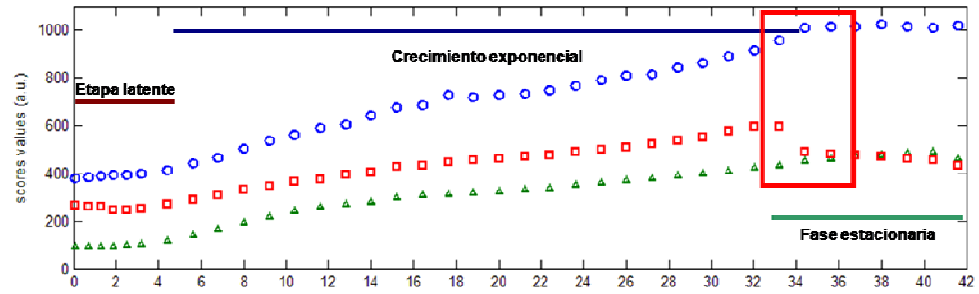
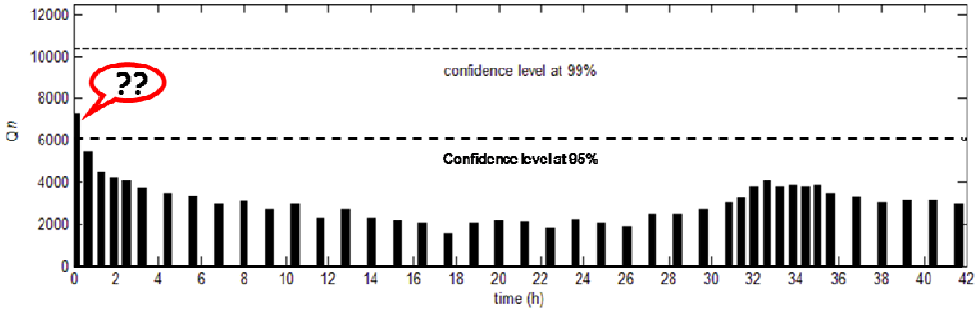




Capítulo 5

RESULTADOS

PARTE I:

DETERMINACIÓN SIMULTÁNEA DE HIPOXANTINA, XANTINA Y ÁCIDO ÚRICO EN ORINA HUMANA.

5.1. INTRODUCCIÓN

A lo largo de esta memoria se ha resaltado el interés en desarrollar nuevos procedimientos de calibración multivariable para la determinación simultánea de varios analitos a través del uso de enzimas de grupo utilizando metodologías que requieren una instrumentación relativamente sencilla y poco tiempo de análisis **[Petterson 1997, Crouch 2000, Coello 2000, Ni 2004]**.

El uso de enzimas que muestran selectividad de grupo se convierten así en una alternativa para la resolución cinética de compuestos similares con grupos funcionales comunes catalizables por la misma enzima. Esta potenciabilidad ya ha sido mostrada en trabajos anteriores, como por ejemplo en la determinación de mezclas de metanol y etanol usando el enzima alcohol oxidasa **[Blanco 1999]**.

El sistema estudiado en esta primera parte está formado por dos reacciones consecutivas ($A \rightarrow B \rightarrow C$) catalizadas por la misma enzima, donde A y B son los analitos

y C, el producto final de la reacción, actúa como inhibidor de la reacción **[Escribano 1988]**.

Sin embargo, a pesar de la complejidad del sistema enzimático, se dispone de una gran cantidad de información debido a que todos los componentes absorben en la región espectral del Ultravioleta-Visible, aunque con un gran solapamiento espectral. Tanto la información espectral como la información cinética contribuyen a discriminar la diferente respuesta de cada analito y resolver las mezclas.

5.2. SISTEMA ENZIMÁTICO

Los niveles de oxipurinas (hipoxantina y xantina) y ácido úrico en diferentes fluídos biológicos (sangre, plasma y orina) son un indicativo de ciertos estados patológicos como, por ejemplo, xantinuria, gota, fallos renales, toxaemia en mujeres embarazadas, entre otras **[Tawa 1981]**. La similitud en sus estructuras químicas y la falta de selectividad espectral (figura 5.1) han obligado al uso de técnicas analíticas de separación, como la Cromatografía Líquida de Alta Resolución (HPLC) **[Putterman 1979, Czauderna 1997, Di Pietro 2001]** o Electroforésis Capilar de Alta Resolución (HPCE) **[Chen 2002]** para su determinación simultánea y diferentes métodos de análisis electroquímico **[Zen 2002, Pei 2000]**.

El enzima xantina oxidasa (xantina:oxigeno oxidoreductasa, E.C. 1.17.3.2) ha sido usada como enzima de grupo para desarrollar metodologías de análisis para la determinación de mezclas complejas de analitos **[Foppoli 1997]** y, también, para la determinación de hipoxantina y xantina **[Tawa 1981, Pei 2000, Carsol 1998]**. Este enzima es uno de los más complejos dentro del grupo de las flavoproteínas, siendo un enzima de baja especificidad, catalizando la oxidación de muchos analitos, tales como purinas, pirimidinas, pteridinas y aldehidos **[Escribano 1988, Massey 1969, Jeżwska 1973]**.

