses. The evaluation of such chromatograms forced to search a method capable of adjusting very complex baselines. Ideally, the method should be reliable and require few or no user interaction. The assisted BEADS provided very satisfactory results in all assayed examples (about 65 chromatograms), and needs little supervision.

The decomposition of the net signal in sparse chromatogram, baseline and noise shows a certain level of mutual dependence, so that the net chromatogram has peaks significantly smaller, even after a correct baseline subtraction. Thus, for certain applications, such as the quantification of peaks, processing the net

chromatogram is risky, and the noise can be overestimated in regions of the chromatogram where peaks are found. Therefore, it is preferable to subtract only the baseline and process the resulting signal by other methods able to eliminate the noise, such as the Savitsky-Golay smoother [14]. The whole process, from loading signals to obtaining the final table of results takes a few seconds.

BEADS, in its original formulation, was also suggested for signals of other nature, such as electrocardiograms. Therefore, the tools developed in this work may also improve signals coming from fields different from chromatography. It should be also mentioned that some of the proposed solutions are also valid for other baseline subtraction algorithms. Thus, for instance, the autocorrelation plots can be useful for configuring other parametric baseline estimation approaches.

As commented, BEADS may suffer from transient artifacts at signal endpoints. These periodicity errors can be solved by reformulating the filter used in the original BEADS algorithm, as recently proposed by Selesnick [15].

8.6. References

[1] S. Fanali, P. Haddad, C.F. Poole, M.L. Riekkola (eds.), *Liquid Chromatography: Fundamentals and Instrumentation*, 2nd ed., Elsevier, Amsterdam, 2017.

[2] A. Felinger, *Data Analysis and Signal Processing in Chromatography*, Elsevier, Amsterdam, 1998.

[3] Ł. Komsta, Y. Vander Heyden, J. Sherma (eds.), *Chemometrics in Chromatography*, CRC Press, New York, 2017.

[4] X. Ning, I.W. Selesnick, L. Duval, Chromatogram baseline estimation and denoising using sparsity (BEADS*)*, *Chemometr. Intell. Lab. Syst.* 139 (2014) 156–167.

[5] https://es.mathworks.com/matlabcentral/fileexchange/49974-beads--baseline-estimation-and-denoising-w--sparsity--chromatogram-signals-?requestedDomain=www.mathworks.com

[6] J.J. de Rooi, P.H.C. Eilers, Mixture models for baseline estimation, *Chemom. Intell. Lab. Syst.* 117 (2012) 56–60.

[7] Z.M. Zhang, S. Chen, Y.Z. Liang, Baseline correction using adaptive iteratively reweighted penalized least squares, *Analyst* 135 (2010) 1138–1146.

[8] V. Mazet, C. Carteret, D. Brie, J. Idier, B. Humbert, Background removal from spectra by designing and minimizing a non-quadratic cost function, *Chemometr. Intell. Lab. Syst.* 76 (2005) 121–133.

[9] M. Dumarey, I. Smets, Y. Vander Heyden, Prediction and interpretation of the antioxidant capacity of green tea from dissimilar chromatographic fingerprints, *J. Chromatogr. B* 878 (2010) 2733–2740.

[10] J.C. Cobas, M.A. Bernstein, M. Martín Pastor, P.G. Tahoces, A new general-purpose fully automatic baseline-correction procedure for 1D and 2D NMR data, *J. Magn. Reson.* 183 (2006) 145–151.

[11] P.H.C. Eilers, Parametric time warping, *Anal. Chem.* 76 (2004) 404–411.

[12] I.E. Frank, R. Todeschini, *The Data Analysis Handbook*, Elsevier, Amsterdam, 1994.

[13] J. Durbin, G.S. Watson, Testing for serial correlation in least squares regression. III, *Biometrika* 58 (1971) 1–19.

[14] A. Savitzky, M.J.E. Golay, Smoothing and differentiation of data by simplified least squares procedures, *Anal. Chem.* 36 (1964) 1627–1639.

[15] I. Selesnick. Sparsity-assisted signal smoothing (revisited), *IEEE Int. Conf. Acoust., Speech, Signal Proc.* (2017) (March).

CHAPTER 9

**STUDY OF THE PERFORMANCE
OF A RESOLUTION CRITERION TO
CHARACTERISE COMPLEX CHROMATOGRAMS
WITH UNKNOWNS OR WITHOUT STANDARDS**

9.1. Abstract

The search of the best conditions in liquid chromatography is routinely carried out with information provided by chemical standards. However, sometimes there are samples with insufficient knowledge about their chemical composition. In other cases, identities of the components are known, but there are no standards available, or the identities of peaks in chromatograms taken under different conditions are ambiguous. Most resolution criteria used to measure the separation performance cannot be applied to these samples. In this work, a global resolution function valid for all situations was developed based on automatic measurements of peak prominences (area fraction exceeding the line that joins the valleys delimiting each peak). The relative performance of this criterion is evaluated against the peak purity criterion (which measures the area free of overlapping). Peak purity provides a truly comprehensive measurement of global resolution since the underlying signals for each compound are used. However, it is only accessible through *in silico* simulations. In contrast, peak prominences are not based on a comprehensive knowledge of the individual signals, and can be obtained from experimental chromatograms or *in silico* simulations. Therefore, this criterion is suitable for the direct evaluation of the resolution of chromatograms with high complexity. A comparison study was carried out based on the agreement of the gradients chosen as Pareto-optimal by both criteria, using information from standards of the 19 primary amino acids found in proteins. The developed global resolution function was applied with success to chromatographic fingerprints of medicinal herbs.

9.2. Introduction

Scientific and technological advances are increasingly demanding more powerful analytical techniques. Therefore, the development of not only more sophisticated instrumentation and materials, but also strategies and chemometric methodologies, are necessary to solve new problems. Reversed-phase liquid chromatography is nowadays the most extended chromatographic technique due to its wide field of application, reliability, robustness, and sensitivity [1]. However, its efficiency is limited in comparison with other analytical techniques such as gas chromatography, capillary electrophoresis, and other electromigration techniques. This is especially detrimental for the analysis of samples with complex compositions. Arbitrarily selected experimental conditions rarely provide enough selectivity. This problem can be minimised with the use of highly selective detectors such as mass spectrometry (MS). Nevertheless, despite the high selectivity of this detection technique, the inclusion of a well-designed separation step is still needed [2–4]. Moreover, MS detection is still out of reach for many laboratories. Thus, an optimisation of the separation conditions is always helpful.

Independently the peaks in a chromatogram being known or unknown, it is required that their mutual separation be as large as possible for both identification and quantification. An extreme case, where the relative peak distribution and magnitude is the relevant feature, is found in chromatographic fingerprints and classification studies. In these cases, better resolution will offer more informative chromatograms.

A satisfactory separation implies obtaining sufficient chromatographic resolution in an acceptable analysis time. This is achieved by the appropriate adjustment of the experimental conditions. When complex samples are analised, the determination of optimal conditions by trial and error can become

an expensive, slow, and inefficient process. *In silico* simulations are the most efficient tools to discover the best separation conditions [5–7]. A practical way to carry out a search is the measurement of the separation quality using a chromatographic objective function (COF) [6–19]. Usually, this type of function requires information about the retention and profile of chromatographic peaks to establish models in order to predict chromatograms in a wide range of conditions, without the need of performing further experimental assays.

Two main strategies have been traditionally used to evaluate the quality of a separation: (i) monitoring expected resolution for the worst resolved peak pair, and (ii) combining the resolution values for all peaks in a chromatogram; both strategies are applied to a set of unassayed conditions involving an experimental design. The second strategy is more interesting when the aim is obtaining complete (or almost complete) separation for all peaks in a chromatogram. However, conventional COFs are not appropriate in all instances to quantify the resolution level, since they are dominated by the worst resolved compounds; when at least one compound appears significantly overlapped under all experimental conditions, the global resolution will always be close to zero, even when all other compounds were fully resolved.

We have proposed new approaches to find the best separation in situations of extremely low chromatographic resolution. These approaches consist of monitoring the number of sufficiently resolved peaks in the chromatograms (the "peak count") [15,20,21]. They are aimed to quantify the degree of success in the separation (they quantify the well resolved peaks), contrary to conventional COFs, which quantify the degree of failure (they attend mainly to the worst resolved peaks).

A further level of complexity consists in optimising the chromatographic resolution for samples where some or even all compounds are unknown, or for which no chemical standards are available. As indicated above, most COFs reported in the literature are based on predictive models, which must be fitted using chromatographic information obtained from standards for all constituents. In the absence of standards, predicting the best separation conditions is not possible using conventional COFs. This is the case of Snyder's R_S resolution and the peak purity (area free of overlapping, see Figure 9.1), which require simulated chromatograms for their evaluation. In contrast, other conventional COFs, such as the so-called "selectivity" (α), and valley-to-peak functions [21], can be straightforwardly measured from experimental chromatograms, without knowledge of the underlying signals of each compound. However, when two or more peaks are overlapped, these functions can offer misleading conclusions. In addition, they are related to peak pairs, and reversals in the elution order give rise to discontinuities. This is also the case for the Snyder's R_S resolution.

In previous work, Álvarez Segura et al. [22] proposed a new COF to solve such problematic situations, which was called "peak prominence". This criterion is based on the measurement of the protruding part of chromatographic peaks (area fraction exceeding the line that joins the valleys delimiting each visible peak, see Figure 9.2). An advantage of this COF is that it is applicable to experimental chromatograms without the need of standards. In that work, peaks in the sample were ranked according to the areas of peak prominences, and a threshold of cumulative area was established to discriminate between significant peaks and those that were irreproducible in replicated injections. The number of significant peaks in fingerprint chromatograms was then used to optimise the extraction and conservation conditions of medicinal herbs [23].
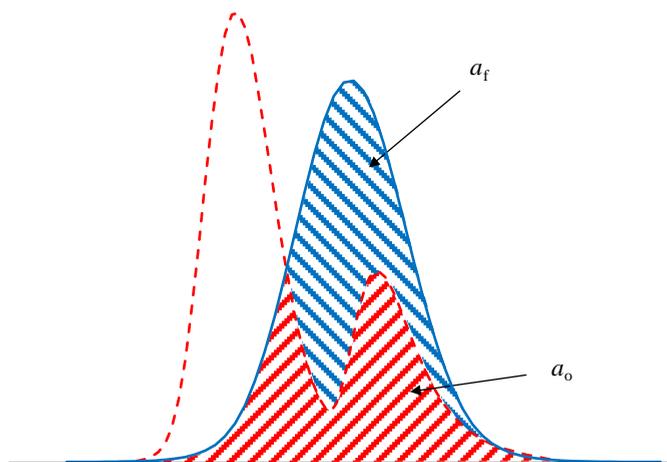
Figure 9.1. Peak purity criterion. The free area fraction (a_f) and overlapped area with other peaks in the chromatogram (a_o) is shown. The total peak area (a_T) is the sum of a_f and a_o.
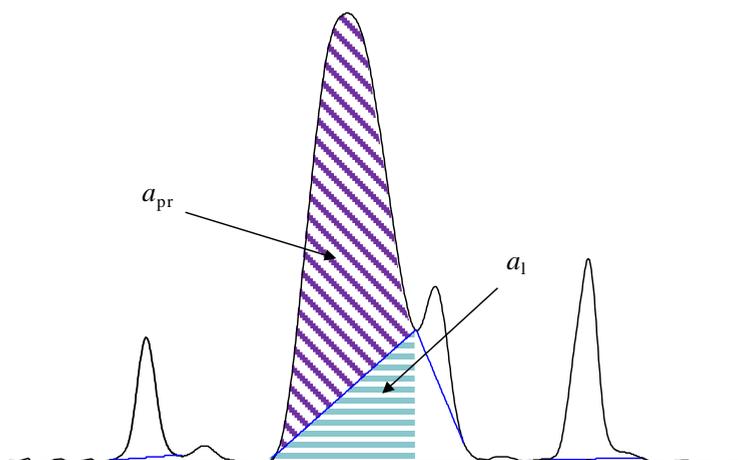
Figure 9.2. Peak prominence criterion. The line that joins the valleys delimiting each peak divides it in two regions of area a_{pr} and a_l. The total peak area (a_T) is approximated to the sum of a_{pr} and a_l.

This work represents another step in the development of the peak prominence as a resolution criterion. The main objective was to check to what extent the results found as optimal agree with those given by a reference function, which intrinsically has an exhaustive knowledge about the number of eluting compounds and profile of the peak of each compound. The peak purity criterion was chosen for this comparison, since it provides accurate and reliable estimations of resolution, and matches closely with the assessment of resolution of experienced analysts [11]. The study implies the development of a proper global function and inspection of its performance in a number of controlled situations gradually closer to reality. Finally, the selected function was adapted to cope with complex chromatograms containing unknown compounds or without available standards.

9.3. Theory

9.3.1. Peak purity

A reliable measurement of the resolution requires information not only about the position of the chromatographic peaks, but also about their full profile. In 1986, Schoenmakers wrote a pioneering work in the field of chromatographic optimisation, where he described the use of the "overlapped peak fraction" to measure accurately the resolution [5]. However, for many years it was no more than a proposal, as it required knowledge of the position and profile for each peak, and complex and laborious numerical calculations with the assistance of software. Fortunately, the proposal of new more practical peak models, together with the development of computers in the last decades and its widespread use in laboratories, have returned the interest in this

criterion. Thus, a function that measures the peak purity (the complement of the overlapped fraction) was proposed in our laboratory [6].

The peak purity quantifies the percentage of peak area for a given analyte free of interference, considering as such all other peaks in the chromatogram (see Figure 9.1). It may be expressed as follows:

$$p_i = \frac{a_f}{a_T} = \frac{a_f}{a_f + a_o} = 1 - \frac{a_o}{a_f + a_o} \tag{9.1}$$

where a_o is the area under the analyte peak overlapped by a hypothetical chromatogram built with the peaks of the accompanying compounds in the sample (the overlapped area), and a_f is the peak area free of interference (the free area). The resolution value obtained in this way tends to zero when the overlap of the analyte peak with the peaks of the interferences is total, and reaches $p_i = 1$ when the peak is fully resolved. However, it should be noted that the peak purity depends on the relative peak areas.

Even in situations where the chromatograms contain peaks remarkably deformed and largely overlapped, the peak purity shows an excellent correlation with the assessment of the resolution of expert analysts. For this reason, it has been considered as the best measurement of resolution [19]. It also has a number of features, which generally make it the most appropriate criterion.

(i)   Its meaning is very intuitive: it correlates with the information the analyst is interested on, that is, the interference level. For example, a value of 0.98 peak purity simply means that 98% of the peak of interest is free of interference (in other words, it shows 2% of interference or overlap).

(ii)   It provides a realistic evaluation of the separation capability of the system, and can be easily applied to situations of diverse complexity, taking into account the full signal (peak profile, size and noise).

(iii) It is an inherently normalised measurement, which facilitates the combination of elementary resolution values into a single global measurement and the combination with other quality criteria.

(iv)   One of the most important features, due to the consequences that it entails, is the qualification of individual peaks rather than peak pairs, so there is no possibility of unambiguous relationships between the identities of the peaks and the numerical resolution values. In addition, knowledge of the identity of the neighbouring peaks is not as important as it is for the criteria related to peak pairs, like the classical R_S criterion. All this allows operations such as peak weighting or exclusion easier, which avoids problems associated to peak identities in situations of peak reversals.

The concept of peak purity has allowed the development of new optimisation strategies. On the one hand, the fact that it is able to anticipate the maximal resolution capability of the separation system is particularly useful for dealing with situations of low resolution, where conventional resolution criteria fail [15]. On the other hand, it allows the simultaneous optimisation of two or more mobile phases, eluents and/or columns, or even separation techniques (complementary situations) [24]. The only drawback is that it is designed for the evaluation of the resolution through simulations and it is hard to apply directly to experimental chromatograms of mixtures, since individual contributions are not available.

9.3.2. Peak prominence

The peak prominence is an elementary resolution criterion recently developed [22], which can be classified in the group of valley-to-peak functions, but with a significant difference: in conventional valley-to-peak functions, the maxima of two adjacent peaks are compared to a property of the valley that lies between them [25], whereas in the new function, the area of a peak is delimited between the two valleys that separate it from other peaks (see Figure 9.2).

The aim of this new resolution criterion is to quantify the relationship between the size of the peak area that is above the valleys that define it (or above the baseline) and its total area:

$$pr_i = \frac{a_{\text{pr},i}}{a_{\text{T}}} = \frac{a_{\text{pr},i}}{a_{\text{pr},i} + a_{\text{l},i}} \tag{9.2}$$

The peak prominence has several advantageous features for measuring the chromatographic resolution:

 (i)   It is a normalised function, which facilitates its interpretation.
 (ii) It allows the inclusion of the size ratio between neighbouring peaks.
(iii) It qualifies individual peaks, instead of peak pairs.
(iv)   It does not require the measurement of the properties of the peaks obtained from standards.

The latter feature differentiates the peak prominence from the peak purity, which requires the information of individual signals obtained through retention and peak profile models, established through design of experiments. Hence, the peak prominence is an ideal criterion for the measurement of the resolution of chromatographic fingerprints and, in general, of the experimental chromatograms of any sample.

9.4. Experimental

Two types of samples were considered:

(i) Solutions of the 19 primary amino acids found in proteins, for which standards were available: (1) Aspartic acid, (2) glutamic acid, (3) asparagine, (4) serine, (5) glutamine, (6) histidine, (7) glycine, (8) arginine, (9) threonine, (10) alanine, (11) cysteine, (12) tyrosine, (13) valine, (14) methionine, (15) isoleucine, (16) tryptophan, (17) phenylalanine, (18) leucine, and (19) lysine. Prior to chromatographic separation with acetonitrile-water mixtures, the amino acids were derivatised with *o*-phthalaldehyde (OPA) and *N*-acetylcysteine (NAC), so they could be monitored at 335 nm. This set of amino acids was used to study the performance of the peak prominence criterion compared to the peak purity. The purpose of using these compounds was to have experimental information about the characteristics of the peaks for each analyte. Experimental information on the chromatographic behaviour of the amino acid derivatives was taken from our laboratory database. More details on the chromatographic procedure are given elsewhere [26].

(ii) Extracts of decaffeinated and horsetail teas (bought at a local supermarket), were prepared according to the recommendations given by Dumarey et al. [27]. These samples were used as representatives of cases where no standards are available to build predictive models of retention and peak profiles. Fingerprints of the extracts were obtained with a modular Agilent chromatograph (Model HP 1100, Waldbronn, Germany), consisting of a quaternary pump, autosampler, thermostatted column compartment, and UV-visible detector set at 210 nm. The

injection volume was 10 µL, and the mobile phase flow rate was kept constant at 1 mL/min. Analyses were carried out with a Zorbax Eclipse XDB-C18 column (150 mm × 4.6 mm i.d. with 5 µm particle size, Agilent), using linear gradients of acetonitrile, buffered with 0.01 M sodium dihydrogen phosphate (Sigma) and 0.01 M HCl (Scharlab) at pH 3.

9.5. Data treatment

A MATLAB (2016b version, The MathWorks Inc., Natick, MA, USA) function was developed for automatic measurement of relevant information in complex chromatograms. This function detects and measures all peaks present in a chromatogram up to a certain critical peak size (area or height) threshold, which depends on the noise level. It also merges secondary signals, and discards spikes and noise peaks. A number of peak properties are measured for those peaks surviving the refinement process. One of them is the peak prominence, which is obtained from the optimised tangent line between the two valleys defining each peak. More details are given elsewhere [22,23].

9.6. Results and discussion

A study was carried out to check whether the peak prominence criterion was able to discover the same optimal conditions as those selected by the peak purity criterion, which in previous work has shown excellent performance [11,15,19,26,28,29]. Obviously, when all peaks are resolved, both criteria will agree, but in conditions where there are fewer visible peaks than compounds, the differences may become notorious. With this aim, several synthetic study cases involving the separation of the OPA-NAC derivatives of amino acids

using 1081 linear acetonitrile-water gradients were generated. Note that for this study we needed a sufficiently complex set of compounds with standards available to be able to predict the resolution according to the peak purity criterion.