Se sabe que la hipoxantina (H) y la xantina (X) son oxidados a ácido úrico (U) en una reacción catalizada por la xantin oxidasa (XO) a través del siguiente esquema:

$$H + O_2 \xrightarrow{XO} X + H_2O_2 \qquad (5.1)$$

$$X + O_2 \xrightarrow{XO} U + H_2O_2 \qquad (5.2)$$

El proceso global (figura 5.1) está influenciado por factores como el pH, la composición del tampón usado y su concentración, la concentración de enzima, la temperatura de trabajo y la presencia inicial de ácido úrico. Se sabe que el ácido úrico actúa de inhibidor de la reacción incluso a niveles de saturación de oxígeno **[Escribano 1988]**. Además, existen evidencias experimentales de que el ácido úrico se oxida a allantoin, analito no absorbente en rango espectral del UV-Visible, en soluciones alcalinas **[Clarck 1979, Lindsey 1981]**.

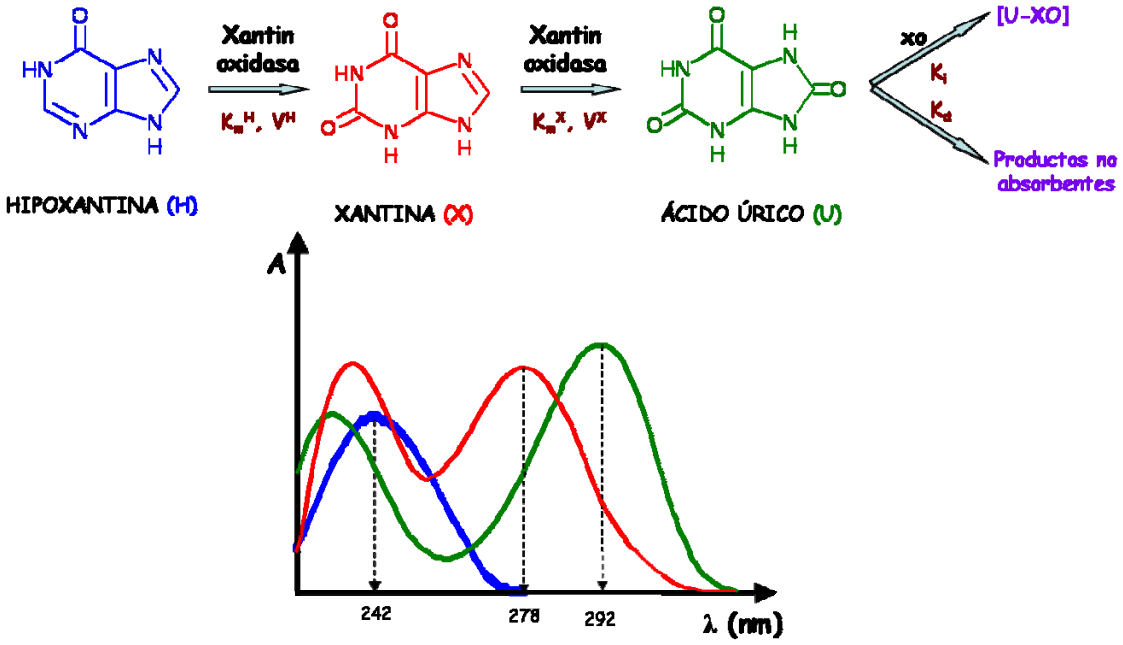


Figura 5.1: Proceso enzimático en dos etapas de la catálisis de la hipoxantina (H) por la xantin oxidasa (XO) y espectros UV-Vis de la hipoxantina (H, azul), xantina (X, rojo) y el ácido úrico (U, verde).

5.3. DETERMINACIÓN SIMULTÁNEA DE HIPOXANTINA Y XANTINA EN ORINA HUMANA MEDIANTE 3W-PLS. COMPARACIÓN CON TÉCNICAS DE DESDOBLAMIENTO *unfolding*-PLS.

5.3.1. OBJETIVOS

En este primer trabajo se muestra con detalle la potenciabilidad analítica del uso combinado de los métodos de calibración multivariable multidimensional con enzimas no selectivos, sin necesidad del conocimiento del modelo enzimático por el cual se rige el sistema. Para ello, se propondrá un método sencillo y preciso para la determinación simultánea de xantina e hipoxantina en orina humana usando la xantin oxidasa como enzima de grupo y metodologías multidimensionales (3W-PCA y 3W-PLS). Los resultados se compararán con las metodologías PCA y PLS aplicados a las matrices desdobladas (*unfolding*-PCA Y *unfolding*-PLS).

5.3.2. DISEÑO DEL CONJUNTO DE CALIBRACIÓN, VALIDACIÓN Y MUESTRAS DE ORINA

Las metodologías de calibración multivariable PLS requieren un número elevado de muestras para realizar modelos robustos de calibración. A parte, se requiere un conjunto de muestras externo para la validación del modelo de calibración.