Three definitions of global peak prominence were examined, investigating the following situations: existence of peaks in a chromatogram with the same or different areas, presence of noise, and presence of unknown compounds. The best global resolution function was further refined to account for situations where the number of components is not well defined, as is the case of chromatographic fingerprints. Since information from standards is required, the peak purity of fingerprints cannot be calculated.

9.6.1. Selection of optimal conditions for a sample with standards available

Chromatographic training data for the amino acid derivatives consisted of a number of isocratic experiments using acetonitrile in the range 5.0–27.5% (*v/v*) [26]. This information yielded very accurate predictions for linear gradients, suitable for the studies designed for this work. The predicted chromatograms were computed following a methodology described elsewhere [28,29] (see also Chapter 3). The complexity of the sample gave rise to the generation of a variety of situations with chromatograms involving multiple peak configurations. This allowed evaluating the performance of the peak prominence criterion under controlled situations. Similar conclusions could be achieved with any other complex sample, provided that full information is available from standards.

The search of optimal chromatographic conditions requires attending simultaneously to both resolution and analysis time, which are opposed to each other: high analysis times (which are undesirable) tend to provide better

resolution (which is favourable); conversely, short analysis times do not favour best resolution. This behaviour implies that instead of a single solution for an optimal separation, there is usually a collection of valid solutions; some of them are preferable because the analysis time is shorter and others because the resolution is higher. The solutions (linear gradients in this work) can be represented in a plot where the axes correspond to the opposite quality measurements to be enhanced: chromatographic resolution and analysis time. Such types of plots were suggested by Pareto et al. in the context of multi-objective optimisations [30]. The purpose of the plots is to reveal Pareto-optimal solutions. A solution is qualified as Pareto-optimal when a response cannot be improved without worsening another [31].

Figure 9.3a shows a Pareto optimality plot for the separation under gradient elution of a mixture containing derivatives of the 19 primary amino acids, assuming that the peak areas for all compounds are the same. In our example, each point in the plot corresponds to a particular linear gradient, which is characterised by its global peak purity –calculated as product of the individual peak purities– and analysis time. The assayed 1081 gradients assumed a gradient time of 60 min and variable initial and final concentrations of acetonitrile in the 5.0–27.5% range, which corresponds to the extreme concentrations in the isocratic experimental design used for modelling the chromatographic behaviour.
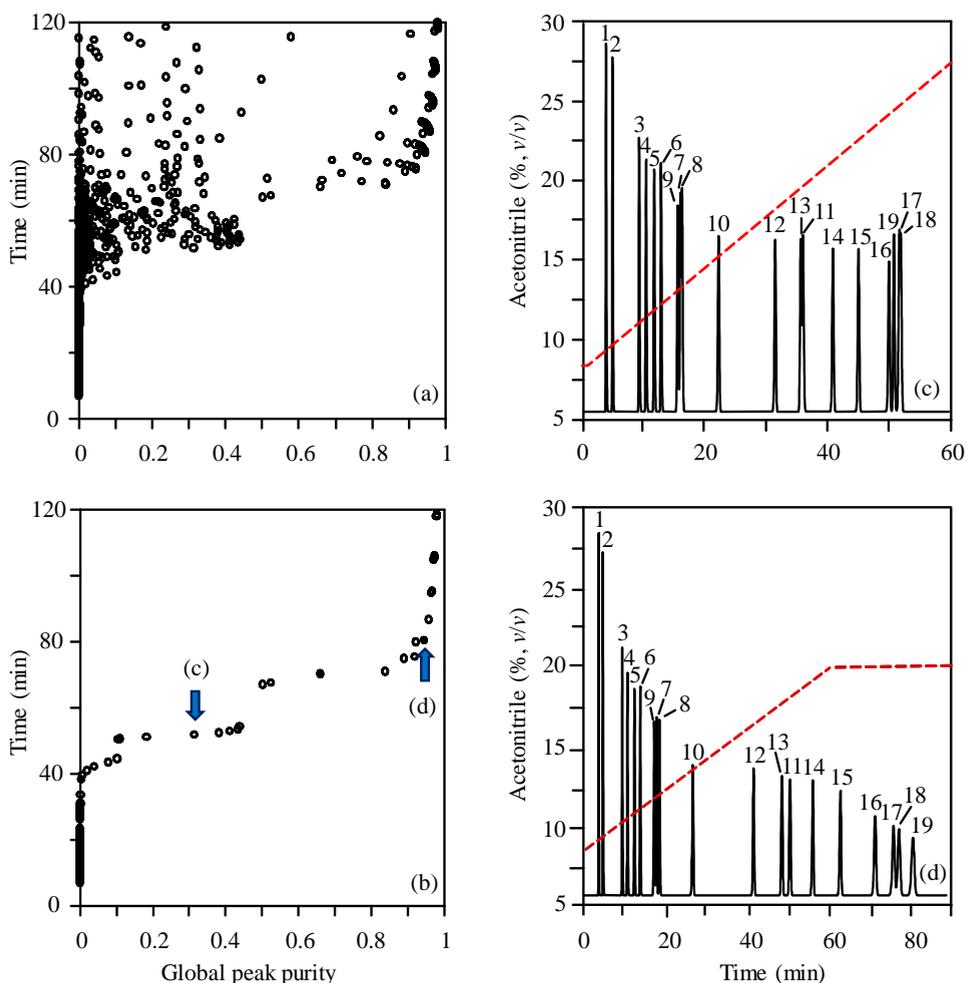
Figure 9.3. Performance of the separation of a sample containing the 19 primary amino acids, derivatised with OPA-NAC, using linear gradients in a selected experimental design: (a) Full Pareto optimality plot, (b) Pareto front, and (c) and (d) chromatograms corresponding to the situations marked with arrows on the Pareto front. The identities of amino acids are given in Section 9.4. The gradient program is represented on each chromatogram.

The border region in Figure 9.3a constitutes the Pareto front (Figure 9.3b), which includes the gradients fulfilling the optimality principle: those experimental conditions (in our case linear gradients), for which the resolution cannot be improved without increasing the analysis time, from which an analyst can make his/her selection. It can be observed that almost complete resolution (global peak purity > 0.95) is only possible at long analysis times. To appraise the separation performance, two points were marked with arrows on the Pareto front, corresponding to situations of incomplete and nearly complete resolution (global peak purity of 0.3 (point c) and above 0.9 (point d), respectively). The corresponding chromatograms are shown in Figures 9.3c and d, respectively.

In the horizontal intermediate region of the Pareto front, the best chromatograms keep an approximately constant analysis time, but very different resolution from low to high. This situation is usual in chromatography, and the solution selected as optimal is the last high resolution condition in the right side of the horizontal region (point (d) in Figure 9.3b). If the analyst tries to decrease the analysis time beyond point (d), the resolution decreases drastically and there is no longer a practical optimal solution.

9.6.2. *Peak prominence versus peak purity*

Peak prominence and peak purity criteria are based on different principles and have different scopes of application. Peak purity is calculated from simulated chromatograms obtained using information from chemical standards. It is possible to predict the individual contribution for each compound, and from this and the contribution from accompanying compounds, the associated resolution. By combining the peak purities for all compounds, a global measurement expressing the quality of the separation of the whole chromatogram is obtained. In this way, it is possible to predict how the

chromatograms would change under different experimental conditions (*in silico* simulation). In contrast, peak prominence is measured directly from experimental chromatograms (which is not possible with the peak purity). Since the aim of this study was the comparison of both functions (peak prominence *versus* peak purity), information obtained from standards (the set of amino acid derivatives) was used. In this study, when peak prominences were computed, the chromatogram of the mixture was straightforwardly processed as if we only had that chromatogram.

A common way to reduce the information from several peaks to a single value is the product of the resolutions of all peak pairs, which qualifies the global resolution reached in a chromatogram. Alternatively, the product of individual resolutions associated with all peaks can be used. This global criterion is used in Figure 9.3. However, only a comprehensive knowledge of the compounds eluting under an apparent peak allows a proper appraisal of the global resolution. When the peak purity criterion is applied, the number of expected compounds in the mixture is always known, and so also the number of underlying peaks (being visible or not). In contrast, when only the overall chromatogram of a sample is available, it is not possible to know with certainty whether a peak hides one or more underlying compounds. Consequently, it can be expected that the products of peak purities and peak prominences show strong differences for chromatograms presenting significant overlaps, where the number of visible peaks is smaller than the existing compounds.

9.6.3. Global resolution function based on peak prominence

As the product of elementary values is not a good choice to compare both resolution criteria, the sum of normalised elementary resolutions was used

instead. Similarly to the product, the higher the summation, the more are the resolved compounds. However, the product is extremely sensitive to peak overlapping, whereas the summation allows improvements of resolution in spite of existing overlapped peaks. Another advantage of the summation is that the maximal global value for a set of experimental conditions will give an indication of the number of compounds in the sample (i.e., the sample will contain that number –or more– compounds). The sum of resolutions has been used in the literature and participates, for instance, in the COFs reported by Duarte and Duarte [21] and Berridge [32].

For the peak prominence, three variants were considered for the global resolution function:

$$PR1_j = \sum_{i=1}^{nd} a_{\mathrm{pr},i,j} \tag{9.3}$$

$$PR2_j = \sum_{i=1}^{nd} \frac{a_{\mathrm{pr},i,j}}{a_{\mathrm{T},i,j}} \tag{9.4}$$

$$PR3_j = nd \times \frac{\displaystyle\sum_{i=1}^{nd} a_{\mathrm{pr},i,j}}{\displaystyle\sum_{i=1}^{nd} a_{\mathrm{T},i,j}} \tag{9.5}$$

nd being the number of detected peaks, $a_{\mathrm{pr},i,j}$ the area of the protruding part of peak i, and $a_{\mathrm{T},i,j}$ its total area (Figure 9.2), for a chromatogram obtained with gradient j. Equation (9.3) considers directly areas of the protruding part of the peaks, and makes sense when applied to a numerical optimisation based on simulated chromatograms, where all peaks have unit areas. Equations (9.4) and (9.5) are aimed to make the resolution function insensitive to differences among peak areas, and additionally achieve better equivalence between peak

prominence and peak purity in the selection of optimal gradients. Equation (9.4) considers normalised values referred to each individual peak, so that if the peaks are fully resolved, *PR*2 matches the number of eluted compounds. Equation (9.5) also provides a normalised measurement, but in this case referred to the whole chromatogram, so that if it is multiplied by the number of peaks, it gives values similar to *PR*2, but giving more importance to the resolution of major components.

For comparison purposes, the global peak purity was calculated as:

$$P_j = \sum_{i=1}^{n} p_{i,j} \qquad (9.6)$$

where $p_{i,j}$ is the elementary peak purity for compound i and gradient j, and n the number of eluted compounds.

9.6.4. *Comparison methodology to check the good performance of peak prominence*

In Section 9.6.3, three different definitions of global peak prominence (Equations (9.3) to (9.5)), and one definition for the peak purity (Equation (9.6)) were proposed. The plot in Figure 9.4 represents Pareto plots for the peak purity (left) and peak prominence criteria according to the three definitions (right). The most suitable definition for peak prominence will be that one for which the optimal selected gradients (those for the Pareto front) agree the best with those gradients selected by the peak purity under conditions of incomplete resolution. To facilitate a comparison between peak prominence and peak purity, a sort of "projection" of the Pareto fronts for one criterion was made on the Pareto plot for the other criterion, and vice versa.

Peaks with different area
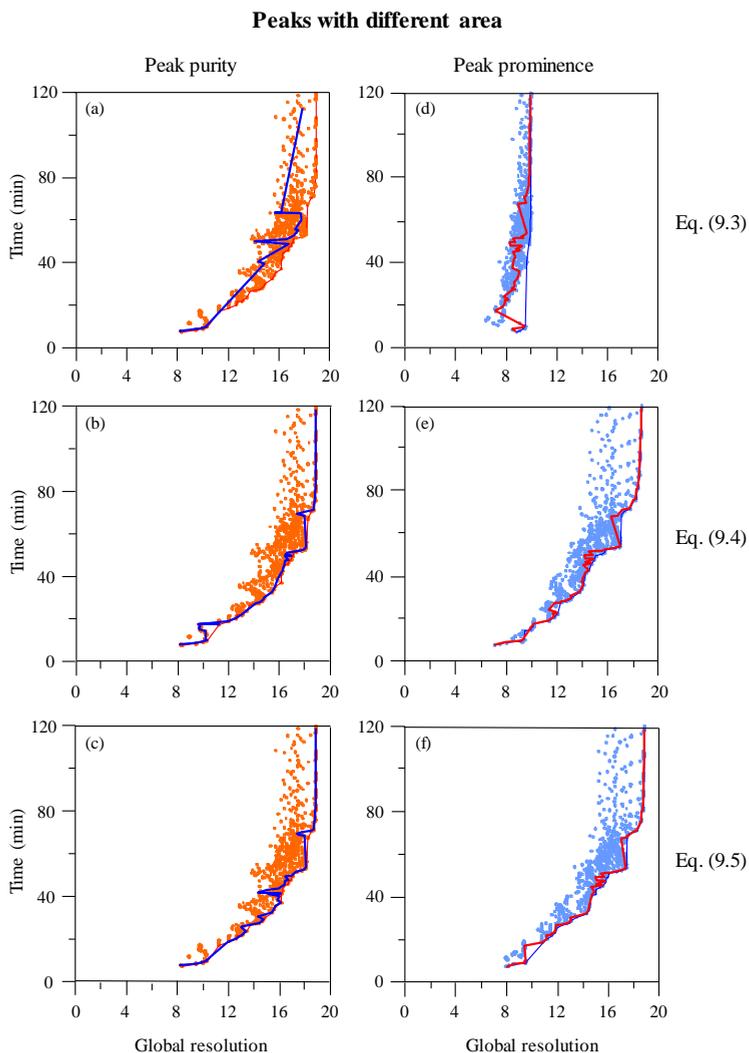
Figure 9.4. Pareto plots for chromatograms of samples containing the 19 amino acids with differentiated peak areas. Global resolution criteria: (a to c) sum of peak purities, (d) sum of areas of the protruding part of each peak, (e) sum of peak prominences, and (f) number of detected peaks multiplied by the ratio of the sum of the protruding parts of each peak and sum of total areas. The Pareto fronts for each criterion are shown as thin lines, and the projections of the gradients from one criterion to the other as thick lines.

For this purpose, the gradients selected as Pareto optimal in one criterion (drawn as a thin line) were searched on the Pareto plot for the other criterion, and the corresponding points were joined with a thick line. If the thin and thick lines agreed in the Pareto plots for both criteria (prominence and purity), the analyst would select the same gradients as optimal, or at least gradients performing similarly, using both criteria. In other words, a global function for peak prominence will be considered ideal if the "projection" of the gradients representing the performance of the Pareto front for the peak prominence matches with the Pareto front obtained for the peak purity.

The three left Pareto plots for peak purity in Figure 9.4 are exactly the same (same dots and thin line), but the projection of the Pareto front (thick line) for each definition of the peak prominence (Equations (9.3) to (9.5)) is different. Each plot in Figures 9.4, 9.5 and 9.6 shows what happens in different situations. The goal is to find what definition of peak prominence (which lacks information about the number of compounds under the global signal) selects the same optimal gradients as the criterion that has full knowledge about the real number of compounds in the sample.

It can be expected that in the region of the Pareto plot where the number of visible peaks coincides with the number of compounds, the Pareto front for one criterion and the projection of the other will agree, meaning that the same gradients are selected as optimal by both criteria. Therefore, the region where the resolution is poorer (the number of peaks is less than the number of compounds) is more meaningful for the comparison of both criteria.

Same area for all peaks
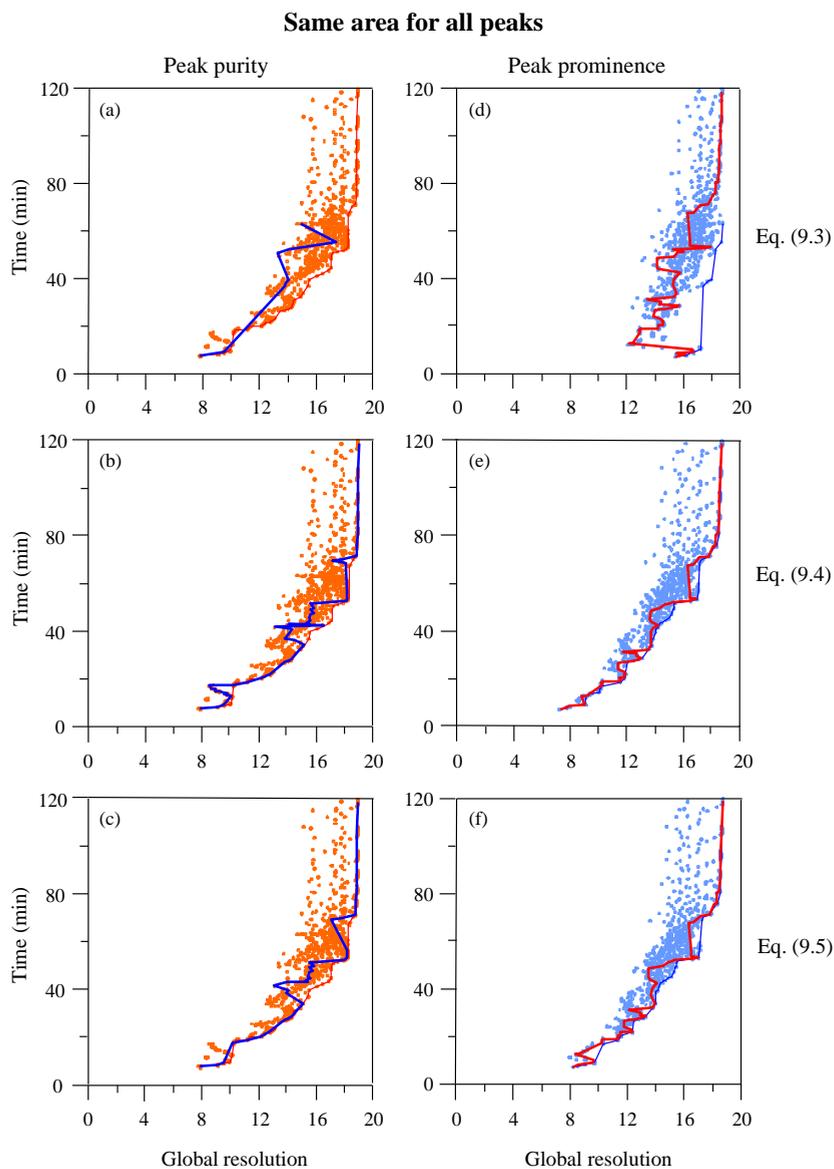
Figure 9.5. Pareto plots according to different global resolution criteria, corresponding to the separation of the OPA-NAC derivatives of the 19 amino acids assuming normalised peak areas. See Figure 9.4 and text for other details.
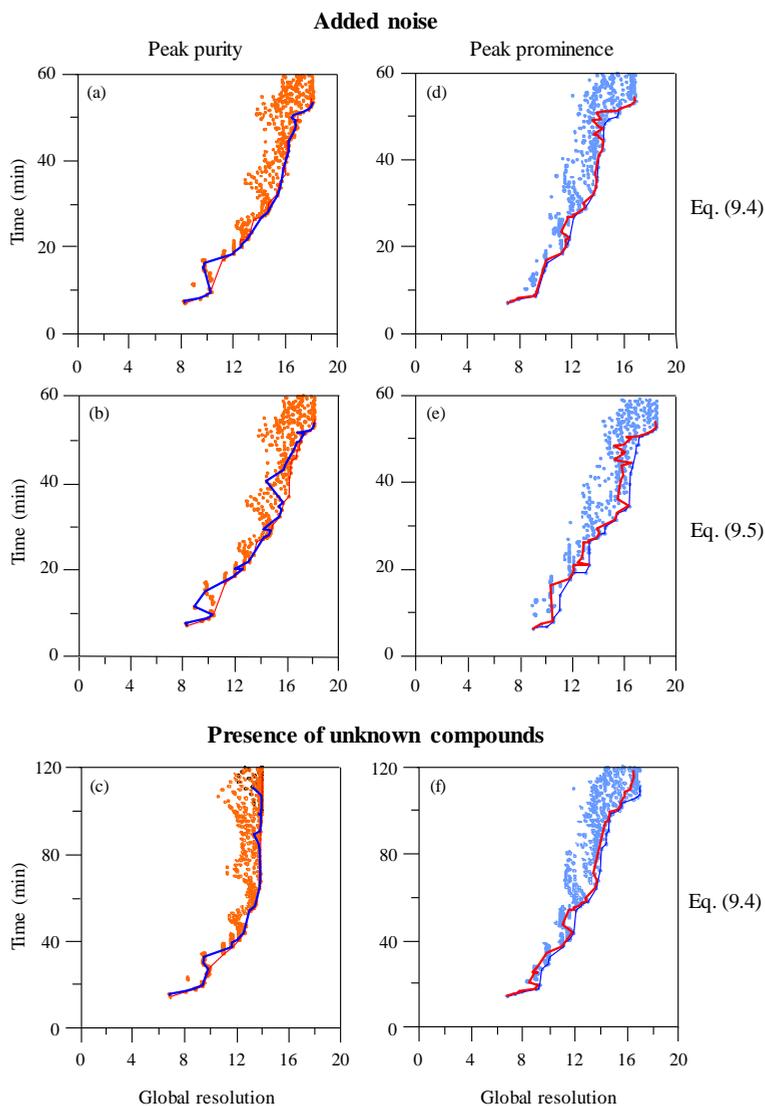
Figure 9.6. Pareto plots for chromatograms of samples containing the 19 amino acids with differentiated peak areas, added noise and unknown compounds: (a to c) sum of peak purities, (d,f) sum of peak prominences, and (e) number of detected peaks multiplied by the ratio of the sum of the protruding parts of each peak and sum of total areas. In (c,f), five compounds were considered as unknowns. See Figure 9.4 and text for other details.