Aunque el objetivo es la determinación de hipoxantina y xantina, el efecto inhibidor del ácido úrico obligó a tener en consideración a este analito a la hora de preparar los diferentes grupos de muestras (calibración, validación y orina dopada). En este aspecto, se realizó un diseño experimental de calibración con los tres analitos a 5 niveles de concentración (figura 5.2). Se prepararon 23 mezclas para la realización del modelo de calibrado, 15 mezclas para la validación de los modelos y 15 muestras de orina dopada con diferente concentración de los tres analitos. Los rangos de

concentración para los tres analitos fueron de 0.0–3.0 mmol de hipoxantina por litro de orina, 0.0–1.5 mmol de xantina por litro de orina y 0.0–3.0 mmol de ácido úrico por litro de orina (tabla 5.1).

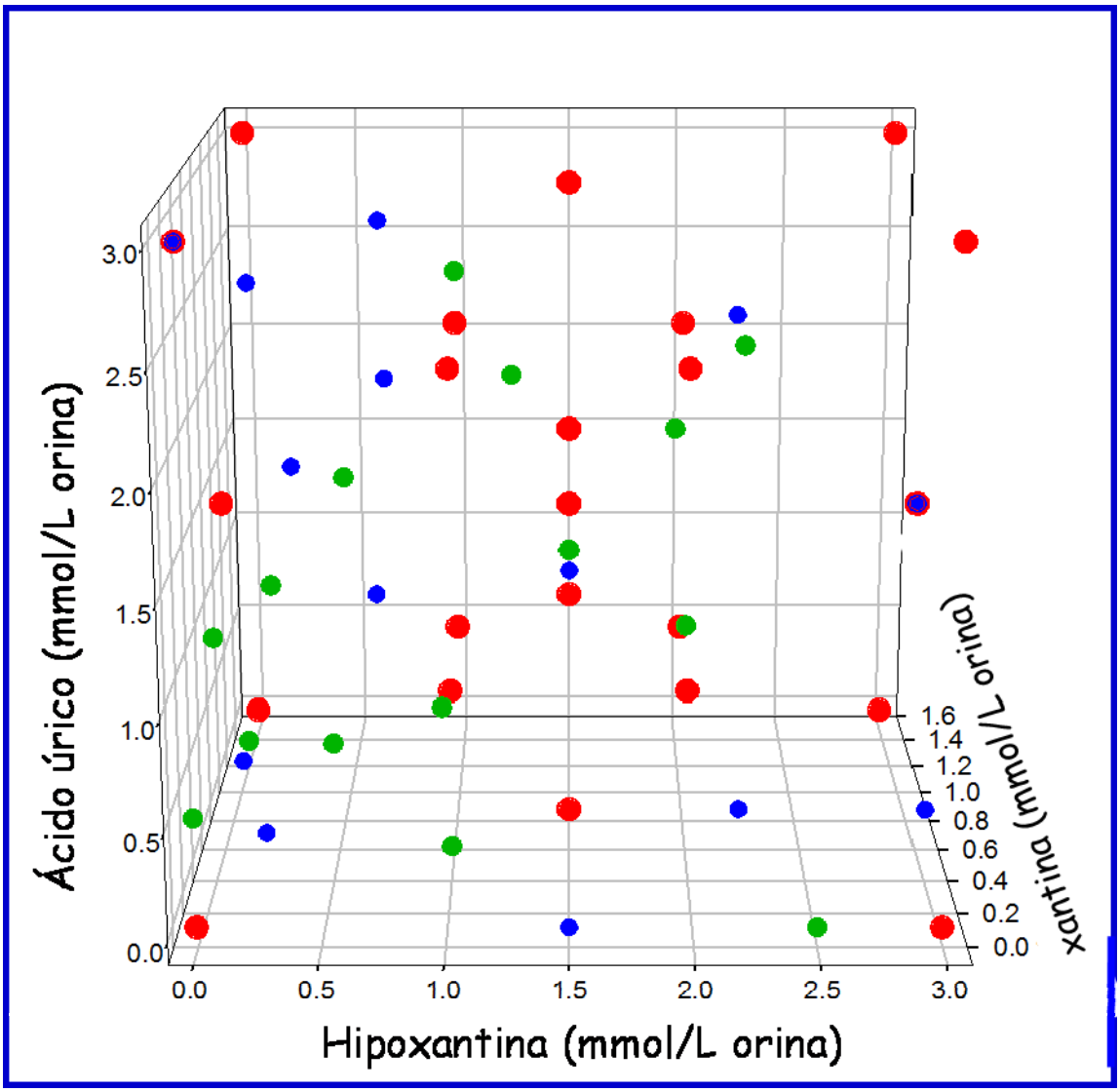


Figura 5.2: Diseño experimental de las concentraciones de H, X y U para el conjunto de calibración (rojo), validación (azul) y muestras de orina dopadas (verde)