9.6.5. Study of peak prominence performance

In this section, the most suitable definition of global peak prominence is investigated in situations progressively closer to reality. For this purpose, the set of 1081 predicted chromatograms for the mixture of the OPA-NAC derivatives of the 19 amino acids, obtained using linear gradients, were processed. Those gradients yielding analysis times exceeding 120 min were not considered in the Pareto plots.

9.6.5.1. Effect of peak area

In a first step of the study, an ideal situation implying noise-free predicted chromatograms was considered. Figure 9.4 shows the corresponding Pareto plots, where the areas of the involved peaks were different (a similar figure for chromatograms containing peaks with the same area, is given in Figure 9.5). Relative peak areas, in the range 0.04–0.95, were randomly selected for each amino acid. The same values were used for the 1081 assayed linear gradients and kept for the next studies. As shown in Figures 9.4 and 9.5, correspondence between the Pareto fronts and the "projections" for both resolution criteria is very satisfactory when Equations (9.4) and (9.5) were used, especially for the former, but very poor for Equation (9.3), in spite of involving normalised resolution measurements. Since the results obtained with Equation (9.3) were not acceptable, this global function was discarded for further studies.

9.6.5.2. Effect of noise

Figure 9.6 shows the Pareto plots obtained according to the peak purity criterion (Figures 9.6a and b) and peak prominence criteria (Figures 9.6d and e), for peaks of different size including significant noise, which was fixed for

all gradients in the experimental design. The signal-to-noise ratio was 6.45 for the smallest peak obtained with the slowest gradient, expressing the noise band as 2×1.96×0.015, where 0.015 was the standard deviation of the normal noise.

It can be observed again that Equation (9.4) gave rise to a better matching with peak purity compared to Equation (9.5). Therefore, Equation (9.4) seems to be the most suitable measurement of global resolution due to the high agreement between the Pareto front for the peak purity and the projection of optimal gradients for the peak prominence, and vice versa.

9.6.5.3. Presence of unknown compounds

Figures 9.6c and f constitutes a case of study even closer to reality since it includes, besides peaks of different size and noise, the presence of unknowns: compounds 2, 4, 7, 13, and 17 (see identities in Section 9.4) were randomly chosen among those with peaks of low magnitude, in order to simulate the presence of unknown impurities or matrix components. Interference of the five unknown compounds was taken into account for calculating the peak purity, but the global resolution was limited to the remaining 14 compounds. In contrast, the peak prominence straightforwardly attended to all visible peaks, independently of being analytes, impurities, or matrix components.

For the peak purity, as there were only 14 target compounds to be resolved (the remaining five were unknowns), the maximal sum of elementary peak purities tended to 14. Meanwhile, for peak prominences, the maximal number of visible peaks tended to 18. In spite that the measurements of global resolution are different, the Pareto fronts for both criteria agreed satisfactorily in the regions of incomplete resolution, where the differences between them are magnified. This means that the prominence criterion selects as optimal practically the same gradients as the peak purity, and consequently, it is

possible to carry out an optimisation using the peak prominence criterion with a guarantee of finding the same optimal conditions as the peak purity, independently of the resolution level.

9.6.6. Measurement of the mean resolution from Equation (9.4)

9.6.6.1. Chromatograms containing a limited number of peaks

From the above discussion (Section 9.6.5), the sum of peak prominences according to Equation (9.4) was concluded to be the best choice for monitoring the resolution in complex chromatograms. In order to convert this measurement to a mean resolution and make it independent of the number of peaks, the sum of peak prominences should be somehow normalised. For instance, it can be divided by the number of compounds in the sample (19 for the case of study involving the amino acid derivatives).

However, for many samples, the number of compounds is unknown, although still limited. This is the situation that happens when a non-excessively complex sample with similar concentrations for all solutes is analysed using a set of gradients, so that the elution conditions are forced to reduce the analysis time. One way of transforming the sum of prominences into a normalised resolution in such a situation is by dividing the sum of prominences by the maximal number of detected peaks, considering all assayed experimental conditions. In the separation example of the amino acid derivatives, taking into account different areas and the 1081 inspected gradients, the maximal observed sum of prominences was 18.85, which means that there should be at least 19 compounds in the sample. In this way, the maximal representative resolution would be $18.85/19 = 0.992$ (almost full resolution). Similarly, the respective

summations for all other gradients can be converted to a normalised resolution dividing by 19.

9.6.6.2. Chromatograms containing an undefined number of peaks

Up to now, we considered cases where the number of compounds present in a sample was known, or at least, limited. However, the ultimate goal of this research was to measure the resolution in chromatograms of real complex samples, where the identity of some or even all compounds is unknown, and the number of peaks is not well defined (i.e., the number varies largely depending on the experimental and detection conditions, and with sample concentration). This is the case of chromatographic fingerprints, such as those obtained from medicinal herbs [33–35]. Appraising properly the resolution level in such samples is particularly difficult, owing to the extreme disparity in signal size. Figure 9.7 depicts the fingerprints of extracts of decaffeinated and horsetail teas. In this type of sample, there is a high number of analytes in a wide range of concentrations.

For such samples, the selected global function for the peak prominence criterion (Equation (9.4)) requires additional adaptations. Without them, resolution values would depend on the number of considered peaks. It should be noted that, owing to undefinedness in the number of analyte peaks in the fingerprints, the sum of peak prominences does not provide unequivocal values, since it is conditioned by the number of terms included in the summation. Thus, if the detected peaks are prematurely cropped (i.e., the detection criterion only selects major peaks), significant peaks will be neglected, and oppositely, if the peaks are cropped too late, noise peaks will be included in the measurement. In contrast, a normalised resolution (if it could be calculated somehow) would be less prone to such variability.
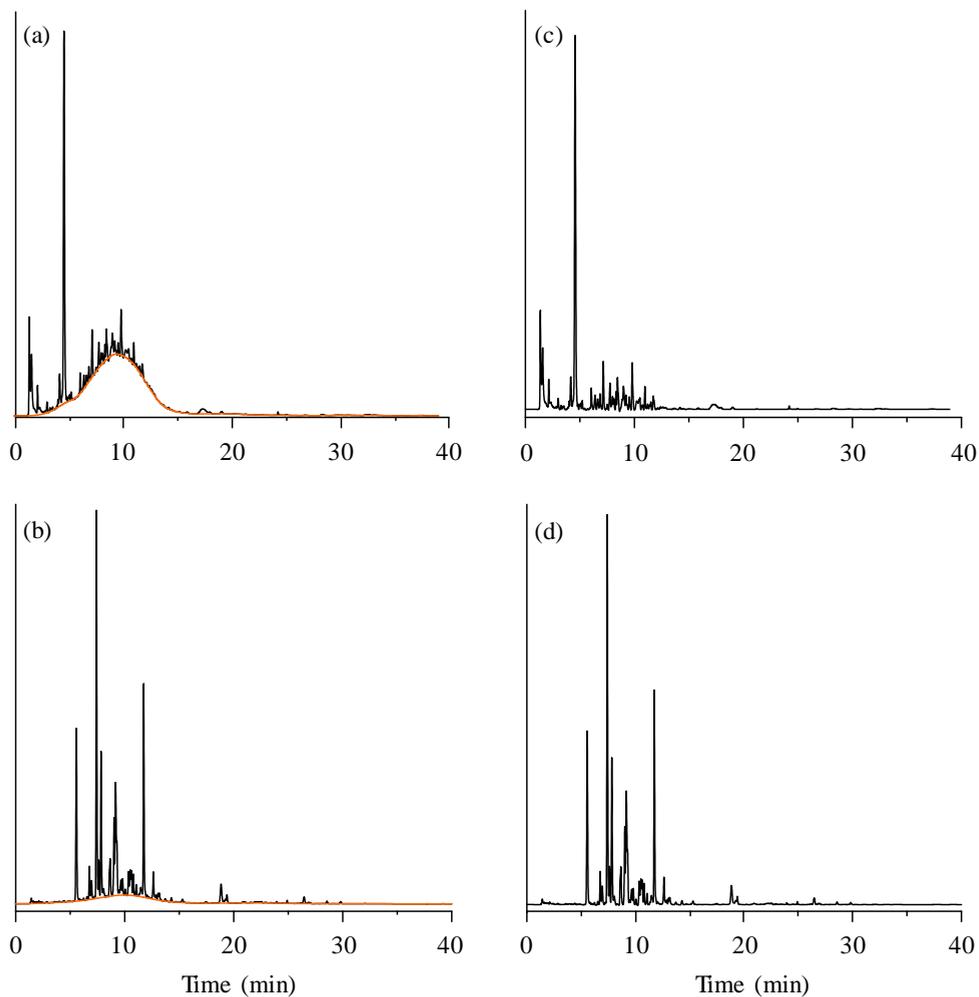
Figure 9.7. Chromatographic fingerprints for extracts of decaffeinated tea (a,c), and horsetail tea (b,d) obtained with a 20–60% (*v/v*) acetonitrile linear gradient using a gradient time of 10 min. (a,b) Raw chromatograms depicting the baseline found by the BEADS algorithm, and (c,d) chromatograms obtained after subtracting the baseline.

A solution to this issue could be computing the summation in Equation (9.4), after sorting the peak prominences according to a second property, such as the respective peak area, or alternatively, the area of the protruding part of the peaks. The operation of sorting the results allows introducing a secondary factor (namely, signal size), which does not participate in the calculation of peak prominence, but whose importance in estimating the resolution is decisive when the number of peaks is not well defined, and the peaks span a wide range of magnitude. Thus, if the sorting is carried out by decreasing peak areas, the influence of residual peaks tends to be neglected. If the sorted summation is done according to the protruding part of the peaks, less visible peaks will have smaller influence. Both options are valid and the results similar.

Raw chromatograms for the extracts of two tea samples analysed in our laboratory are shown in Figure 9.7 (left). As observed, both chromatograms present humps, which are constituted by the accumulation of tens of thousands of unresolved co-eluting compounds, as described by Kuhnert et al. [36]. The presence of humps is a very common phenomenon in samples of natural products and cannot be left aside. Depending on the way these humps are processed, the conclusions can be very different. Processing a chromatogram whose baseline preserves the hump will attend less to the protruding parts of the peaks over the hump, whereas processing a chromatogram corrected with an apparent baseline that cancels the hump would give more importance to those peaks protruding over the hump. Thus, an optimisation based on hump-corrected chromatograms will tend to magnify the visibility of the peaks over the hump. Obtaining a maximal amount of the protruding peaks (over the hump) is clearly the desirable situation. Therefore, the apparent baseline should be subtracted prior to any measurement of resolution.

Peak	Start	Max	End	upperArea	TotalArea	%prominence
47	7.25	7.37	7.50	44.1947	46.0797	95.9092
70	11.59	11.70	11.92	25.0524	25.8730	96.8282
38	5.32	5.52	5.73	23.6213	24.0047	98.4029
55	9.05	9.12	9.21	8.1423	18.0342	45.1493
49	7.68	7.79	7.94	16.4605	17.6514	93.2532
54	8.87	8.98	9.04	5.5213	9.6100	57.4541
53	8.47	8.62	8.81	6.9138	7.0641	97.8724
56	9.21	9.23	9.31	1.0225	5.1679	19.7859
48	7.50	7.57	7.68	3.3904	4.8106	70.4768
62	10.37	10.47	10.53	1.5123	4.3965	34.3977
94	18.59	18.80	19.02	3.7953	4.3812	86.6265
74	12.41	12.59	12.76	3.4190	4.0301	84.8367
44	6.62	6.71	6.82	3.2870	3.7736	87.1047
63	10.53	10.59	10.67	1.2482	2.9714	42.0082
61	10.21	10.30	10.37	1.7257	2.7107	63.6617
45	6.82	6.91	7.08	1.9356	2.5553	75.7484
58	9.49	9.61	9.69	1.6198	2.5467	63.6027

(........)

Figure 9.8. Screenshot for horsetail tea.

For this purpose, we developed a modification [37] of the BEADS algorithm [38], which assists in its application to real samples. Figure 9.7 shows the corrected chromatograms on the right. Corresponding screenshots of the chromatograms processed with the developed MATLAB function are shown in Figures 9.8 and 9.9. The peaks are numbered according to their elution order, and only those that exceed a relative peak area of 0.05% are shown. Blue tangents define the optimal protruding region for each peak, which is marked in red.

The inserted tables in each chromatogram show some of the parameters that the developed application provides: peak index, start, maximum and end peak

times, area of the protruding part of the peaks ($a_{pr,i}$), total area ($a_{pr,i} + a_{l,i}$) (see Figure 9.2), and prominence (Equation (9.4)) expressed as percentage.

Peak	Start	Max	End	upperArea	TotalArea	%prominence
17	4.27	4.53	4.75	214.6562	225.7066	95.1041
2	1.26	1.33	1.45	27.1850	39.1718	69.3993
3	1.45	1.55	1.71	18.9650	38.4690	49.2995
45	9.69	9.79	9.95	23.5967	24.9107	94.7254
31	7.01	7.13	7.32	21.7505	22.7277	95.7005
16	4.04	4.13	4.27	12.9503	18.1462	71.3663
39	8.37	8.43	8.59	14.7884	16.6141	89.0113
41	8.90	8.99	9.13	8.0888	15.1822	53.2783
75	16.79	17.32	17.81	11.1705	14.6291	76.3581
57	11.61	11.73	11.99	8.7822	12.6879	69.2168
34	7.62	7.75	7.84	10.8195	12.3899	87.3247
52	10.87	10.96	11.06	7.9768	9.9098	80.4949
4	2.04	2.11	2.19	6.7948	9.4053	72.2452
30	6.74	6.83	7.00	6.1522	7.9577	77.3108
27	5.95	6.03	6.17	7.4785	7.9066	94.5856
28	6.20	6.37	6.49	7.3242	7.4541	98.2563
44	9.43	9.52	9.66	5.8388	6.3465	91.9992

(........)

Figure 9.9. Screenshot for decaffeinated tea.

In Figure 9.10, mean resolution values for the two corrected fingerprints are plotted, after sorting the sum of peak prominences according to the respective peak areas. As can be seen, the horsetail extract not only has a smaller drop in resolution when new peaks are added, but also includes more detected peaks.

In order to make resolution values for intrinsically different samples comparable, a common reference is needed. In previous work [23], we observed that sorted relative areas of the protruding parts of the peaks of replicates of fingerprints diverged beyond 99.95% of the cumulative sum. The same is valid for the total peak areas. Thus, for the measurement of global resolution, a logical choice consists in summing the sorted peak prominences

according to Equation (9.4) up to this level, which would include most minor peaks in the computation of the resolution. However, this threshold can be decreased or increased according to user needs and the sample features.
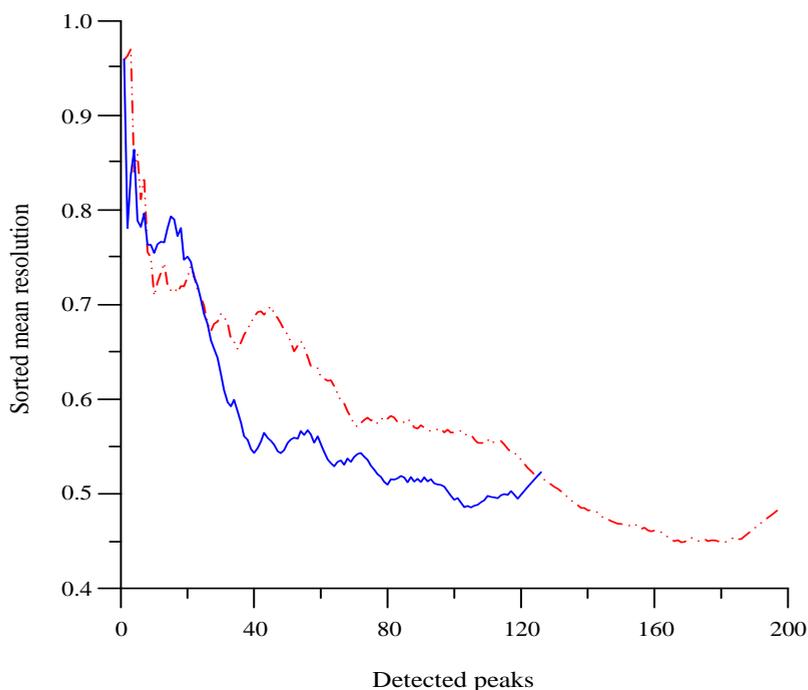
Figure 9.10. Sorted mean resolution plot for the fingerprints of the extracts of medicinal herbs: decaffeinated tea (continuous line), and horsetail tea (dotted dashed line). See text for meaning.

For the horsetail and decaffeinated tea (Figure 9.7), the number of peaks included at a threshold of 99.95% total peak area was 120 and 102, respectively. The sum of peak prominences sorted according to the areas was 64.33 and 50.05, and therefore, the mean global peak prominence was 64.33/120 = 0.536, and 50.05/102 = 0.491 for the two samples. Note, however,

that if our objective is selecting the chromatogram giving more resolved peaks, the sum of peak prominences at a given threshold is the best measurement.

9.7. Conclusions

This work demonstrates that experimental conditions selected as the best by the global peak prominence agrees with those chosen by the global peak purity. Peak purity is a function that has shown excellent features for measuring the resolution level in a chromatogram. This criterion is able to find the best separation conditions, even when complete resolution is not achieved, as is the case of the OPA-NAC derivatives of the amino acids found in proteins at short analysis times. However, it requires comprehensive information on the individual signals for each chromatogram to be computed. Peak prominence is based on very different principles, but shares some of the best features of the peak purity criterion, with the additional advantage of being evaluable directly from experimental chromatograms of the sample, without going through steps of modelling, prediction and simulation based on the information obtained from standards, as required for the peak purity.

The representation of Pareto plots allowed a thorough inspection of a large number of separation conditions, and the evaluation of several global resolution functions versus analysis time. After discarding the product of peak prominences as a resolution measurement, three functions represented by Equations (9.3) to (9.5) were compared with the sum of peak purities. The mutual projections of Pareto plots in cases of increased realism allowed a pairwise comparison of resolution functions, a strategy that can be useful in evaluating the performance of other COFs.

Our comparison study showed that Equation (9.4) was the best definition of global peak prominence, since the same gradients were selected as optimal by both criteria (peak prominence and peak purity), which is particularly noteworthy in the low resolution region. This agreement allows having a valid resolution criterion in situations where it is not possible to evaluate the peak purity. Some of such situations include measuring the quality of a separation in a trial and error optimisation, or when a sample contains unknown matrix components, or when there are no standards available for some (or all) analytes. The studies carried out in this work indicate that a resolution function based on the measurement of peak prominences can be applied to any type of sample, in the presence and absence of standards, with chromatograms including peaks of similar size or with very different magnitude, in the presence or absence of noise, in the presence of unknown matrix components, and even when the number of constituents is not well defined (e.g., natural products with extreme disparity in the concentration of components).

This work is dedicated to developing and validating a global resolution function that can be used for screening studies and optimisation in a design of experiments (DOE) framework. The validation was outlined through optimisation studies, by checking whether experimental conditions selected as optimal by the peak prominence agree with those selected by the peak purity, in spite of lacking information from underlying peaks. The results suggest that the sum of peak prominences is a good choice for optimising fingerprint chromatograms. The practical use of the developed function for optimisation purposes with fingerprints obtained by gradient elution is currently being developed in our laboratory, and it will be the subject of a future report.

9.8. References

[1] M.C. García Álvarez-Coque, J.J. Baeza Baeza, G. Ramis Ramos, Reversed phase liquid chromatography, in *Analytical Separation Science Series*, Vol. 1 (edited by J.L. Anderson, A. Stalcup, A. Berthod, V. Pino), Wiley, New York, 2015.

[2] R. Oertel, J. Pietsch, N. Arenz, S.G. Zeitz, L. Goltz, W. Kirch, Simultaneous determination of drugs in human autopsy material using phase-optimized liquid chromatography, *Biomed. Chromatogr.* 26 (2012) 1608–1616.

[3] K.R. Chalcraft, B.E. McCarry, Tandem LC columns for the simultaneous retention of polar and nonpolar molecules in comprehensive metabolomics analysis, *J. Sep. Sci.* 36 (2013) 1379–1388.

[4] R.J. Vonk, A.F.G. Gargano, E. Davydova, H.L. Dekker, S. Eeltink, L.J. de Koning, P.J. Schoenmakers, Comprehensive two-dimensional liquid chromatography with stationary-phase-assisted modulation coupled to high-resolution mass spectrometry applied to proteome analysis of *Saccharomyces cerevisiae*, *Anal. Chem.* 87 (2015) 5387–5394.

[5] P.J. Schoenmakers, *Optimisation of Chromatographic Selectivity: A Guide to Method Development*, Elsevier, Amsterdam, The Netherlands, 1986.

[6] M.C. García Álvarez-Coque, J.R. Torres Lapasió, J.J. Baeza Baeza, Models and objective functions for the optimisation of selectivity in reversed-phase liquid chromatography, *Anal. Chim. Acta* 579 (2006) 125–145.

[7] R. Cela, E.Y. Ordoñez, J.B. Quintana, R. Rodil, Chemometric-assisted method development in RPLC, *J. Chromatogr. A* 1287 (2013) 2–22.

[8] R. Cela, C.G. Barroso, J.A. Pérez Bustamante, Objective functions in experimental and simulated chromatographic optimization: Comparative study and alternative proposals, *J. Chromatogr.* 485 (1989) 477–500.

[9] J.P. Foley, Resolution equations for column chromatography, *Analyst* 116 (1991) 1275–1279.

[10] M.R. Schure, Quantification of resolution for two-dimensional separations, *J. Microcol. Sep.* 9 (1997) 169–176.

[11] S. Carda Broch, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Evaluation of several global resolution functions for liquid chromatography, *Anal. Chim. Acta* 396 (1999) 61–74.

[12] A.M. Siouffi, R. Phan-Tan-Luu, Optimization methods in chromatography and capillary electrophoresis, *J. Chromatogr. A* 892 (2000) 75–106.

[13] E.J. Klein, S.L. Rivera, A review of criteria functions and response surface methodology for the optimization of analytical scale HPLC separations, *J. Liq. Chromatogr. Relat. Technol.* 23 (2000) 2097–2121.

[14] S. Peters, G. Vivó Truyols, P.J. Marriott, P.J. Schoenmakers, Development of a resolution metric for comprehensive two-dimensional chromatography, *J. Chromatogr. A* 1146 (2007) 232–241.

[15] A. Ortín, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Finding the best separation in situations of extremely low chromatographic resolution, *J. Chromatogr. A* 1218 (2011) 2240–2251.

[16] R.M.B.O. Duarte, J.T.V. Matos, A.C. Duarte, A new chromatographic response function for assessing the separation quality in comprehensive two-dimensional liquid chromatography, *J. Chromatogr. A* 1225 (2012) 121–131.

[17] J.T.V. Matos, R.M.B.O. Duarte, A.C. Duarte, A generalization of a chromatographic response function for application in non-target one- and two-dimensional chromatography of complex samples, *J. Chromatogr. A* 1263 (2012) 141–150.

[18] J.T.V. Matos, R.M.B.O. Duarte, A.C. Duarte, Chromatographic response functions in 1D and 2D chromatography as tools for assessing chemical complexity, *Trends Anal. Chem.* 45 (2013) 14–23.

[19] E. Tyteca, G.A. Desmet, Universal comparison study of chromatographic response functions, *J. Chromatogr. A* 1361 (2014) 178–190.

[20] J.R. Torres Lapasió, M.C. García Álvarez-Coque, E. Bosch, M. Rosés, Considerations on the modelling and optimisation of resolution of ionisable compounds in extended pH-range columns, *J. Chromatogr. A* 1089 (2005) 170–186.

[21] R.M.B.O. Duarte, A.C. Duarte, A new chromatographic response function for use in size-exclusion chromatography optimization strategies: Application to complex organic mixtures, *J. Chromatogr. A* 1217 (2010) 7556–7563.

[22] T. Álvarez Segura, A. Gómez Díaz, C. Ortiz Bolsico, J.R. Torres Lapasió, M.C. García Álvarez-Coque, A chromatographic objective function to characterise chromatograms with unknown compounds or without standards available, *J. Chromatogr. A* 1409 (2015) 79–88.

[23] T. Álvarez Segura, E. Cabo Calvet, J.R. Torres Lapasió, M.C. García Álvarez-Coque, An approach to evaluate the information in chromatographic fingerprints: Application to the optimisation of the extraction and conservation conditions of medicinal herbs, *J. Chromatogr. A* 1422 (2015) 178–185.

[24] C. Ortiz Bolsico, J.R. Torres Lapasió, M.C. García Alvarez-Coque, Approaches to find complementary separation conditions for resolving complex mixtures by high performance liquid chromatography, *J. Chromatogr. A* 1229 (2012) 180–189.

[25] R.D. Caballero, S.J. López Grío, J.R. Torres Lapasió, M.C. García Alvarez-Coque, Single-peak resolution criteria for optimization of mobile phase composition in liquid chromatography, *J. Liq. Chromatogr. Rel. Technol.* 24 (2001) 1895–1919.

[26] V. Concha Herrera, G. Vivó Truyols, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Limits of multi-linear gradient optimisation in reversed-phase liquid chromatography, *J. Chromatogr. A* 1063 (2005) 79–88.

[27] M. Dumarey, I. Smets, Y. Vander Heyden, Prediction and interpretation of the antioxidant capacity of green tea from dissimilar chromatographic fingerprints, *J. Chromatogr. B* 878 (2010) 2733–2740.

[28] C. Ortiz Bolsico, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Optimization of gradient elution with serially-coupled columns. Part I: Single linear gradients, *J. Chromatogr. A* 1350 (2014) 51–60.

[29] C. Ortiz Bolsico, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Optimisation of gradient elution with serially-coupled columns. Part II: Multi-linear gradients, *J. Chromatogr. A* 1373 (2014) 51–60.

[30] A.K. Smilde, A. Knevelman, P.M.J. Coenegracht, Introduction of multi-criteria decision making in optimization procedures for high-performance liquid chromatographic separations, *J. Chromatogr.* 369 (1986) 1–10.

[31] V. Pareto, *Manual of Political Economy*, Augustus M. Kelley Publishers, New York, USA, 1971.

[32] J.C. Berridge, Unattended optimisation of reversed-phase high-performance liquid chromatographic separations using the modified simplex, *J. Chromatogr.* 244 (1982) 1–14.

[33] B.Y. Li, Y. Hu, Y.Z. Liang, P.S. Xie, Y.P. Du, Quality evaluation of fingerprints of herbal medicine with chromatographic data, *Anal. Chim. Acta* 514 (2004) 69–77.

[34] P. Xie, S. Chen, Y.Z. Liang, X. Wang, R. Tian, R. Upton, Chromatographic fingerprint analysis: A rational approach for quality assessment of traditional Chinese herbal medicine, *J. Chromatogr. A* 1112 (2006) 171–180.

[35] X.M. Liang, Y. Jin, Y.P. Wang, G.W. Jin, Q. Fu, Y.S. Xiao, Qualitative and quantitative analysis in quality control of traditional Chinese medicines, *J. Chromatogr. A* 1216 (2009) 2033–2044.

[36] N. Kuhnert, F. Dairpoosh, G. Yassin, A. Golon, R. Jaiswal, What is under the hump? Mass spectrometry based analysis of complex mixtures in processed food: Lessons from the characterisation of black tea thearubigins, coffee melanoidines and caramel, *Food Funct.* 4 (2013) 1130–1147.

[37] J.A. Navarro Huerta, J.R. Torres Lapasió, S. López Ureña, M.C. García Álvarez-Coque, Assisted baseline subtraction in complex chromatograms using the BEADS algorithm, *J. Chromatogr. A* 1507 (2017) 1–10.

[38] X. Ning, I.W. Selesnick, L. Duval, Chromatogram baseline estimation and denoising using sparsity (BEADS), *Chemom. Intell. Lab. Syst.* 139 (2014) 156–167.

CHAPTER 10

CLASSIFICATION OF OLIVE LEAF AND PULP EXTRACTS BY COMPREHENSIVE TWO-DIMENSIONAL LIQUID CHROMATOGRAPHY OF POLYPHENOLIC FINGERPRINTS

10.1. Abstract

The development of a new comprehensive two-dimensional liquid chromatographic method is described, to obtain the profiles of polyphenolic compounds present in olive (*Olea europaea* L.) leaves and pulps from different genetic origin. Optimisation of the stationary phase nature, particle size, column length and internal diameter, as well as other separation conditions, was performed. Along the study, three stationary phases (C18, PFP and phenyl) in the first dimension (^{1}D), and five (C18, amide, cyano, phenyl and PFP) in the second dimension (^{2}D) were combined to obtain the maximal number of resolved peaks. The optimised method successfully characterised the presence of 26 and 29 common polyphenols in olive leaf and pulp extracts, respectively. Peak volume ratios were used to develop linear discriminant analysis models able to distinguish olive leaf and pulp extracts among seven cultivars from several Spanish regions. The results demonstrated that polyphenolic profiles were characteristic of each cultivar.

10.2. Introduction

Olive trees (*Olea europeaea* L.) are cultivated throughout the Mediterranean area. Olive leaf and pulp extracts are complex mixtures containing hundreds of different compounds, and their composition can change as a function of the cultivar [1,2], geographic origin [3], and maturity index [4]. Consequently, rapid and reliable methods for guaranteeing the quality and origin of these products are highly demanded [5]. These matrices contain a high number of polyphenolic components, which are highly appreciated by consumers due to their contribution to the nutritional, sensorial, and commercial characteristics of food. These compounds can be grouped in phenolic acids, flavan-3-ols, flavanones, flavones, flavonoids and lignans [6,7].

One-dimensional liquid chromatography (1D-LC) is commonly used to obtain the profile of polyphenols in olive leaves and fruits [6,8–10]. Mass spectrometry is frequently coupled to 1D-LC to reach enough sensitivity, and detect as many compounds as possible [3,11]. Some authors have described coupling to diode array detection (DAD) to carry out the analysis of polyphenols in extra virgin olive oil [12], olive leaves [10], and pulp of olive fruit [9].

Two-dimensional LC (2D-LC) in the comprehensive mode (LC×LC), coupled to DAD, has also been successfully applied to the analysis of polyphenols in food samples [6,13–17]. In this technique, discrete fractions of the first dimension ([1]D effluent) are collected during a short period of time, and transferred one at a time to a column in the second dimension ([2]D), where the selectivity of the system must be different to obtain good orthogonality [18,19]. The [2]D chromatograms from such experiments can reveal peak patterns similar for groups of specific analyte functionalities. Several types of columns have been used to perform the analysis of polyphenols in diverse types of samples.

Kivilompolo and Hyötyläinen [13] reported an LC×LC method to analyse polyphenols in *Lamiaceae* herbs by coupling C18 and cyano columns, obtaining excellent peak capacity. An amide column as [1]D coupled to a C18 column as [2]D resolved critical polyphenol pairs in a mate sample [20]. Hydrophilic interaction liquid chromatography (HILIC) combined with RPLC has been also used for the characterisation of polyphenols in apples [21], licorice [22], green cocoa beans [14], and grapevine [15].

In this work, an LC×LC method with UV detection is applied to the profiling of the polyphenolic fraction of olive leaves and pulps from several genetic origin. For this purpose, a detailed study of the best column combination in the 2D-LC instrument was carried out at different experimental conditions. Linear discriminant analysis (LDA) was successfully conducted using LC×LC peak volume ratios as predictors, to classify the matrices (olive leaves and pulps), according to their cultivar. To our knowledge, these diagnostic fingerprints for such samples have not previously been described for pattern recognition with discrimination purposes, with the purpose of cultivar discrimination.

10.3. Experimental

10.3.1. Reagents and samples

The following reagents were used: acetonitrile (ACN), methanol (MeOH), hydrochloric acid (HCl), trifluoroacetic acid (TFA), and 3,5-di-tert-butyl-4-hydroxy-toluene (BHT), all from Sigma-Aldrich (St. Louis, MO, USA). Deionised water was obtained with a B30 water purification system (Adrona, Riga, Latvia).

Table 10.1. Cultivar, geographical origin (Spain), and crop year of the olive leaves and pulps included in this study.

Cultivar	Geographical origin	Crop year[a]
Arbequina	Altura (Castellón)	2018
	Torres Segre (Lérida)	2016
Blanqueta	Muro de Alcoi (Alicante)	2018
	Pobla del Duc (Valencia)	2016
Cornicabra	Altura (Castellón)	2018
	Daimiel (Ciudad Real)	2016
Hojiblanca	Requena (Valencia)	2018
	Antequera (Málaga)	2015
Picual	Puente Genil (Córdoba)	2018
	Jumilla (Murcia)	2016
Serrana	Altura (Castellón)	2018
		2016
Villalonga	Muro de Alcoi (Alicante)	2018
		2016

[a] Three replicated extractions from samples collected in the indicated years were processed. In all cases, the crop month was November.

The olive leaves and fruits analysed in this study (Table 10.1) were kindly donated by different olive oil manufacturers. Correct sampling was assured, in both olive leaves and fruits, by collecting them directly from trees located in different Spanish regions in the same period (usually end of November 2016 and November 2018). The cultivar of samples was guaranteed by the olive oil

producer companies. Both olive leaves and fruits were randomly selected, washed with water to remove dust and airborne particles settled on the olive, and then stored at –20 ºC up to their use.

10.3.2. Preparation of polyphenolic extracts

The polyphenol extraction procedure, selected to get the extracts to be injected into the chromatographic system, was adapted from Jerman et al. [23] and Martí et al. [24]. This procedure offers an adequate extraction of polyphenols from complex tissues without any degradation, or chemical modification, as demonstrated in the literature [23].

Briefly, for olive leaves, 1 g of sample was weighted and 15 mL of a 40% (*v/v*) MeOH aqueous solution, containing 0.1% BHT (*w/v*) to avoid the oxidation of polyphenols, was added for the extraction. For olive pulps, 1.5 g of sample was weighted and 25 mL of pure MeOH was added instead. The achieved mixtures were sonicated in an ultrasonic bath at 45 ºC for 1 h (S15H Elma Electronics AG, Wetzikon, Switzerland) at a frequency of 37 kHz. Then, the extracts were centrifuged at 5000 rpm during 10 min (EBA 20, Hettich, Tuttlingen, Germany). In the case of olive leaves, the supernatant was filtered through a 0.22 μm Nylon syringe filter (Análisis Vínicos, Tomelloso, Spain), and then injected into the chromatographic system. For olive pulp samples, the supernatant was 4-fold pre-concentrated, using a miVac sample concentrator (Genevac™, Ipswich, UK), and filtered through 0.22 μm Nylon syringe filter prior to chromatographic analysis.

10.3.3. LC×LC instrument

An 8-port/2-position switching valve (1290 series, Agilent Technologies, Waldbronn, Germany) was used to interface both dimensions in the LC×LC instrument, by collecting fractions from ^{1}D in two identical 40 µL-sampling loops, and transferring them to ^{2}D. The first dimension consisted of a 1260 series HPLC system (Agilent), equipped with autosampler, binary pump, online vacuum degasser, thermostated column compartment, and DAD. The column in ^{1}D was connected to the entrance of the switching valve, which allowed the injection of sample fractions into the second dimension (^{2}D), composed of an Agilent 1290 series HPLC system, equipped with binary pump, degasser, thermostated column compartment, and DAD.

10.3.4. Comprehensive LC×LC conditions

In the LC×LC optimised method, the polyphenolic compounds were separated in ^{1}D using a pentafluorophenyl (PFP) Kinetex F5 column (50 mm × 2.1 mm i.d., 2.6 µm particle size, Phenomenex, Torrance, CA, USA). Gradients were obtained by mixing water (solvent A) and MeOH (solvent B), each of them containing 0.05% (*v/v*) TFA. The optimised ^{1}D gradient was conducted by varying the proportion of MeOH as follows: 30% (*v/v*) MeOH during 10 min, followed by an increase of MeOH up to a ratio of 60% within 15 min, and a final increase to 95% within additional 15 min, which was kept during 2 min. A re-equilibration step to reach the initial conditions was carried out during 14 min. The injection volume was 2 µL, column temperature was fixed at 40 ºC, and the flow rate was set at 0.1 mL/min.

For the optimised ^{2}D, a Zorbax Eclipse Plus C18 column (50 mm × 3 mm i.d., 1.8 µm, Agilent), and gradients of water (solvent A) and ACN (solvent B),

containing both 0.05% TFA were used. The elution was performed using a shifted gradient (see Section 10.4.1.2); column temperature was 50 ºC, and the flow rate was set at 2.5 mL/min. The valve was switched automatically after each modulation cycle of 0.25 min. UV detection was performed at 280 ± 4 nm and 80 Hz for both dimensions.

The optimal conditions were obtained after testing several columns in both dimensions. In ^{1}D: Zorbax SB C18 (100 mm × 2.1 mm, 1.8 μm, Agilent) and ACE 5 Phenyl (75 mm × 4.6 mm, 5 μm, Aberdeen, Scotland, UK). In ^{2}D: ACE 5 CN (75 mm × 4.6 mm, 5 μm), ACE 5 Phenyl (75 mm × 4.6 mm, 5 μm), ACE 5 C18-PFP (75 mm × 4.6 mm, 5 μm), Zorbax Bonus RP (amide, 50 mm × 2.1 mm, 1.8 μm, Agilent), Kinetex F5 (PFP, 50 mm × 2.1 mm, 2.6 μm, Phenomenex), and Zorbax Eclipse Plus C18 column (50 mm × 3 mm, 1.8 μm, Agilent).