Tabla 5.1: Composición de las muestras de calibración, validación y orina dopada (μmol mL^{-1} de orina)

CALIBRACIÓN			
mezclas	hipoxantina	xantina	Ácido úrico
c01	0	0	0
c02	3015	0	0
c03	0	1515	0
c04	0	0	3027
c05	3015	0	3027
c06	3015	1515	0
c07	0	1515	3027
c08	3015	1515	3027
c09	503	404	757
c10	503	404	2270
c11	503	1111	757
c12	503	1111	2270
c13	2262	404	757
c14	2262	404	2270
c15	2262	1111	757
c16	2262	1111	2270
c17	1508	757	0
c18	1508	757	3027
c19	1508	0	1513
c20	1508	1515	1513
c21	0	757	1513
c22	3015	757	1513
c23	1508	757	1513

VALIDACIÓN				ORINA DOPADA			
mezclas	hipoxantina	xantina	Ácido úrico	mezclas	hipoxantina	xantina	Ácido úrico
v01	1508	0	0	s01	2513	0	0
v02	0	1111	0	s02	0	1262	0
v03	0	0	3027	s03	0	0	504
v04	2262	757	0	s04	0	505	1009
v05	754	0	1513	s05	2262	0	1009
v06	0	1515	2270	s06	3015	505	0
v07	251	1111	1513	s07	2010	606	1009
v08	754	252	3027	s08	1256	1010	2018
v09	3015	404	252	s09	503	1010	1513
v10	754	404	2270	s10	2262	757	2270
v11	1508	757	757	s11	1005	1010	2522
v12	251	252	252	s12	503	505	504
v13	2262	1111	2270	s13	1508	1262	1009
v14	1508	1515	757	s14	2010	1515	1513
v15	3015	757	1513	s15	251	505	1261

La composición de las 15 mezclas de validación y de orina estaba comprendida entre los rangos de concentración definidos por el conjunto de calibración. Se registró una réplica de cada mezcla para verificar la repetibilidad de la metodología.

5.3.3. PROCESADO DE LOS DATOS

Los espectros UV-Vis para cada muestra *i* se registraron cada *j* tiempos a *k* diferentes longitudes de onda para construir la matriz tridimensional **M** (*i* x *j* x *k*). Esta matriz se desdobló para obtener la clásica matriz bidimensional de tal manera que cada fila contenía los espectros para una muestra unidos secuencialmente (capítulo 3, figura 3.3).

Teniendo en cuenta la relación intrínseca entre la hipoxantina y la xantina, se realizaron modelos 3W-PLS2 y *unfold*-PLS2 y se compararon con modelos 3W-PLS1 y *unfold*-PLS1. Los modelos de calibración fueron validados internamente por validación cruzada (*cross-validation*) *leave-one-out* segmentada por replicados de la misma mezcla.

El número de factores PLS escogido para cada modelo de calibración fue seleccionado como el mínimo número de factores que ofreciesen un valor de PRESS (*Prediction Error Sum of Squares*) no significativamente diferente del mínimo (ecuación 5.3)

$$PRESS = \sum_{i=1}^{n} (\hat{c}_i - c_i)^2 \qquad (5.3)$$

Donde *n* es el número de muestras, $\hat{c}_i$ es la concentración calculada de la muestra *i* y c_i es el valor de referencia.

Para evaluar la precisión y la capacidad predictiva de los modelos se contrastaron los valores de error estándar de calibración y predicción, SEC y SEP, respectivamente (*Standard Error of Calibration / Prediction*) (ecuaciones 5.4 y 5.5, respectivamente)

teniendo en cuenta el *bias* del modelo (ecuación 5.6). La Desviación Estándar entre Replicados, SDBR (*Standard Deviation Between Replicates*) (ecuación 5.7) se calculó entre las diferentes concentraciones calculadas entre los dos replicados de cada muestra. El SDBR es una medida de la repetibilidad experimental y puede ser considerada como la precisión "objetivo" de la máxima precisión esperable bajo unas determinadas condiciones experimentales. También se puede decir que valores de SEP claramente superiores a valores de SDBR son indicativos de falta de ajuste del modelo, por lo que se tendría que buscar un modelo más adecuado.

$$SEC = \sqrt{\frac{\sum_{i=1}^{n}(\hat{c} - c_i - bias)^2}{n - f - 1}} \tag{5.4}$$

$$SEP = \sqrt{\frac{\sum_{i=1}^{n}(\hat{c} - c_i - bias)^2}{n - 1}} \tag{5.5}$$

$$bias = \frac{\sum_{i=1}^{n}(\hat{c} - c_i)}{n} \tag{5.6}$$

$$SDBR = \sqrt{\frac{\sum_{i=1}^{n}(\hat{c}_{i,1} - \hat{c}_{i,2})^2}{2n}} \tag{5.7}$$

En estos parámetros de evaluación del modelo, c_i es la concentración de referencia y $\hat{c}_i$ es la concentración calculada. n es el número de muestras utilizadas para la calibración o la predicción. f es el número de factores PLS.

5.3.4. RESULTADOS Y DISCUSIÓN

5.3.4.1. Comparación entre las muestras de calibración, validación y las muestras de orina

Se realizó un modelo 3W-PCA a la matriz conjunta, que contenía las 23 muestras de calibración, 15 muestras de validación y 15 muestras de orina, más sus correspondientes replicados (modelo A en la tabla 5.2). El modelo con 4 factores explica más de un 99% de varianza en cada uno de los modos descompuestos (modo de concentraciones, modo cinético y modo espectral). En la misma tabla se aprecia que la mayor variabilidad viene explicada por el primer factor.

Tabla 5.2: Porcentaje de varianza explicada por el análisis 3W-PCA de cada uno de los modos. El modelo A hace referencia al modelo con el rango de longitudes de onda completo. El modelo B hace referencia al modelo sin las 15 primeras longitudes de onda.

	concentración		**tiempo**		**Longitud de onda**	
	modelo A	modelo B	modelo A	modelo B	modelo A	modelo B
PC1	90.24	91.20	93.26	93.19	88.71	91.42
PC2	6.12	7.97	5.91	5.98	8.31	8.18
PC3	**2.99**	**0.58**	**0.66**	**0.67**	**2.57**	**0.39**
PC4	0.42	0.10	0.12	0.12	0.38	0.01

La figura 5.3 muestra los *scores* obtenidos para los tres primeros factores en los tres modos. Observando el *score* de del tercer factor en el modo de concentraciones (fig. 5.3a) se puede observar cómo las concentraciones de las muestras de orina difieren de las concentraciones de las muestras que no tienen orina. Observando la figura de *scores* para el modo espectral (fig. 5.3c) se aprecia una diferencia significativa en las primeras 15 longitudes de onda en el *score* del tercer factor. Sin embargo, no existe ninguna desviación significativa en el *score* del tercer factor en el modo cinético (fig. 5.3b).

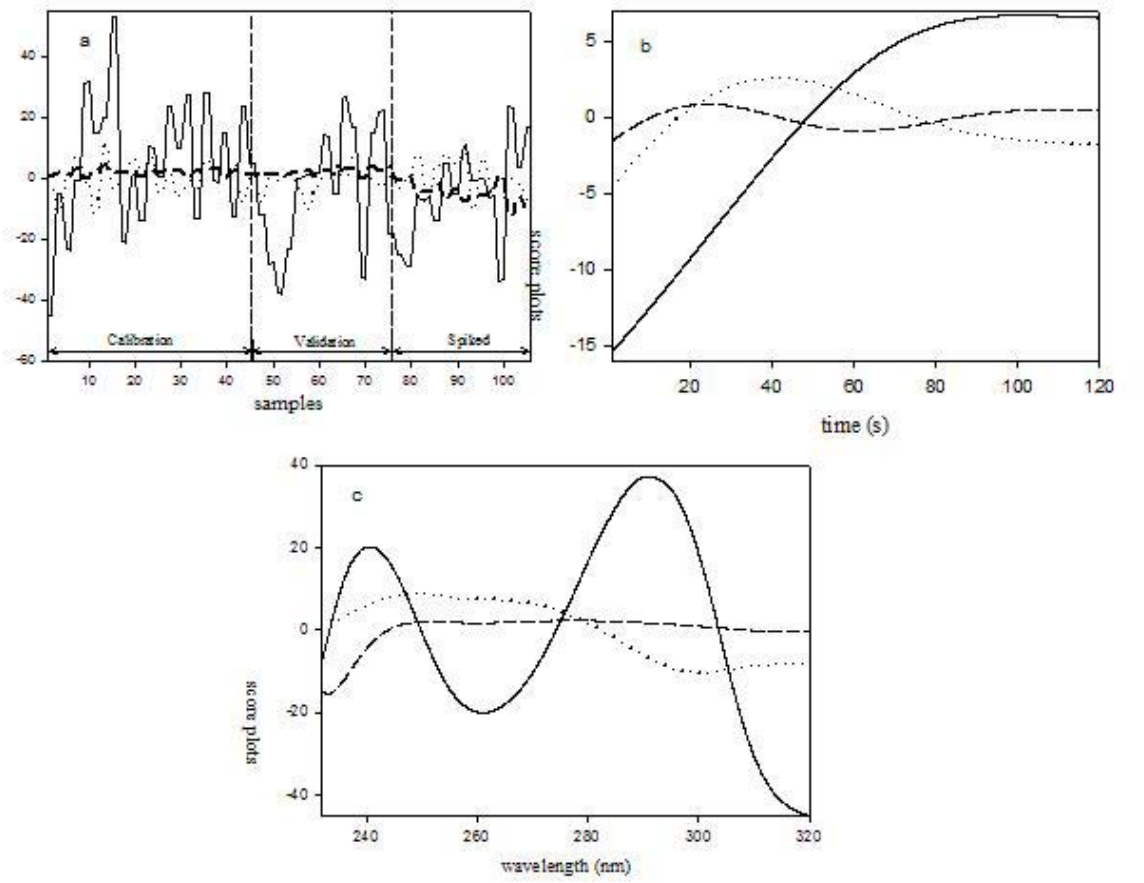


Figura 5.3: Gráficos de *scores* obtenidos para los modos de a) concentración, b) cinético y c) espectral. Primer factor, línea continua. Segundo factor, línea punteada. Tercer factor, rayas discontinuas