10.3.5. Acquisition of raw LC×LC data and statistical analysis

Data were acquired by an Agilent MSD ChemStation (C.01.07 SR2), and processed using the GC Image LC×LC software (version 2.4, GC Image, LLC, Lincoln, NE, USA). Peak detection was performed for each sample to select the 2D-LC common peaks. For this purpose, each 2D-LC data file (chromatogram) was previously baseline corrected using the GC Image software. The peak volume ratios were used to construct LDA models, able to distinguish samples of different genetic origin.

LDA analysis was carried out by using the Statistical Package for the Social Sciences (SPSS, version 19.0, Chicago, IL, USA). This is a supervised classificatory technique, which is considered an outstanding tool to obtain vectors at the maximal distance between a set of previously defined categories.

Up to $N-1$ discriminant vectors are created, N being the lowest value between the number of predictors and categories.

LDA models were constructed by using a stepwise algorithm to select the predictors. The Wilks' lambda λ_w was used as selection criterion. Values of λ_w close to 0 correspond to well resolved categories, while for overlapped categories λ_w approaches 1. According to the stepwise algorithm, a predictor is included in the model if the λ_w value after its inclusion does not exceed a pre-selected value, F_{in} (the entrance threshold of a test of comparison of variances, F_{-test}). The inclusion of a new predictor modifies the significance of those predictors already present in the model. After the inclusion of a new predictor, a rejection threshold, F_{out}, is used to decide if other predictor(s) should be removed from the model, without modifying the variance using pre-selected F_{in} and F_{out} values. The process ends when no more predictors enter or are eliminated from the model. The probability values $F_{in} = 0.001$, and $F_{out} = 0.10$ were adopted.

10.4. Results and discussion

10.4.1. Selection of the separation conditions

An olive leaf extract from the Serrana cultivar was analysed for the optimisation of the separation conditions of the polyphenolic compounds, in both 1D-LC and 2D-LC.

10.4.1.1. ^{1}D separation conditions

Three stationary phases were evaluated for the ^{1}D separation: C18, phenyl and PFP (see Section 10.3.4). C18 columns have been extensively used in the analysis of polyphenolic compounds in different food and vegetal samples

[3,11–13], obtaining in all cases satisfactory peak retention. Phenyl and PFP stationary phases also effectively retain these compounds via π-π interactions of delocalised electrons on the analytes and the phenyl group in the stationary phase [7,25]. The assayed columns (C18, phenyl and PFP) had different lengths (100, 75 and 50 mm), diameters (2.1, 4.6 and 2.1), and particle sizes (1.8, 5.0 and 2.6 μm), respectively.

For ^{1}D, the best gradients were built with MeOH and water containing both 0.05% TFA. Gradient optimisation was performed for each column, attending to the number of detected peaks. It was found, however, that the same gradient (detailed in Section 10.3.4) offered the best separation of polyphenolic compounds in the olive leaf extracts. In Figure 10.1, the polyphenolic profiles obtained with the C18, phenyl and PFP columns in 1D-LC (Figures 10.1a, b and c, respectively) are depicted. It should be noted that slow flow rates (in this case, 0.1 mL/min) were required for the ^{1}D separation, resulting in long analysis times (*ca.* 50 min).

As can be observed, the polyphenolic profiles obtained with the C18 (Figure 10.1a) and PFP (Figure 10.1c) columns presented better peak distribution and a higher number of visible peaks, compared to the phenyl column (Figure 10.1b). This fact can be partially attributed to the smaller internal diameter of the C18 and PFP columns (2.1 mm), combined with the smaller stationary phase particle size (1.8 and 2.6 μm for C18 and PFP, respectively). This provided better efficiency. However, although both columns offered good polyphenolic profiles in 1D-LC, it was found convenient to further increase the number of resolved peaks to develop LDA models able to distinguish samples according to their cultivar. In this regard, LC×LC should offer advantages, and was next investigated.
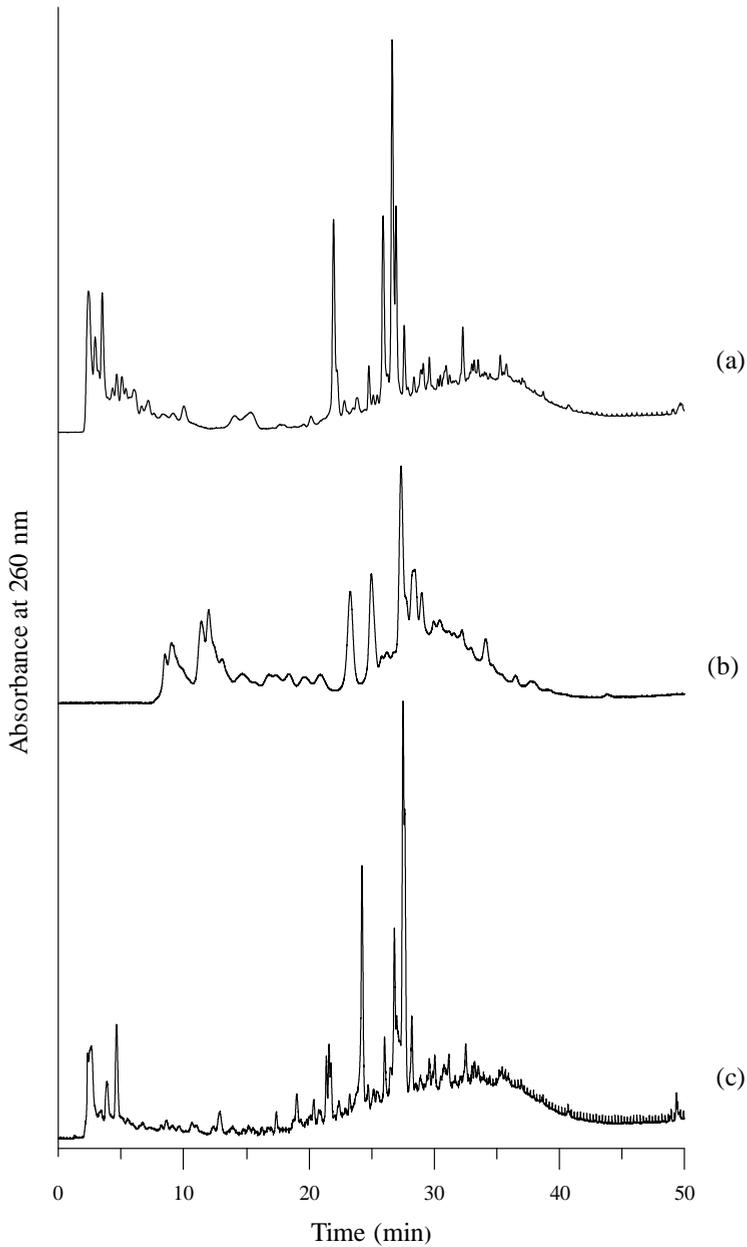
Figure 10.1. 1D-LC chromatograms showing the polyphenolic profile of the Serrana olive leaf extract, obtained with C18 (a), phenyl (b), and PFP (c) columns. The methanol-water gradient is described in Section 10.3.4.

10.4.1.2. ^{2}D separation conditions

The ^{2}D separation of polyphenolic compounds in the olive leaf extracts was next optimised. Considering the good performance of the C18 and PFP stationary phases (Section 10.4.1.1), only these columns were used in ^{1}D. They were combined in ^{2}D with either of the six following columns, with different dimensions, (see Section 10.3.4): C18 (submicro, 1.8 μm), amide (submicro, 1.8 μm), cyano (conventional, 5 μm), phenyl (conventional, 5 μm), and PFP (conventional and submicro, 5 and 2.6 μm). The aim of this study was to test the performance of a range of column chemistries. Conventional columns were used when no alternative submicro column was available. The first trials were done with the 1.8 μm C18 column in ^{1}D.

We should remind that the purpose of this analysis was to obtain the maximal number of detected peaks, without needing any information about their identities. For peak selection, the baseline was subtracted from the signal, and a minimal threshold of detected peak volume of 25 a.u. per min^2 established. The GC Image software automatically delimits the peaks exceeding the peak volume threshold, and indicates the total number of peaks that meet the condition. Several parameters should be optimised in a ^{2}D set-up, such as gradient time, modulation time, flow rate, and type of gradient. One of the most critical parameters is the rate at which a peak is sampled (i.e., modulation time). The peak eluting from the first column should be divided at least in three fractions to maintain the ^{1}D separation in the second column. For this reason, the flow rate in ^{1}D should be quite low, whereas the separation in ^{2}D should be carried out at a high flow rate [13].

The first column combination was formed with C18 and cyano stationary phases in ^{1}D and ^{2}D, respectively. The flow rates were 0.1 and 2.5 mL/min for each column. In ^{2}D, the gradient time was 0.3 min (18 s), followed by a time

period of 0.1 min (6 s) to return to the initial conditions. This gave rise altogether to a modulation time of 0.4 min. Under these conditions, the 40 µL-loop was totally filled. The ^{2}D separation was carried out using a shifted gradient, which occupied the maximal area of the ^{2}D separation space [26]. The initially assayed gradient (Figure 10.2a) offered the LC×LC separation illustrated in Figure 10.3a, where 29 insufficiently resolved peaks were detected. To improve the separation, the shifted gradient was readjusted by decreasing the ACN composition, as can be seen in Figure 10.2b. As a result, peaks eluting between 20 and 40 min were resolved, which gave rise to an increase in the number of detected peaks (39 compounds) (Figure 10.3d).

An LC×LC set-up based on the combination of C18 (^{1}D) and phenyl (^{2}D) columns was next assayed. The initial conditions were based on the optimal ones found for the combination of the C18 and cyano columns described above (gradient shown in Figure 10.2b). The achieved separation is depicted in Figure 10.3b. As observed, the compounds were not completely eluted in the assayed conditions, and the peaks appeared on the upper part of the 2D-LC space when a new ^{2}D cycle was started. For this reason, the ^{2}D gradient was modified by increasing the proportion of ACN up to get adequate elution (Figure 10.2c). The results obtained are shown in Figure 10.3e, where no improvement in the number of detected peaks was found (36 and 37 before and after the optimisation, respectively).
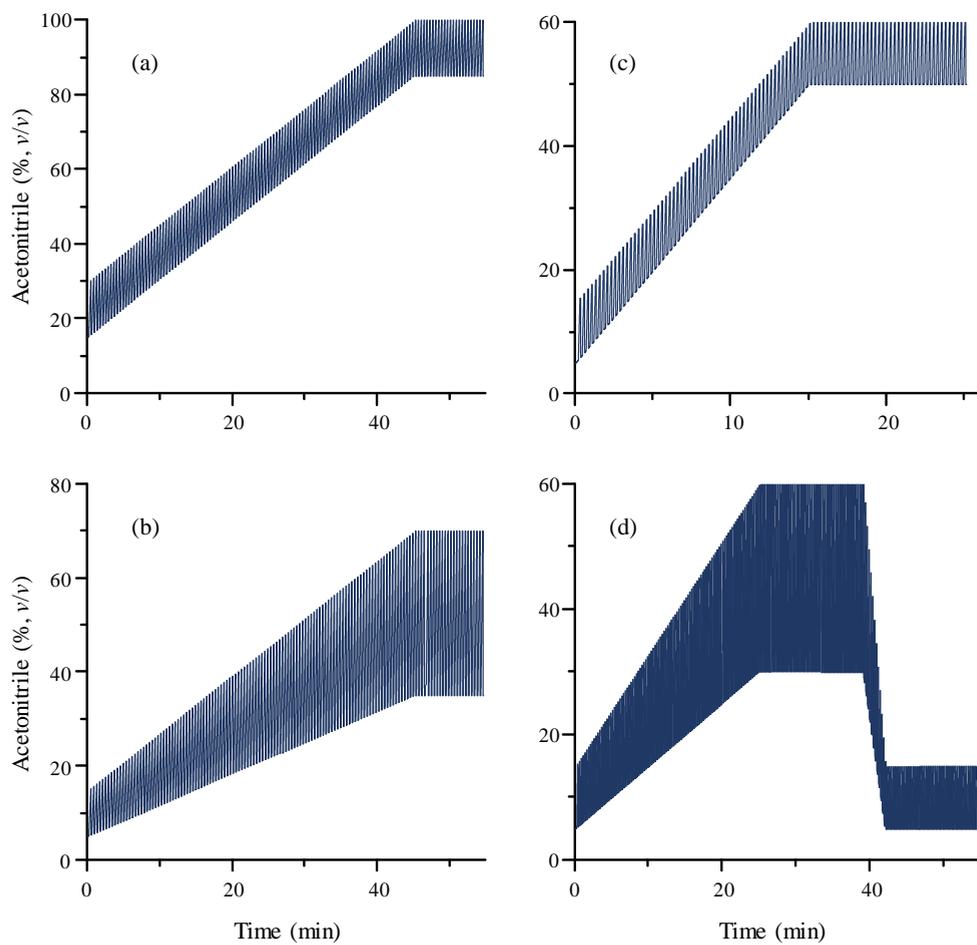
Figure 10.2. Acetonitrile-water shifted gradients used in the separations carried out in ^{2}D.
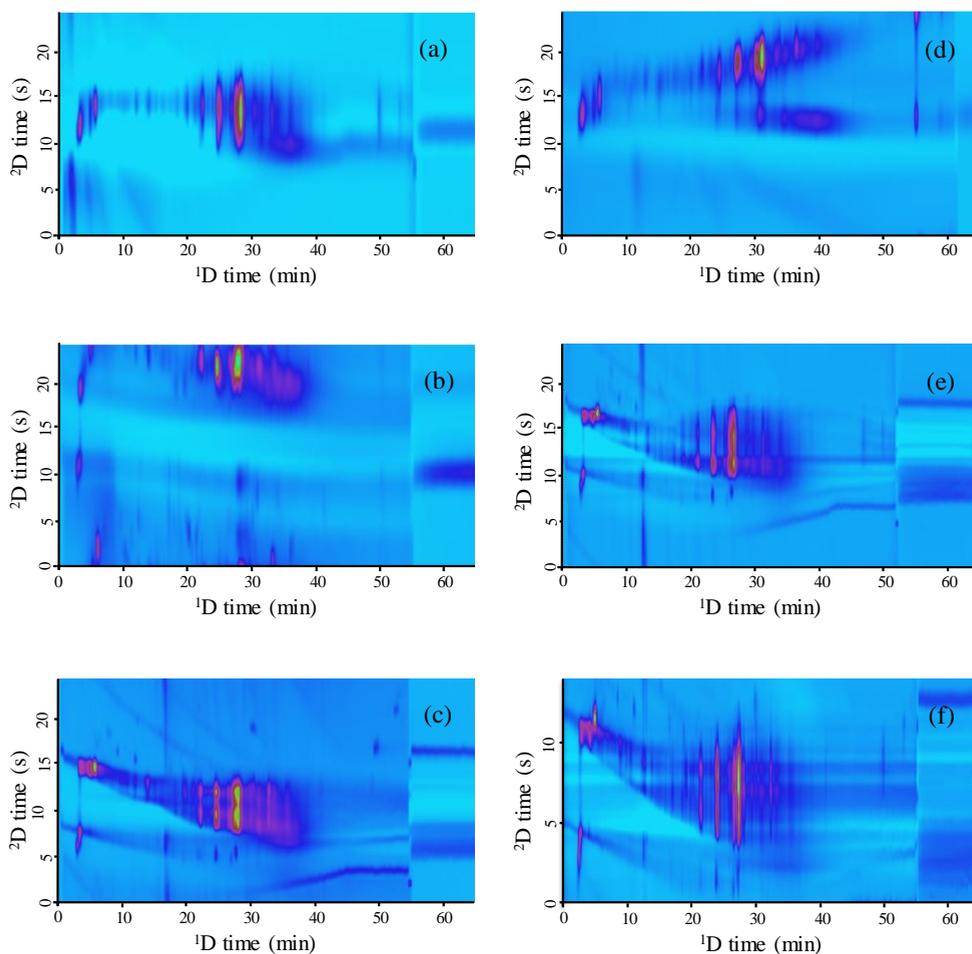
Figure 10.3. 2D-LC separation of polyphenolic compounds from a Serrana olive leaf extract using a submicro C18 column in ^{1}D, and the following conventional columns in ^{2}D: cyano before (a) and after optimisation (d); phenyl before (b) and after optimisation (e), and PFP before (c) and after optimisation (f). The methanol-water gradient used in ^{1}D is that optimised for 1D-LC. The optimised acetonitrile gradient for ^{2}D is indicated in Section 10.4.1.2 (see also Figure 10.2).

A third combination was assayed with the C18 column in ^{1}D and the conventional PFP in ^{2}D. The first gradient assayed (Figure 10.2a) provided the best results (Figure 10.3c), with a higher number of detected peaks with regard to the previous column combinations (48 peaks were detected). In view of the satisfactory results found with this combination, the possibility of increasing the number of peaks by improving peak sampling was considered. With a reduction in the modulation time to 0.25 min (15 s), only 63% of the 40-μL loop was filled. Operating in this way, most peaks were spread all over the separation space (Figure 10.3f), making the detection of minor compounds possible: 73 peaks were visible instead of the 48 peaks obtained in the previous conditions.

The three column combinations described above were formed with a C18 column in ^{1}D and three conventional columns in ^{2}D. It should be noted that, in LC×LC, short and narrow columns, with small particles, are usually recommended for ^{2}D, in order to provide short analysis times and high efficiencies [6,17,20]. Consequently, we decided to investigate the effect of a stationary phase with a smaller particle size (submicro columns), which require the use of UHPLC pumps and high temperature (50 ºC). The performance of the amide (1.8 μm) and PFP (2.6 μm) stationary phases in ^{2}D was thus evaluated. For the C18×amide set-up, the gradient shown in Figure 10.2a was first assayed, which resulted in the LC×LC profile depicted in Figure 10.4a. In the 2D-LC chromatogram, the polyphenolic compounds did not show significant retention, owing to the high ACN concentrations in the gradient (57 peaks were detected). As expected, readjusting the ACN percentages as shown in the gradient in Figure 10.2b, longer retention was possible (Figure 10.4d). This allowed the number of visible peaks be increased to 76. This number is similar to that obtained with the combination of the C18 column and conventional PFP.
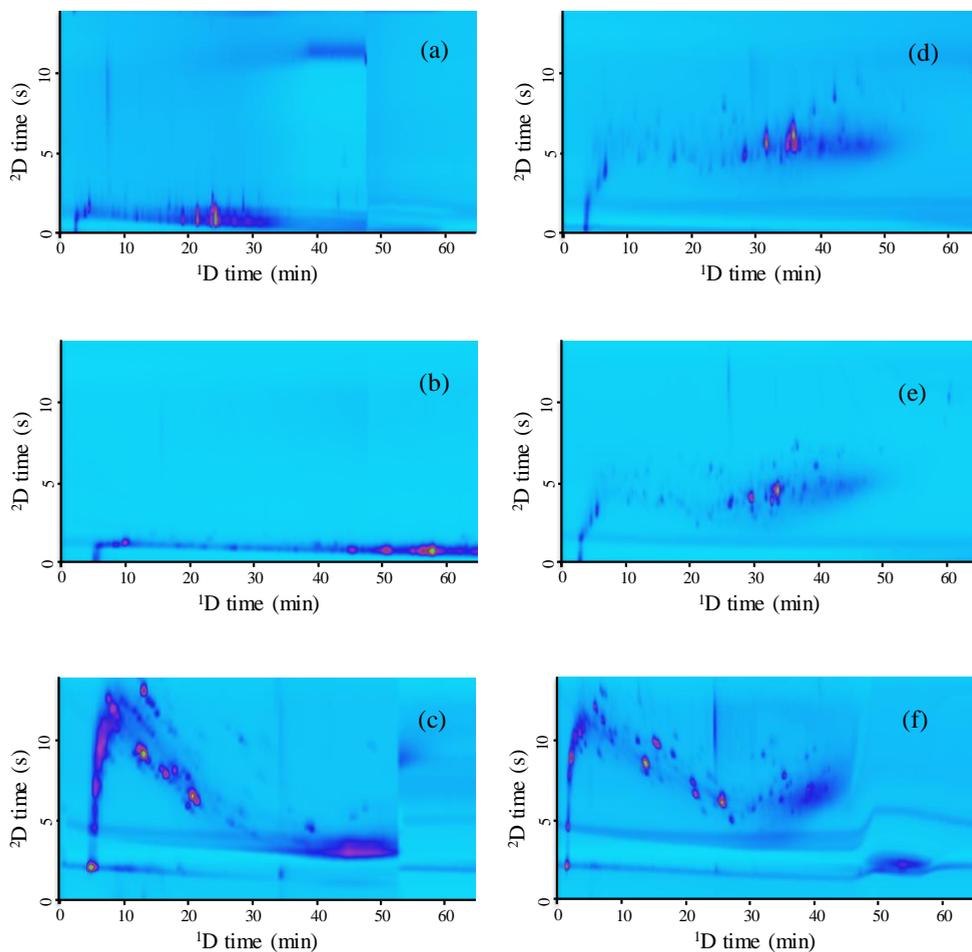
Figure 10.4. 2D-LC chromatograms for polyphenolic compounds from a Serrana olive leaf extract, using the following combinations of submicro columns: C18×amide before (a) and after optimisation (d); C18×PFP before (b) and after optimisation (e), and PFP×C18 before (c) and after optimisation (f). The optimised acetonitrile gradient for ^{2}D is indicated in Section 10.4.1.2 (see also Figure 10.2).

Next, the previous separation with the C18×PFP configuration was thought could be improved by replacing the conventional PFP column with a similar submicro. The gradient depicted in Figure 10.2a resulted again in an inadequate separation, since all analytes were eluted close to the solvent front (Figure 10.4b) (only 25 peaks were detected). As previously observed, by decreasing the ACN composition along the ^{2}D separation (Figure 10.2b), the retention of the polyphenolic compounds was increased. This allowed higher retention with better resolution, resulting in the detection of 83 peaks (Figure 10.4e).

Up to this point, the best 2D-LC separation and highest number of detected peaks corresponded to the combination of the submicro C18 and PFP columns (C18×PFP). We still thought it would be interesting to check the effect of changing the order of the stationary phases, by placing the PFP column in ^{1}D and C18 column in ^{2}D (PFP×C18). The length of the C18 column in previous combinations was 10 cm. For this assay, a 5-cm C18 column was used instead to guarantee a fast gradient in a short time. Initially, the separation in ^{2}D was carried out using a shifted gradient (Figure 10.2c), which offered a good LC×LC profile for the polyphenolic compounds (Figure 10.4c). However, an empty region between 20 and 40 min appeared, with a large unresolved peak cluster between 40 and 50 min in ^{1}D. To solve both problems, a segmented gradient (Figure 10.2d) was used in order to get better distribution of the peaks in the ^{2}D region (Figure 10.4f). This ^{2}D profile allowed the detection of a total of 104 peaks, being the best chromatographic performance obtained along this work. Consequently, this LC×LC system was chosen for further analysis of the olive leaf and pulp extracts from several cultivars.

10.4.2. Profiling of polyphenolic fraction for olive leaves and pulps from different cultivars

Comprehensive 2D-LC is considered a potential tool to discriminate among samples from different genetic origin, since it provides a high amount of information from complex matrices derived from the LC×LC chromatographic profiles [20–22,27]. This work is focused on differentiating olive leaves and pulps, according to their genetic origin. For this purpose, polyphenolic extracts of the olive leaves and pulps included in Table 10.1 were analysed using the optimal LC×LC set-up (the PFP×C18 combination) described in Section 10.4.1.2. In Figure 10.5, representative LC×LC profiles of Blanqueta olive leaf (Figure 10.5a) and pulp extracts (Figure 10.5c), and Cornicabra olive leaf (Figure 10.5b) and pulp extracts (Figure 10.5d), are shown, where the detected peaks are marked. As observed, the polyphenolic profiles for the olive leaves and pulps show differences, since many components are present in leaves, but not in pulps, and viceversa. The number of detected peaks is given in Table 10.2 for all analysed samples.
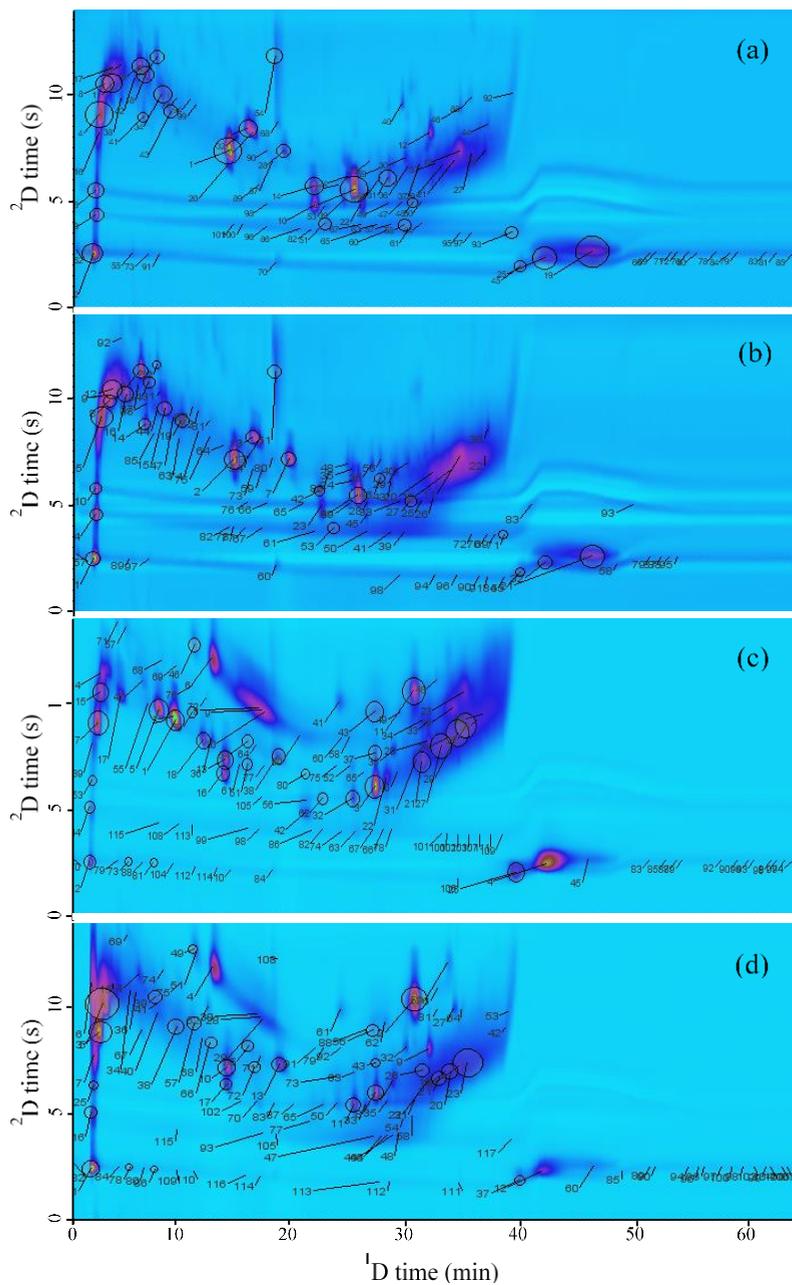
Figure 10.5. Comprehensive 2D-LC separation for Blanqueta and Cornicabra olive leaf ((a) and (b)) and pulp ((c) and (d)) extracts, using the optimised conditions (see Section 10.3.4).

Table 10.2. Number of detected peaks under the optimised LC×LC conditions for all the olive leaves and pulps included in this study.[a]

Cultivar	Leaves	Pulp
Arbequina	89	101
Blanqueta	102	109
Cornicabra	99	109
Hojiblanca	112	92
Picual	88	97
Serrana	106	100
Villalonga	86	85

[a] Mean values from three replicated extractions for each crop year are given.

10.4.3. Construction of LDA models from optimal peak volume ratios

The detected peaks for the olive leaf and pulp polyphenolic extracts were compared for all samples (Table 10.1). Those exhibiting the same retention time in both dimensions were selected. Data treatment was applied to the 26 and 29 peaks, which were common for all olive leaves and pulps. As example, the peaks selected for Blanqueta and Cornicabra are marked with a circle on the LC×LC chromatograms depicted in Figure 10.5.

LDA models were built using peak volume ratios as predictors, instead of the peak volumes. Peak volume ratios have the advantage of emphasising differences among components, whereas account to a certain extent for the differences associated to the extraction process. The ratios were calculated by dividing the volume of each peak in the LC×LC chromatogram by each of the volumes for the other peaks. Taking into account that each pair of peaks should be considered only once, the total number of predictors are $n(n-1)/2$, where n is the number of original variables. For the case of study, the number of non-redundant peak ratios available to be used as predictors were: $(26×25)/2 = 325$ and $(29×28)/2 = 406$, for leaf and pulp samples, respectively.

The samples corresponded to seven cultivars (from different geographical origin), obtained at two cropping times, each of them three-fold replicated (Table 10.1). Since samples consisted of leaves and pulps, two data matrices were constructed with 325 and 406 columns (predictors), respectively. Both matrices contained 42 rows (or objects: 7 cultivars × 2 crop years × 3 extractions). A response variable was added to each matrix, denoting the category to which the sample belongs from the seven cultivars of olive leaves and pulps. Since the cultivar tags are arbitrary, the associated variable should be categorical.

For data treatment, each of the two final matrices (for leaves and pulps) was divided in two groups: the training and validation sets. The training set was composed by 35 objects, which corresponded to 5 objects randomly selected for each cultivar, while the remaining samples (7) were assigned to the validation set. The SPSS algorithm selected 15 predictors (from the 325 available) for leaves, and 14 predictors (from the 406 available) for pulps, as optimal set with the largest discriminant capability.

Table 10.3. Selected predictors and corresponding standardised coefficients of the LDA model built to classify olive leaves and pulp samples, according to their cultivar.

	Olive leaves samples				
Predictors[a]	f_1	f_2	f_3	f_4	f_5
1/8	-1.69	-0.30	-0.25	-0.12	-0.39
3/6	0.46	0.31	1.56	0.87	-0.22
3/8	1.34	-0.23	-1.25	0.55	0.29
4/18	1.77	-0.07	0.54	-0.67	0.49
6/11	2.63	-0.56	-0.80	1.28	0.42
6/22	-1.59	0.56	1.34	-0.33	0.67
7/10	0.92	2.42	-0.99	0.360	0.73
7/14	-2.32	0.04	1.33	0.86	0.42
8/11	-1.47	-1.12	-0.24	0.40	-0.33
8/17	1.10	0.12	-1.27	-0.57	0.50
8/24	-1.86	-1.87	-0.10	0.65	1.20
10/22	6.12	-0.34	-1.73	1.19	0.58
13/22	-4.25	1.62	0.05	-0.23	-0.39
14/24	-0.24	2.02	2.04	1.37	-0.73
20/23	0.88	-0.62	0.61	-3.71	0.15

	Olive pulp samples					
Predictors[a]	f_1	f_2	f_3	f_4	f_5	f_6
1/12	-1.17	0.59	0.31	-0.12	-0.27	-0.26
3/8	-0.60	0.40	-0.96	-0.60	0.96	0.74
4/9	4.42	0.32	-0.24	0.13	0.18	0.17
5/17	2.81	-0.06	-0.03	0.14	-0.06	0.17
5/19	-1.82	0.05	-0.34	1.39	-0.28	-0.04
6/19	3.92	-0.51	0.56	-0.61	0.01	0.40
7/28	-1.01	1.65	-0.76	-0.50	-0.22	-1.29
9/16	3.90	-0.15	1.78	0.96	0.40	-0.84
14/16	-0.97	0.84	-1.07	0.42	0.15	0.75
15/29	0.87	0.97	0.79	-0.53	0.34	0.22
17/26	0.06	0.09	-0.25	0.47	-0.75	-0.08
19/27	0.27	-1.88	0.53	0.52	-0.22	1.70
20/27	0.62	1.48	0.87	-0.39	-0.003	0.55
23/28	0.09	-0.51	0.65	0.94	0.94	-0.53

[a] Polyphenolic peak volume ratios.
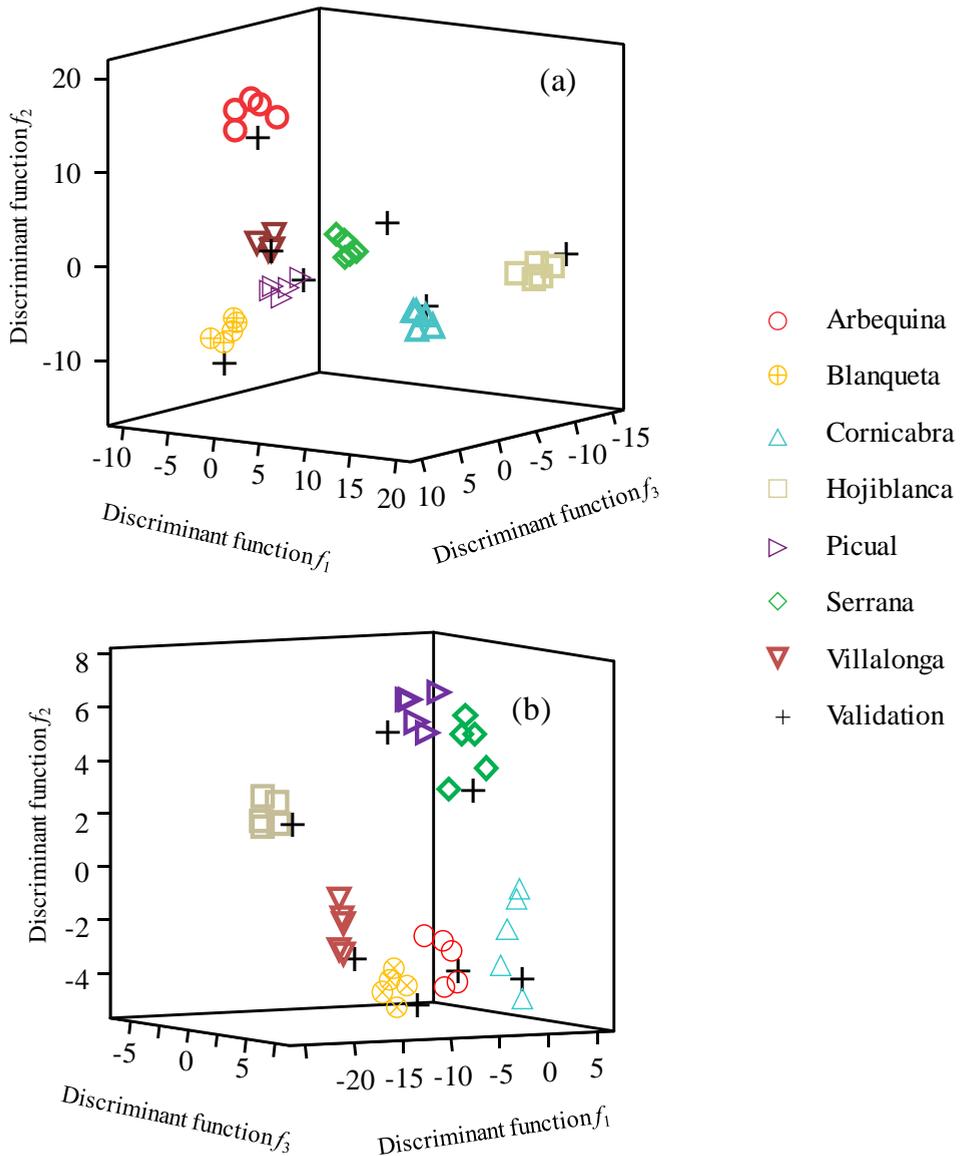
Figure 10.6. Three-dimensional plot showing the scores associated to the three first discriminant functions constituting the LDA model used for the classification of olive leaves (a) and pulps (b), according to their cultivar. Validation samples are labelled with a cross symbol. The cubes were tilted in order to get a proper perspective.

The predictors with the highest discriminant capability and the standardised coefficients of the discriminant functions for each selected predictor are indicated in Table 10.3.

Figure 10.6a shows a three dimensional plot drawn according to the three first discriminant functions, where each olive leaf sample is represented according to its scores, using the information provided by the LC×LC chromatogram. The plot shows that all samples in a given class are grouped in compact clusters, and neatly separated from the other classes. Concerning the prediction capability of the model, all objects in the validation set (represented with crosses in Figure 10.6a) were correctly assigned with a 95% probability level. Also, an excellent separation was achieved ($\lambda_w < 0.01$), for all category pairs.

Similar satisfactory results were obtained when the LDA algorithm was applied to the classification of the olive pulp extracts (Figure 10.6b). It was found that all category pairs were well separated with a λ_w value below 0.01 and all the objects (training and validation set) were correctly classified.

10.5. Conclusions

The possibility of distinguishing olive leaves and pulps, according to their cultivar, by using polyphenolic profiles obtained by comprehensive 2D-LC (LC×LC), is demonstrated. Polyphenolic profiles were obtained for the two types of samples after optimisation of both dimensions in the chromatographic set-up, in terms of column type and length, particle size and gradient profile. The polyphenolic extracts were analysed using several combinations of conventional (5 μm) and submicro (1.8 and 2.6 μm) columns. A particular optimisation of the gradient conditions was carried out with the purpose of

getting the maximal number of detected peaks. The optimisation process, assisted by the GC Image software that delimited the peaks exceeding a pre-selected peak volume threshold, is described in detail. The study started using a standard combination of a C18 column in ^{1}D and a cyano column in ^{2}D, and a conventional gradient, which resulted in a poor number of peaks (29) insufficiently resolved. Change of the cyano column by a conventional PFP column, together with the reduction in the modulation time increased the number of peaks (73). Finally, by using submicro columns in both dimensions, changing the order of the columns in the set-up (PFP column in ^{1}D and C18 in ^{2}D), and applying a shifted gradient, 112 and 109 peaks could be clearly detected for the olive leaf and pulp extracts, respectively.

The comparison of the polyphenolic profiles of samples from different cultivars showed 26 and 29 common peaks, for leaf and pulp samples, respectively, which were selected for the data treatment. The peak volume ratios of polyphenolic compounds were used as predictors to construct the LDA models able to discriminate the samples according to the cultivar. Volume ratios were used in order to increase the separation between the different cultivars, reducing the extraction recovery effect. Both olive leaf and pulp samples, belonging to seven cultivars from different Spanish regions, were correctly classified with an excellent separation among all categories, with assignment probabilities above 95%. This demonstrates that polyphenolic profiles are characteristic of each cultivar.

10.6. References

[1] M. Vergara Barberán, M.J. Lerma García, J.M. Herrero Martínez, E.F. Simó Alfonso, Use of protein profiles established by CZE to predict the cultivar of olive leaves and pulps, *Electrophoresis* 35 (2014) 1652–1659.

[2] M. Vergara Barberán, M.J. Lerma García, J.M. Herrero Martínez, E.F. Simó Alfonso, Classification of olive leaves and pulps according to their cultivar by using protein profiles established by capillary gel electrophoresis, *Anal. Bioanal. Chem.* 406 (2014) 1731–1738.

[3] A. Taamalli, D. Arráez Román, A.M. Gómez Caravaca, M. Zarrouk, A. Segura Carretero, Geographical characterization of Tunisian olive tree leaves (cv. Chemlali) using HPLC-ESI-TOF and IT/MS fingerprinting with hierarchical cluster analysis, *J. Anal. Methods Chem.* (2018) (https://doi.org/10.1155/ 2018/6789704).

[4] L.M. Crawford, D.M. Holstege, S.C. Wang, High-throughput extraction method for phenolic compounds in olive fruit (*Olea europaea*), *J. Food Compos. Anal.* 66 (2018) 136–144.