Esto concuerda con la varianza explicada por el tercer factor en los modos de concentraciones y espectral, alrededor del 3%, mientras en el modo cinético sólo explica un 0.66% de varianza (tabla 5.2). La figura de *scores* del primer factor frente al tercer factor para las muestras (figura 5.4) hace evidente que el tercer factor diferencia claramente entre las muestras que contienen orina y las que no la contienen (muestras de laboratorio).

Esta diferencia afecta al modo de concentraciones y al modo espectral, pero no afecta a la cinética del sistema. La figura 5.5 muestra que esa diferencia se encuentra en las primeras 15 longitudes de onda. Esta región espectral es la zona donde la orina presenta su máximo de absorbancia.

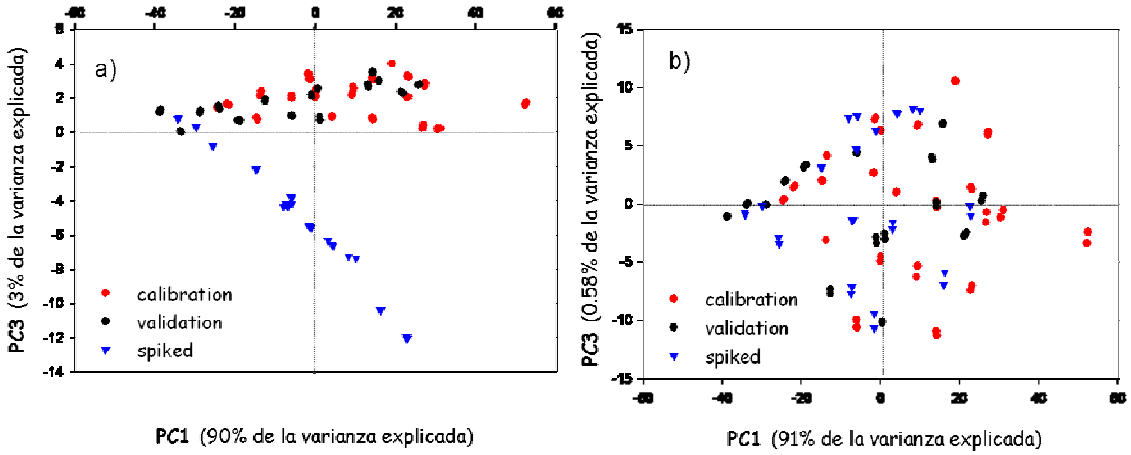


Figura 5.4: Gráfico de *scores* del primer factor frente al tercer factor para el modo de concentración (calibración en rojo, validación en negro y muestras de orina en azul). a) modelo A. b) modelo B.

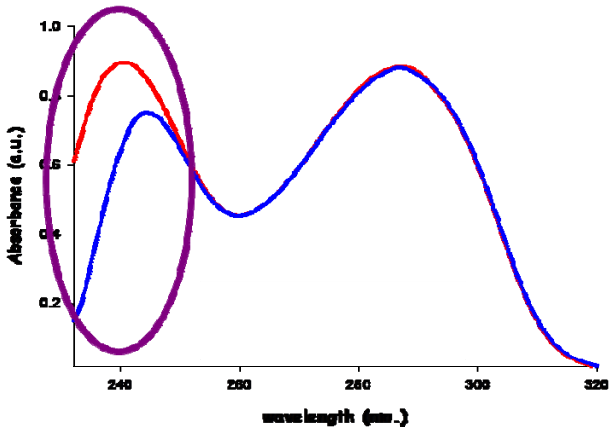


Figura 5.5: Diferencia espectral entre muestras que contenían orina (rojo) y muestras sin orina (azul)

Después de la eliminación de las primeras 15 longitudes de onda (modelo B), se repite el análisis 3W-PCA. En la figura 5.4b se observa cómo ahora las muestras con orina y las muestras de laboratorio son indistinguibles. Ahora el tercer factor explica alrededor de un 0.67% de varianza para los tres modos (tabla 5.2).

El primer factor en cada uno de los tres modos del nuevo modelo explica el comportamiento del ácido úrico. Este hecho no es sorprendente debido a que, de acuerdo

con el sistema cinético, el hecho predominante es la formación de la banda de absorción del ácido úrico a 290 nm (figura 5.3c).

5.3.4.2. Determinación de hipoxantina y xantina

Los resultados generales para la determinación de la hipoxantina y la xantina se muestran en las tablas 5.3 y 5.4, respectivamente. Las muestras de validación y muestras con orina presentan valores similares de SEP y SDBR, con valores relativamente bajos, por lo que se concluye que no existen diferencias significativas entre estos dos grupos. Este hecho es un indicativo de una buena corrección del efecto matriz de la orina y que los modelos ofrecen una buena capacidad predictiva y buena repetitividad.