[5] M. Esteki, J. Simal Gandara, Z. Shahsavari, S. Zandbaaf, E. Dashtaki, Y. Vander Heyden, A review on the application of chromatographic methods, coupled to chemometrics for food authentication, *Food Control* 93 (2018) 165–182.

[6] F. Cacciola, S. Farnetti, P. Dugo, P.J. Marriott, L. Mondello, Comprehensive two-dimensional liquid chromatography for polyphenol analysis in foodstuffs, *J. Sep. Sci.* 40 (2017) 7–24.

[7]   C. Ancillotti, S. Orlandini, L. Ciofi, B. Pasquini, C. Caprini, C. Droandi, M. del Bubba, Quality by design compliant strategy for the development of a liquid chromatography-tandem mass spectrometry method for the determination of selected polyphenols in Diospyros kaki, *J. Chromatogr. A* 1569 (2018) 79–90.

[8]   P.J. Xie, L.X. Huang, C.H. Zhang, Y.L. Zhang, Phenolic compositions, and antioxidant performance of olive leaf and fruit (*Olea europaea* L.) extracts and their structure-activity relationships, *J. Funct. Foods* 16 (2015) 460–471.

[9]   M. Cabrera Bañegil, T. Schaide, R. Manzano, J. Delgado Adámez, I. Durán Merás, D. Martín Vertedor, Optimization and validation of a rapid liquid chromatography method for determination of the main polyphenolic compounds in table olives and in olive paste, *Food Chem.* 233 (2017) 164–173.

[10] A. Romani, S. Mulas, D. Heimler, Polyphenols and secoiridoids in raw material (*Olea europaea* L. leaves) and commercial food supplements, *Eur. Food Res. Technol.* 243 (2017) 429–435.

[11] G. Tóth, A. Alberti, A. Sólyomváry, C. Barabás, I. Boldizsár, B. Noszál, Phenolic profiling of various olive bark-types and leaves: HPLC–ESI/MS study, *Ind. Crops Prod.* 67 (2015) 432–438.

[12] M. Ricciutelli, S. Marconi, M.C. Boarelli, G. Caprioli, G. Sagratini, R. Ballini, D. Fiorini, Olive oil polyphenols: A quantitative method by high-performance liquid-chromatography-diode-array detection for their determination and the assessment of the related health claim, *J. Chromatogr. A* 1481 (2017) 53–63.

[13]   M. Kivilompolo, T. Hyötyläinen, Comprehensive two-dimensional liquid chromatography in analysis of *Lamiaceae* herbs: Characterisation and quantification of antioxidant phenolic acids, *J. Chromatogr. A* 1145 (2007) 155–164.

[14]   S. Toro Uribe, L. Montero, L. López Giraldo, E. Ibáñez, M. Herrero, Characterization of secondary metabolites from green cocoa beans using focusing-modulated comprehensive two-dimensional liquid chromatography coupled to tandem mass spectrometry, *Anal. Chim. Acta* 1036 (2018) 204–213.

[15]   L. Montero, V. Sáez, D. von Baer, A. Cifuentes, M. Herrero, Profiling of *Vitis vinifera L. canes* (poly)phenolic compounds using comprehensive two-dimensional liquid chromatography, *J. Chromatogr. A* 1536 (2018) 205–215.

[16]   L.T. Vaz Freire, M.D. da Silva, A.M. Freitas, Comprehensive two-dimensional gas chromatography for fingerprint pattern recognition in olive oils produced by two different techniques in Portuguese olive varieties Galega Vulgar, Cobrançosa e Carrasquenha, *Anal. Chim. Acta* 633 (2009) 263–270.

[17]   F. Cacciola, P. Donato, L. Mondello, P. Dugo, Recent advances in comprehensive two-dimensional liquid chromatography for the analysis of natural products, in *Handbook of Advanced Chromatography/Mass Spectrometry Techniques*, AOCS Press, London, 2017, pp. 287–307.

[18]   N.S. Wilson, M.D. Nelson, J.W. Dolan, L.R. Snyder, P.W. Carr, Column selectivity in reversed-phase liquid chromatography: II. Effect of a change in conditions, *J. Chromatogr. A* 961 (2002) 195–215.

[19] J. Pellett, P. Lukulay, Y. Mao, W. Bowen, R. Reed, M. Ma, R.C. Munger, J.W. Dolan, Loren Wrisley, K. Medwid, N.P. Toltl, C.C. Chan, M. Skibic, Kallol Biswas, K.A. Wells, L.R. Snyder, "Orthogonal" separations for reversed-phase liquid chromatography, *J. Chromatogr. A* 1101 (2006) 122–135.

[20] P. Dugo, F. Cacciola, P. Donato, R.A. Jacques, E.B. Caramão, L. Mondello, High efficiency liquid chromatography techniques coupled to mass spectrometry for the characterization of mate extracts, *J. Chromatogr. A* 1216 (2009) 7213–7221.

[21] L. Montero, M. Herrero, E. Ibáñez, A. Cifuentes, Profiling of phenolic compounds from different apple varieties using comprehensive two-dimensional liquid chromatography, *J. Chromatogr. A* 1313 (2013) 275–283.

[22] L. Montero, E. Ibáñez, M. Russo, R. di Sanzo, L. Rastrelli, A.L. Piccinelli, M. Herrero, Metabolite profiling of licorice (*Glycyrrhiza glabra*) from different locations using comprehensive two-dimensional liquid chromatography coupled to diode array and tandem mass spectrometry detection, *Anal. Chim. Acta* 913 (2016) 145–159.

[23] T. Jerman, P. Trebše, B.M. Vodopivec, Ultrasound-assisted solid liquid extraction (USLE) of olive fruit (*Olea europaea*) phenolic compounds, *Food Chem.* 123 (2010) 175–182.

[24] R. Martí, M. Valcárcel, J.M. Herrero Martínez, J. Cebolla Cornejo, S. Roselló, Fast simultaneous determination of prominent polyphenols in vegetables and fruits by reversed phase liquid chromatography using a fused-core column, *Food Chem.* 169 (2015) 169–179.

[25] I. Tomaz, L. Maslov, Simultaneous determination of phenolic compounds in different matrices using phenyl-hexyl stationary phase, *Food Anal. Methods* 9 (2016) 401–410.

[26] F. Bedani, W.T. Kok, H.G. Janssen, Optimal gradient operation in comprehensive liquid chromatography × liquid chromatography systems with limited orthogonality, *Anal. Chim. Acta* 654 (2009) 77–84.

[27] M. Kula, D. Głód, M. Krauze-Baranowska, Application of on-line and off-line heart-cutting LC in determination of secondary metabolites from the flowers of *Lonicera caerulea* cultivar varieties, *J. Pharm. Biomed. Anal.* 131 (2016) 316–326.

SUMMARY AND CONCLUSIONS

Reversed phase liquid chromatography (RPLC) is usually the technique of choice for the analysis of a wide range of organic compounds, due to its versatility, robustness and sensitivity. However, the selectivity and analysis time depend in a complex way on several experimental factors that interact each other, such as the concentration of organic solvent, pH and temperature. Due to the difficulty in finding experimental conditions that simultaneously separate all the compounds in a sample, optimisations based on trial and error are laborious and sometimes unsuccessful, and there is no guarantee of finding the true optimal separation.

The best separation conditions should be preferably obtained using the extracted information from a reduced set of carefully planned experiments, covering the entire space of the experimental factors. The data from these experiments are used with the purpose of fitting a retention model for each analyte, in order to predict the retention times at any new arbitrary condition within the experimental domain and simulate chromatograms for mixtures of compounds. Finally, the best conditions are selected using computer-assisted methodologies in the so-called interpretive optimisations. The fitted models can also give information on the interactions established inside the chromatographic column.

This PhD. Project includes fundamental studies to improve interpretive optimisation methodologies and their application to the analysis of physiological fluids and natural products (olive leaf and pulp extracts, and medicinal herbs). The determination of several groups of compounds was considered: alkylbenzenes, sulphonamides, β-adrenoceptor antagonists, amino acids, phenols and polyphenols, and unknown compounds in a wide range of polarities contained in medicinal herbs. Most analysis were carried out with mobile phases of acetonitrile-water in isocratic and gradient elution, but the

presence of secondary equilibria when a surfactant was added to the mobile phase was also investigated.

Along the work, new strategies and tools, some without previous antecedents, were developed, which required the construction of diverse software. The performance of the new developments was compared with others in published reports, when available. The work is outlined in two parts:

- Part I: Increasing the modelling capability in liquid chromatography
- Part II: Improving the separation performance for chromatographic fingerprints

Next, the general conclusions from each chapter are summarised.

C.1. Increasing the modelling capability in liquid chromatography

The reliability of interpretive strategies depends significantly on the accuracy of the models used in the prediction of retention times and peak profiles, which are built from the information obtained from standards. Part I gathers contributions dedicated to optimise the experimental designs needed to build the models. It also contains several proposals on the application of the models to obtain information about the interactions that take place inside a chromatographic column, estimate the peak capacity in both isocratic and gradient elution, and optimise gradient elution using eluents that contain a surfactant in micellar and submicellar conditions. The most relevant aspects of each proposal are described below.

C.1.1. Modelling retention and peak shape of small polar solutes analysed by nano-HPLC using methacrylate-based monolithic columns

- Several polymeric monolithic columns containing different amounts of hydrophobic and hydrophilic monomers were prepared and tested, using alkylbenzenes (non-polar) and sulphonamides (polar) as probe compounds: a column containing lauryl methacrylate (LMA), which confers a dominant hydrophobic character; a column of intermediate polarity formed with a mixture of hydrophobic (LMA) and ionisable monomers (methacrylic acid, MAA); and a column with a more polar monomer (hexyl methacrylate, HMA), combined with MAA.

- Among the columns, a monolith composed of HMA, MAA, and ethylene dimethacrylate (EDMA), was selected, based on its better chromatographic resolution and reasonable analysis times, for both sets of compounds. In spite of the presence of moderately polar methacrylic acid groups in the poly(HMA-co-MAA-co-EDMA) monolithic column, the elution order observed for the alkylbenzenes (with a regular distribution of the retention times) proved the importance of the hydrophobic interactions. In contrast, the behaviour of the polar sulphonamides was irregular, with the compounds distributed in three groups according to their retention, with co-elution in most assayed experimental conditions and peak reversals at high organic solvent contents. However, the resolution of sulphonamides was highly improved with respect to previous monolithic columns.

- The chromatographic behaviour of the probe compounds with the best monolithic column was analysed by modelling the retention times and peak profiles. The accuracy of several retention models was studied (Equations (2.1) to (2.12)), including a model describing a mixed retention mechanism. The fitted parameters for this model suggested that the retention mechanism was based mainly on adsorption for both sets of compounds (alkylbenzenes and sulphonamides). All assayed models provided acceptable predictions, with relative errors often below 1.0%. The model performance for the monolithic column was similar or better, compared to that found with conventional RPLC columns, for the same compounds.

- The correlations between the parameters (S_1, S_2 and q) in the logarithmic quadratic model that includes the P_M^N transformation (Equation (2.7)), instead of the organic solvent contents, gave information on the retention behaviour of sulphonamides with the monolithic column. The high scattered correlations observed between the S_1 and S_2 model parameters (which quantify the elution strength of the mobile phase and the deviation of the model from linearity), and the intercept q (which quantifies the retention level of the solutes), indicated a significant variability in the retention behaviour for each sulphonamide, with regard to alkylbenzenes. This was explained by the existence of different proportions of hydrophilic and hydrophobic interactions between sulphonamides (with different molecular structures), and the polar and non-polar monomers in the monolithic column.

- The correlations of the peak half-widths with the retention times, for the peaks obtained with the monolithic and conventional C18 columns, revealed a diversity of interactions for the studied alkylbenzenes and sulphonamides. The significant dispersion observed in the correlation of the right half-widths, for the sulphonamides analysed with the monolithic column, denoted particular kinetics for each compound. This indicated again the diverse participation of polar and non-polar monomers in the monolithic column, in their interaction with sulphonamides.

C.1.2. Benefits of solvent concentration pulses in retention time modelling of liquid chromatography

- Isocratic experimental designs provide the richest information about the behavior of solutes to fit retention models, giving rise to the most accurate model parameters with narrower confidence intervals. However, the use of isocratic designs is hampered by the long retention times of the less polar solutes in mixtures with other analytes, especially at low organic solvent contents. The usual solution is the use of experimental designs formed with gradients of organic solvent, where its content is gradually increased to reduce the retention times. However, designs containing gradient experiments give rise to less accurate models, and consequently, their prediction performance is worst.

- As an alternative, the use of isocratic experimental designs, including sudden increments (i.e., pulses) of organic solvent in the mobile phases of lowest elution strength, was explored. Runs containing pulses are a type of multi-isocratic gradients that allow obtaining chromatographic information for non-polar solutes at low organic solvent contents. The effect of the pulse is moving in block the retention times of late eluting compounds, in isocratic elution, to earlier times. Faster solutes elute before the pulse, and the most retained solutes after the pulse in acceptable times.

- Mixed designs can be easily constructed by replacing the slowest isocratic runs with runs containing one or two pulses of short duration, at intermediate times. The pulse location can be set arbitrarily, but the best choice is locating it in an empty intermediate region of the chromatogram. Since runs containing a pulse has important effects on the selectivity and retention of the solutes eluting after the pulse, its position, duration and sudden increase in organic solvent content must be adapted to each sample.

- The inclusion of pulses is not practical with optimisation purposes, due to the increased peak overlapping, especially in the pulse region, and because solutes suffer important drops in efficiency after the pulse. However, they are beneficial for retention modelling.

- The predictions of the elution conditions for runs containing pulses were made using the fundamental equation for gradient elution. The retention times calculated numerically showed notable deviations for solutes eluted close to the pulse, even using a retention model with low prediction errors. When the intra-column delay (i.e., the time required

for the solvent front to reach the solute from the column inlet) was taken into account, the predictions were improved and agreed satisfactorily with the experimental chromatograms.

- When the predictions with designs containing pulses or gradients were carried out inside the experimental domain, the difference between predicted and experimental times was below 0.01 min. The designs with pulses provided parameters for the retention models similar to those obtained with isocratic designs, which as commented, are considered the most accurate for predictions. Designs with a single pulse were checked to be the most accurate. For out-of-domain predictions, the predictive performance of designs containing pulses was also similar to the performance of designs with only isocratic experiments.

- In general, designs containing pulses were proved to be very competitive with regard to gradient designs, in terms of analysis time and solvent consumption. Although gradient designs with variable gradient time yielded the smallest analysis time and solvent waste, the errors in the model parameters and the deviations in the extrapolated predictions were larger.

C.1.3. Testing experimental designs in liquid chromatography: Development and validation of a method for the comprehensive inspection of experimental designs

- Designs containing gradient runs are preferred by many analysts, not only for making the analyses, but also to build experimental designs for modelling purposes. However, finding a design with an optimal distribution of gradients is not straightforward. In order to find the best experimental designs (formed of isocratic runs or gradients), a universal methodology for assessing their quality was developed. The approach makes use of the G-optimality principle, which is based on the error propagation theory, and relates the mathematical properties of a retention model with a given distribution of points in an experimental design. To our knowledge, there is no such powerful methodology to evaluate experimental designs.

- The methodology estimates the variance associated to the prediction of retention times using an expression that considers two Jacobian matrices, associated to training ($\mathbf{J}_{train}$) and sampling experiments ($\mathbf{J}_{pred}$) (Equation (4.19)). The Jacobian matrices imply the calculation of partial derivatives of the retention models for a large set of conditions. For gradient elution, the computing time may be particularly long since it needs the prediction of the retention time by integration of the fundamental equation, which can imply massive calculations. However, the computation time was significantly reduced taking advantage of recent developments in the laboratory of the research group.

- The proposed methodology was validated by checking the capability of five training designs, common in RPLC, to build models used in the prediction of the retention of 14 sulphonamides, according to the runs in sampling designs for isocratic and gradient elution. The equation proposed by Neue-Kuss to describe the retention gave better prediction accuracy than the Linear Solvent Strength (LSS) model, with relative errors in predictions below 0.7%. The LSS model, which is extensively used for gradient elution, was found to yield lack of fit, and was discarded.

- The comparison of the training designs was assisted by maps, where the relative uncertainties in the predictions according to the runs in the sampling designs were plotted for each compound: for the isocratic designs against the mobile phase compositions with increments of 1% acetonitrile, and for the gradient designs against the ramp slopes with constant angular increments of 3º. Relative uncertainties provided more meaningful and interpretable results than absolute uncertainties, which were strongly variable and dependent on solute retention.

- The accuracy level in the calculation of the gradient time was found critical for the calculation of the derivatives in the Jacobian matrices. When the accuracy level was insufficient, the uncertainty maps contained noisy curves. With an accuracy level of around 10^{-15}, the curves were smooth, and in most cases, a characteristic U-pattern with increments at both extremes and minor errors in between were obtained for gradient elution.

- For all training designs, the intermediate regions of the uncertainty maps showed a systematic change at decreasing solute polarity. The magnitude of the minimal uncertainty, for equivalent isocratic and gradient plots, was similar. However, gradients were predicted generally with smaller uncertainties for any experimental design, and were less sensitive to mobile phase composition than isocratic predictions.

- A training design consisting of a set of isocratic experiments, gradually concentrated at low organic solvent contents (ISO1), was confirmed as the best for both isocratic and gradient predictions. Gradient designs at fixed gradient time and variable final organic solvent content (G1), and at fixed final organic solvent content and variable gradient time (G2), exhibited insufficient performance in most situations, being only acceptable for the slowest eluents and fastest solutes. The G3 design, which combined some features of designs G1 and G2, provided reasonable good performance for all probe compounds, only surpassed by design ISO1.

C.1.4. *Estimation of peak capacity based on peak simulation*

- Peak capacity is a key concept in chromatographic analysis, which refers to the maximal number of peaks that ideally are completely resolved in a given time window. In RPLC, chromatograms tend to have uneven peak distributions, with overlapped peaks and large gaps. Therefore, peak capacity is just a theoretical concept. In spite of this, it is useful to evaluate the possibilities of a column to get peak resolution, and has attracted great attention.

- Several authors have proposed algorithms to estimate the peak capacity in isocratic conditions. Neue also developed an algorithm to make estimations in gradient elution. However, these are limited to symmetrical peaks, the assumption of a constant theoretical plate number, simple linear gradients, absence of delays and extra-column effects. To overcome these limitations, an approach was developed based on the simulation of chromatograms containing a series of peaks for fictitious compounds, with the same type of behaviour as the target analytes when separated with a given column. The peaks for the fictitious compounds are generated based on the prediction of retention times and peak half-widths, and are arranged to fulfil the definition of peak capacity.

- The prediction of peaks is performed with models fitted from the information obtained for standards of a set of structurally-related compounds with varying polarity. The approach is illustrated using a set of 15 sulphonamides, analysed with three columns using isocratic elution, and linear and multi-linear gradients. The process begins by generating a high number of fictitious peaks with widths according to their retention times. The retention behaviour is obtained from the correlation of the parameters in the logarithmic-quadratic model with $P_{\mathrm{M}}^{\mathrm{N}}$ transformation, fitted with the standards, while the peak widths are predicted from the correlation of the peak half-widths with the retention times. Once the peaks are generated, they are moved up to be connected at the required height, usually assuming a peak width of 4σ free of overlapping.

- The approach was validated by observing the good agreement when the simulated chromatograms were overlapped with real chromatograms, for the mixture of sulphonamides at the same separation conditions. Also, the values of peak capacity were observed to agree with those estimated with classical equations. The approach based on simulation has the advantage, against previous approaches, of being applicable to a variety of situations where previous methods cannot be used, including complex multi-linear gradients and the presence of asymmetrical peaks.