Tabla 5.3: Resultados para la determinación de hipoxantina. [a]Los resultados están expresados como μmol L^{-1} de orina.

		PLS1	PLS2	3W-PLS1	3W-PLS2
Calibración	**Num. PC**	4	7	4	7
	SEC[a]	2.07	1.28	1.94	0.96
	SDBR[a]	0.39	0.63	0.49	0.62
Validación	**SEP[a]**	1.60	0.68	2.04	0.61
	SDBR[a]	0.34	0.47	0.42	0.44
Orina	**SEP[a]**	2.26	1.41	2.20	1.03
	SDBR[a]	0.98	1.06	0.97	0.96

Los resultados para la hipoxantina son ligeramente mejores que para la xantina. Los resultados para la xantina indican una ligera no linealidad en el sistema (los valores de SEP son alrededor del doble de los valores de SDBR), pero la capacidad predictiva de los modelos es de suficiente calidad para cuantificar simultáneamente los dos analitos con buenos resultados.

Tabla 5.4: Resultados para la determinación de xantina. [a]Los resultados están expresados como μmol L^{-1} de orina.

		PLS1	PLS2	3W-PLS1	3W-PLS2
Calibración	**Num. PC**	7	7	7	7
	SEC[a]	1.54	1.6	1.08	1.09
	SDBR[a]	0.50	0.48	0.61	0.54
Validación	**SEP[a]**	0.95	1.30	0.94	0.96
	SDBR[a]	0.59	0.60	0.65	0.54
Orina	**SEP[a]**	1.49	1.54	1.31	1.26
	SDBR[a]	0.73	0.60	0.71	0.56

No existen diferencias significativas en la cuantificación de hipoxantina y xantina entre muestras conteniendo diferente concentración inicial de ácido úrico, indicando la mínima influencia del ácido úrico o que esta influencia se ha modelado correctamente en el conjunto de calibración. Este hecho refleja que, a pesar de que la principal fuente de variabilidad en el sistema es el ácido úrico, los modelos 3W-PLS descomponen la matriz **M** teniendo en cuenta la covarianza con la parte no explicada de las variables dependientes, esto es la matriz de concentraciones (formada por las concentraciones iniciales de hipoxantina y xantina).

El algoritmo PLS2 ofrece mejores resultados que PLS1 debido a la relación intrínseca entre la hipoxantina y la xantina. También se ha de destacar que los modelos 3W-PLS1 y 3W-PLS2 ofrecen mejores resultados que los modelos bilineales.

La figura 5.6 muestra los *loadings* obtenidos en el análisis 3W-PLS2 para los 4 primeros factores en los tres modos. El perfil del primer factor en el modo cinético (fig. 5.6b) indica que todos los tiempos aportan información útil en la determinación de hipoxantina y xantina. El *loading* cinético del segundo factor indica una disminución importante en tiempos comprendidos entre los 20 y los 90 segundos, llegando a un valor de 0. Se puede observar un comportamiento similar en los *loadings* cinéticos del tercer y

cuarto factor, indicando que la información más relevante del sistema para la cuantificación se encuentra en los primeros tiempos de reacción.