- The proposed approach allowed the optimisation of the elution conditions, in a wide range of conditions, according to the predicted values of peak capacity. For this purpose, Pareto plots were built that included the predictions for isocratic conditions, and linear or multi-linear gradients (a solution is qualified as Pareto-optimal when a response cannot be improved without worsening another). As expected, isocratic separations presented the smallest peak capacity, while multi-linear gradients offered the highest values with minimal analysis time. A chromatographic system cannot provide peak capacity values outside the region limited by the isocratic trend and the upper boundary for gradient elution.

- However, for the set of sulphonamides, it was found that the separation conditions giving rise to the best resolution were far from those with maximal peak capacity. This means that an optimisation based on peak capacity becomes only meaningful for very complex samples. In samples where the number of compounds is relatively small, the specific resolution requirements of each peak should be attended instead.

C.1.5. Secondary chemical equilibria in reversed-phase liquid chromatography and interpretive search of optimal isocratic and gradient separations in micellar liquid chromatography in extended organic solvent domains

- Compounds in a wide range of structures and polarities can be analysed by RPLC. However, ionised organic compounds and inorganic anions or metals, which are highly polar, show little or no retention. Other analytes may show too low or too high retention. One way to solve these problems is the preparation of new stationary phases, but a simpler solution is the addition of reagents to the mobile phase, which gives rise to a variety of secondary equilibria with both stationary and mobile phases.

- The use of surfactants at concentrations where micelles are formed has become the most popular solution to modify the retention with additives in RPLC. This has given rise to a chromatographic mode called micellar liquid chromatography (MLC), which has been especially successful for the analysis of physiological samples, which do not require pre-treatment, since the proteins are solubilised in the presence of surfactant, and elute close to the void volume.

- Most reported procedures in MLC make use of the anionic surfactant sodium dodecyl sulphate (SDS). Since the elution strength of aqueous solutions of SDS is low for most solutes, a relatively small amount of organic solvent is added to decrease the retention. More recently, a chromatographic mode has been developed, where the concentration of organic solvent is increased to get sufficiently short times for highly retained compounds in the presence of surfactant. This RPLC mode has

been called high submicellar liquid chromatography (HSLC), since micelles are not formed, in spite of the relatively high concentration of surfactant.

- The procedures developed in MLC are usually implemented in the isocratic mode, since the general elution problem in RPLC (i.e., the exponential increase of retention at decreasing polarity) is less troublesome. However, gradient elution may be still useful to analyse, in shorter times, mixtures of compounds within a wide range of polarities. The analyses of physiological samples, in gradient elution, can be performed starting with a mobile phase that contains micelles, keeping the organic solvent contents low in order to provide better protection to the column against protein precipitation. Once the proteins are eliminated from the column, the elution strength can be increased using a positive gradient of organic solvent to reduce the retention times of highly retained compounds. This gives rise to the transition from the micellar to the submicellar modes.

- To appraise the convenience of the use of gradients against isocratic elution in MLC, considering an extended range of organic solvent, it was still necessary to develop an interpretive optimisation method for gradient elution, based on the accurate description of the retention. For this purpose, the screening of a set of eight basic compounds (β-adrenoceptor antagonists) in urine samples was considered, using direct injection, C8 and C18 columns, and aqueous solutions of SDS with added organic solvent.

- The performance of three organic solvents (acetonitrile, ethanol and 1-propanol) was checked at varying concentration of SDS. Acetonitrile offered complete resolution, but an excessive analysis time. Ethanol and 1-propanol offered acceptable analysis time, but the maximal resolution reached with ethanol was too low. Therefore, 1-propanol was selected for the analyses.

- The accuracy of nine retention models (some of them previously proposed for MLC and HSLC), using the concentrations of SDS and 1-propanol as variables, was compared. Equation (7.11) was selected, owing to its good prediction capability when the organic solvent domain was extended, with relative errors between 0.3 and 1.7%.

- When physiological samples are analysed making direct injection, besides the administered drugs, the chromatograms contain a prominent peak corresponding to an endogenous compound, which elutes at relatively short retention times. This compound, whose identity was unknown, should be modelled to be considered in the optimisation of the resolution. The information on its retention behaviour was obtained from the injection of urine, maintaining the concentration of 1-propanol low enough to avoid protein precipitation. Owing to the limited number of experiments available for this compound, Equation (7.5) was preferable to model its retention.

- A detailed study was carried out to know the capability of C8 and C18 columns in the analysis of the basic compounds with direct injection of urine samples, using isocratic mobile phases, or linear and multi-linear gradients. The optimisation of the elution conditions in the isocratic mode resulted in good resolution at a reasonable analysis time (around

25 min), for both columns, using high SDS concentration and organic solvent contents below 15%, which avoided the precipitation of the proteins in the sample. Good agreement was observed between predicted and experimental chromatograms for both columns.

- Single linear gradients yielded a significant reduction in the analysis time with regard to isocratic elution. The inclusion of an initial isocratic step at low organic solvent content was found detrimental to achieve good resolution. Baseline problems were observed with the C18 column, giving rise to deviations in the prediction of the signals. In contrast, the agreement between predicted and experimental chromatograms was excellent for the C8 column. This behaviour can be explained by the higher capability of the C18 column to adsorb surfactant, with regard to the C8 column, which is gradually desorbed by the organic solvent along the gradient.

- In general, the implementation of multi-linear gradients with solutions containing surfactant and sudden changes in the slopes give rise to important baseline disturbance, particularly with the C18 column. For the C8 column, multi-linear gradients were able to reduce the analysis time significantly with good resolution, and good agreement between predicted and experimental chromatograms. Therefore, the use of linear gradients with the C8 column is preferable for these analyses.

C.2. Improving the separation performance for chromatographic fingerprints

As commented above, the search of the best separation conditions in liquid chromatography can be carried out using the information obtained from standards of the analytes. However, obtaining useful information for samples with a large number of compounds is still a challenge. The larger difficulty corresponds to samples for which no prior information on the chemical composition is available, at least for some compounds. There is also the possibility that standards of the analytes, needed to predict the optimal separation conditions, with conventional interpretive strategies, be not available.

Disregarding the identity of the compounds giving rise to peaks in a chromatogram is known or unknown, their mutual separation should be as large as possible for both qualitative and quantitative purposes. An extreme case is found in chromatographic fingerprints, where the relative peak distribution and magnitude is the relevant feature. In these samples, better resolution could offer more informative chromatograms. Part II includes proposals for improving signal processing for complex chromatograms, the estimation of the resolution for fingerprints of medicinal herbs in one-dimensional liquid chromatography, and the optimisation of the separation of polyphenolic compounds in fingerprints of olive leaf and pulp extracts, using two-dimensional liquid chromatography.

C.2.1. Assisted baseline subtraction in complex chromatograms using the BEADS algorithm

- Data processing of the signals in chromatograms of complex samples may constitute a bottleneck to obtain significant information. An important problem that should be addressed before treating the signals is the subtraction of the baseline, which can be notably irregular and, ideally, should be make without supervision.

- An interesting tool, recently developed for baseline subtraction, is the BEADS algorithm, which makes a full decomposition of the chromatograms by using high pass frequency filters that separate the pure signals of the compounds (described as sparse signals), from the baseline (a low frequency signal), and the noise (the high frequency contribution).

- However, the algorithm initially reported needs a careful selection of the working parameters to process properly the signals, especially the cutoff frequency which is the most critical parameter. This should be made using trial and error, which makes the process too slow and unstable. On the other hand, the application of the original BEADS to chromatograms containing peaks with extremely different magnitude gives rise to deformations in the baseline, as ripples under the main peaks associated to the large differences in scale between major and trace components. Also, the presence of negative signals affects severely the subtraction of the baseline.

- Diverse modifications are proposed in this PhD. Project to improve the performance and reliability of BEADS, which was called assisted BEADS, since the selection of the optimal working parameters is simplified based on the use of auxiliary autocorrelation plots. An important characteristic of the assisted BEADS is the logarithmic transformation of the raw signals, which eliminates the irregularities observed in the baseline under the major components. The logarithmic transformation reduces the weight of these peaks, resulting in estimated baselines with a general smooth trend. By making the logarithmic transformation of the signal, stepped plots were obtained for each working parameter, whose optimal value was located close to the inflection point.

- The assisted BEADS can be quite easily adapted to any kind of sample, provided a proper baseline subtraction for all assayed samples, independently of their complexity. It reduces the subjectivity in the selection of the working parameters, and provides always reliable results. The selection of the optimal cutoff frequency, which constitutes the boundary between the baseline and the rest of contributions (sparse signals and noise), was less critical compared to the original BEADS. The effects of sporadic negative signals, after baseline subtraction, were corrected by implementing an iterative process.

- It should be noted that BEADS makes a global fitting of the baseline. This implies losing details in particular regions of the chromatogram, with regard to the fitting of a local baseline (which only considers the surroundings of a peak). However, the magnitude of the errors obtained with the assisted BEADS were acceptable. It is also noteworthy that the

application of the assisted BEADS is not only limited to chromatographic signals.

C.2.2. *Study of the performance of a resolution criterion to characterise complex chromatograms with unknowns or without standards*

- The objective of interpretive optimisation strategies is the search of experimental conditions that yield the best resolution, based on predictions of retention times and peak profiles for the target analytes. With these values, simulated chromatograms are built. Most resolution criteria used to measure the separation performance need standards to fit the models from which the predictions are made. However, for some samples, there are no standards available. Therefore, it was thought that a global resolution function, valid for all situations (with or without standards), was needed.

- The proposed function was based on the measurement of the peak prominence, which is the area fraction exceeding the line that joins the valleys that delimit each peak. The peak prominence criterion was validated by comparison with the peak purity criterion, which measures the peak area free of overlapping and provides reliable estimations of chromatographic resolution. The peak purity criterion requires a comprehensive knowledge of the individual signals for each analyte, at any condition in the experimental design, which is only accessible through simulation based on the information obtained from standards. In contrast, peak prominence can be measured directly from the signals in a real chromatogram, without any prior knowledge of the compounds in the sample.

- To compare the peak prominence and peak purity criteria, the chromatograms for a set of amino acids, derivatised with o-phthalaldehyde and N-acetylcysteine, were obtained for isocratic and gradient conditions. Using the data obtained from standards for 10 isocratic conditions, retention and half-width models were built. With these models, the separation in around 1100 linear and multi-linear gradients was predicted. The amino acid derivatives could only be resolved at high analysis times, even using multi-isocratic and multi-linear gradients. When the analysis time was reduced, significant overlapping was obtained for several compounds. All this behaviour gave rise to interesting cases of study for the evaluation of the resolution functions.

- The comparison study was carried out with the assistance of Pareto optimality plots. The plots were drawn for both peak prominence and peak purity criteria, considering the two opposite quality measurements to be enhanced: chromatographic resolution and analysis time. Plots were built for several simulated situations, gradually closer to reality: signals of different magnitude, inclusion of instrumental noise, real baselines, and presence of unknown compounds.

- Three functions were studied as candidates to measure the global peak prominence (Equations (9.3) to (9.5)), which were compared with the global peak purity expressed as the sum of the individual values (Equation (9.6)). Among the assayed functions for the peak prominence, the sum of normalised individual resolutions (Equation (9.4)) appeared as the best, since the projection of the optimal

gradients for the Pareto front for the peak prominence agreed with the Pareto front obtained for the sum of peak purities.

- The best global prominence function was successfully applied to evaluate the resolution of chromatographic fingerprints for extracts of herbal medicines, which contained a high number of constituents whose identity was unknown. The proposed resolution criterion has the advantage of being evaluable directly from the experimental chromatograms, without going through steps of modelling, prediction and simulation, using information obtained from standards, as required for the peak purity.

C.2.3. *Classification of olive leaf and pulp extracts by comprehensive two-dimensional liquid chromatography of polyphenolic fingerprints*

- Olive leaf and pulp extracts are complex mixtures of hundreds of different compounds. Among these, polyphenols have attracted much attention, due to their healthy benefits. The analyses of polyphenols are usually performed by liquid chromatography using a single column. However, the complexity of the samples makes full resolution not possible. Thus, the possibility of using comprehensive two-dimensional liquid chromatography (LC×LC) to carry out the analyses was investigated. The LC×LC approach combines two columns with different separation mechanisms to get maximal resolution in the analysis of compounds in a complex sample. This gives rise to chromatograms with two dimensions.

- The separation performance of several columns (with different stationary phase, length, internal diameter, and pore and particle sizes) was evaluated to get the maximal number of visible peaks (i.e., peak capacity) in the analysis of polyphenolic fingerprints, using diverse elution conditions. Along the study, three stationary phases were considered in the first dimension (conventional C18, and C18 with phenyl and pentafluorophenyl groups), and five in the second dimension (C18, amide, cyano, phenyl or pentafluorophenyl). The separation in the first dimension was made with methanol-water gradients, whereas acetonitrile-water gradients were used in the second dimension.

- The optimisation of the best column combination started using conventional C18 and cyano columns in the first and second dimensions, respectively, and a conventional gradient, which resulted in a poor number of visible peaks (29 for the olive leaf extracts). Change of the cyano column by a conventional pentafluorophenyl column, together with the reduction in the modulation time (time of collection of the effluent from the first dimension before being injected in the second dimension) increased the number of peaks to 73. Finally, by using submicro columns (1.8 μm C18 in the first dimension and 2.6 μm pentafluorophenyl in the second), changing the column order (pentafluorophenyl in the first dimension), and applying a shifted gradient in the second dimension, more informative fingerprints with 112 and 109 visible peaks could be detected for the olive leaf and pulp extracts, respectively.

- The optimised LC×LC method was successfully applied to characterise the presence of 26 common polyphenols in the olive leaf extracts and 29 in the pulp extracts. The peak volume ratios of the peaks for these compounds (less sensitive to the extraction process than the peak volumes) were selected to develop a linear discriminant analysis (LDA) model, able to distinguish the origin of the extracts.

- Three-dimensional plots were drawn with the scores obtained from the information provided by the LC×LC chromatograms of the olive leaf and pulp extracts, according to the three first discriminant functions. The plots showed that all samples in a given class were grouped in compact clusters. The resulting LDA models allowed the correct classification of seven cultivars of olive leaves and pulps of different genetic origin from several Spanish regions, with an excellent separation among categories and a high level of confidence. This demonstrated that polyphenolic profiles are characteristic of each cultivar.

CONTRIBUTION OF THE PhD WORK
TO PUBLICATIONS

The chapters in this PhD. work correspond to the following publications, listed according to the publication date. The percentage of contribution of José Antonio Navarro Huerta as PhD. student is indicated with each article.

- <u>José Antonio Navarro Huerta</u>, José Ramón Torres Lapasió, Sergio López Ureña, María Celia García Álvarez-Coque

  Assisted baseline subtraction in complex chromatograms using the BEADS algorithm.

  Journal of Chromatography A 1507 (2017) 1–10 (Chapter 8).

  Contribution: 70% José Antonio Navarro Huerta, 30% Sergio López Ureña

- María Celia García Álvarez-Coque, José Ramón Torres Lapasió, <u>José Antonio Navarro Huerta</u>

  Secondary equilibria in reversed-phase liquid chromatography

  Liquid Chromatography: Fundamentals and Instrumentation (edited by S. Fanali, P.R. Haddad, C.F. Poole and M. Riekkola), Elsevier, Amsterdam, Netherlands, 2nd ed., Vol. 1, 2017, pp. 125–146 (chapter written by invitation) (Chapter 6).

  Contribution: 100% José Antonio Navarro Huerta

- Guillermo Ramis Ramos, María Celia García Álvarez-Coque, <u>José Antonio Navarro Huerta,</u>
  Solvent selection in liquid chromatography
  Liquid Chromatography: Fundamentals and Instrumentation (edited by S. Fanali, P.R. Haddad, C.F. Poole and M. Riekkola), Elsevier, Amsterdam, Netherlands, 2nd ed., Vol. 1, 2017, pp. 343–373 (chapter written by invitation) (Chapter 1).
  Contribution: 100% José Antonio Navarro Huerta

- <u>José Antonio Navarro Huerta</u>, Tamara Álvarez Segura, José Ramón Torres Lapasió, María Celia García Álvarez-Coque
  Study of the performance of a resolution criterion to characterise complex chromatograms with unknowns or without standards.
  Analytical Methods 9 (2017) 4293–4303 (Chapter 9).
  Contribution: 50% José Antonio Navarro Huerta, 50% Tamara Álvarez Segura

- <u>José Antonio Navarro Huerta</u>, José Ramón Torres Lapasió, María Celia García Álvarez-Coque
  Estimation of peak capacity based on peak simulation.
  Journal of Chromatography A 1574 (2018) 101−113 (Chapter 5).
  Contribution: 100% José Antonio Navarro Huerta

- José Antonio Navarro Huerta, Adrián Gisbert Alonso, José Ramón Torres Lapasió, María Celia García Álvarez-Coque

  Benefits of solvent concentration pulses in retention time modelling of líquid chromatography.

  Journal of Chromatography A 1597 (2019) 76–88 (Chapter 3).

  Contribution: 50% José Antonio Navarro Huerta, 50% Adrián Gisbert Alonso

- José Antonio Navarro Huerta, Enrique Javier Carrasco Correa, José Ramón Torres Lapasió, José Manuel Herrero Martínez, María Celia García Álvarez-Coque

  Modelling retention and peak shape of small polar solutes analysed by nano-HPLC using methacrylate-based monolithic columns.

  Analytica Chimica Acta 1086 (2019) 142–155 (Chapter 2).

  Contribution: 100% José Antonio Navarro Huerta

- José Antonio Navarro Huerta, Ángel Gamaliel Vargas García, José Ramón Torres Lapasió, María Celia García Álvarez-Coque

  Interpretive search of optimal isocratic and gradient separations in micellar liquid chromatography in extended organic solvent domains.

  Journal of Chromatography A 1616 (2020) 460784 (Chapter 7).

  Contribution: 100% José Antonio Navarro Huerta

- María Vergara Barberán, <u>José Antonio Navarro Huerta</u>, José Ramón Torres Lapasió, Ernesto Francisco Simó Alfonso, María Celia García Álvarez-Coque

  Classification of olive leaves and pulp extracts by comprehensive two-dimensional liquid chromatography of polyphenolic fingerprints.

  Food Chemistry 320 (2020) 126630 (Chapter 10).

  Contribution: 100% José Antonio Navarro Huerta

- <u>José Antonio Navarro Huerta</u>, Adrián Gisbert Alonso, José Ramón Torres Lapasió, María Celia García Álvarez-Coque

  Testing experimental designs in liquid chromatography (I): Development and validation of a method for the comprehensive inspection of experimental designs.

  Journal of Chromatography A 1624 (2020) 461180 (Chapter 4).

  Contribution: 50% José Antonio Navarro Huerta, 50% Adrián Gisbert Alonso

Both supervisors (or at least one of them) of this PhD. work (María Celia Álvarez-Coque and José Ramón Torres Lapasió) appear as co-authors. They are also main researchers of the projects which have funded the research. José Ramón Torres Lapasió has proposed the main lines of several works and has developed the computer programs used in the data processing included in the different articles.

The articles in Chapters 2 and 10 were carried out in collaboration with the CLECEM research group, from the Department of Analytical Chemistry at the University of Valencia. In Chapter 2, José Manuel Herrero Martínez contributed in the development of the capillary monolithic columns, and the

analytical instrumentation to carry out the experiments. In Chapter 10, Ernesto Francisco Simó Alfonso contributed providing the analysed samples and chemometrics analysis of the results.

Tamara Álvarez Segura participated in the initial period of the PhD. Work, training and supervising the experimental work (preparation of solutions, experimental technique and handling of instruments). Sergio Lopez Ureña, from the Department of Mathematics at the University of Valencia, collaborated with the mathematical treatment and data processing. Enrique Javier Carrasco Correa (postdoctoral researcher) participated in the preparation of the columns and handling of the instrument. María Vergara Barberán (postdoctoral researcher) participated in the sample preparation and development of the analytical method. In addition, some research work contributed to the training of the degree student Angel Gamaliel Vargas García, and Master student Adrián Gisbert Alonso.