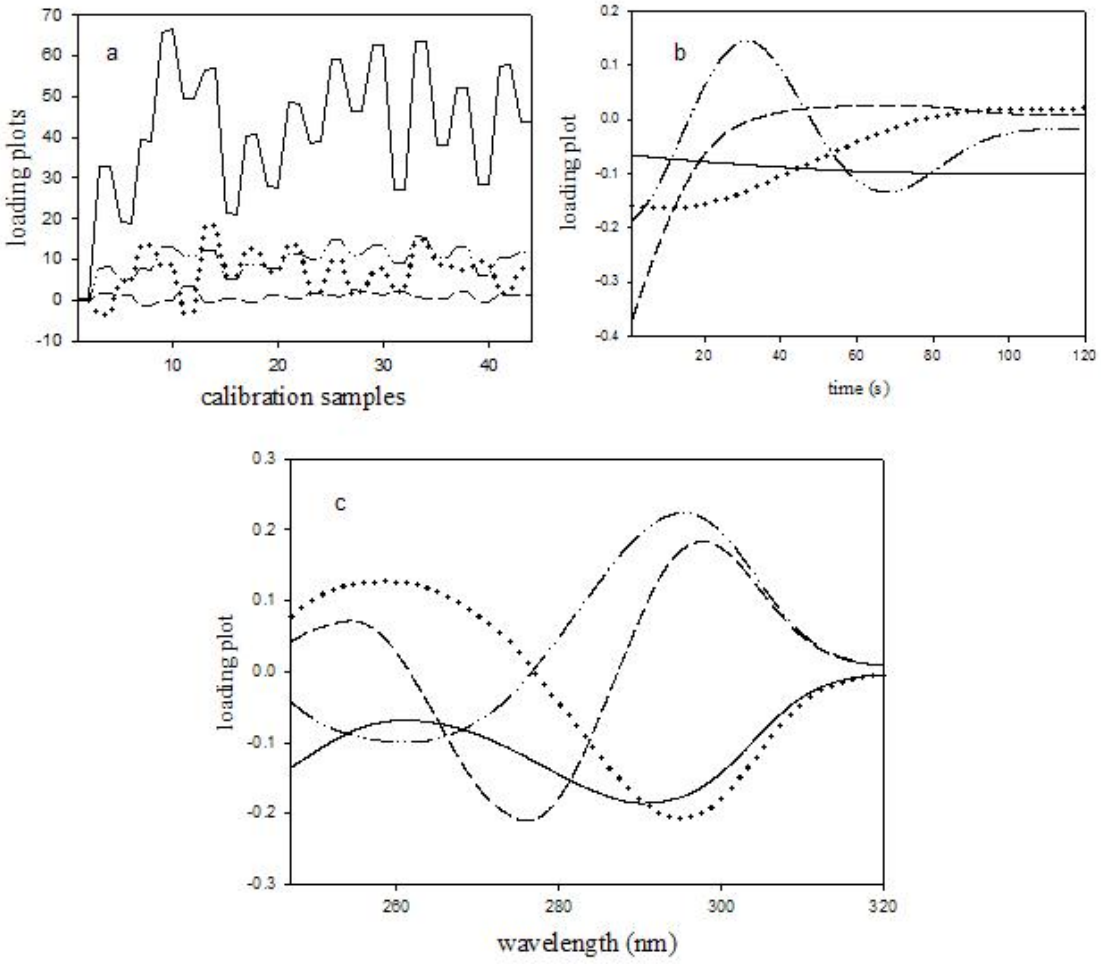


Figura 5.6: *Loadings* obtenidos en el análisis 3W-PLS2 para la cuantificación conjunta de hipoxantina y xantina. a) *Loadings* de concentración. b) *Loadings* cinéticos. c) *Loadings* espectrales. Primer factor, línea continua. Segundo factor, línea punteada. Tercer factor, rayas discontinuas. Cuarto factor, línea de rayas y puntos.

Los *loading*s espectrales para todos los factores muestran las regiones espectrales de importancia. La primera, con un máximo entre 250 y 270 nm, principalmente relacionada con la hipoxantina. La otra, con un máximo alrededor de 290 nm, relacionada con el ácido úrico. Uno de los máximos de la xantina aparece a 280 nm (figura 5.1). Sin embargo no existe ningún *loading* espectral con una gran importancia en esta longitud de onda.

La tabla 5.5 presenta los estadísticos calculados para el modelo 3W-PLS2 para las muestras de calibración, validación y muestras de orina. La desviación estándar de la pendiente y de la ordenada en el origen se muestra entre paréntesis. Tanto la pendiente, como el coeficiente de correlación (R^2) se encuentran muy cercanos a 1 para los tres conjuntos de muestras.

Tabla 5.5: Estadísticos de la regresión del modelo 3W-PLS2 para la cuantificación de hipoxantina y xantina. Desviación estándar entre paréntesis.

		calibración	**validación**	**orina**
HIPOXANTINA	**pendiente**	0.987 (0.006)	1.000 (0.006)	0.978 (0.011)
	ordenada	0.649 (0.217)	-0.438 (0.211)	0.072 (0.340)
	R^2	0.999	0.999	0.998
XANTINA	**pendiente**	1.005 (0.019)	1.001 (0.023)	0.987 (0.030)
	ordenada	-0.248 (0.358)	0.045 (0.385)	0.407 (0.509)
	R^2	0.990	0.989	0.982

En la literatura se pueden encontrar los resultados obtenidos por otros autores para la cuantificación de hipoxantina y xantina en orina humana y en soluciones dopadas de laboratorio. Los resultados obtenidos para la metodología cinética-espectrofotométrica usando 3W-PLS2 son comparables con los resultados obtenidos mediante otras metodologías analíticas (HPLC, HPGC, etc) **[Putterman 1979, Di Pietro 2001, Zen 2002]**.

5.3.5. CONCLUSIONES DE ESTE TRABAJO

Se ha demostrado la capacidad de las metodologías tridimensionales de calibración multivariable, aplicadas a sistemas enzimáticos para la cuantificación simultánea de diferentes analitos, con buenos resultados. Los algoritmos 3W-PCA y 3W-PLS se han confirmado como una alternativa a los métodos de desdoblamiento.

Con el algoritmo 3W-PCA se ha obtenido importante información cualitativa sobre el comportamiento cinético del sistema y se han podido estudiar las fuentes de variabilidad más importantes entre las muestras de laboratorio y las muestras de orina.

Se ha de destacar que la cuantificación simultánea de hipoxantina y xantina se ha realizado si ningún tipo de pretratamiento analítico previo de separación.

El algoritmo 3W-PLS2 ha ofrecido los mejores resultados en la cuantificación de los dos analitos. La influencia del ácido úrico no es significativa en los resultados. Los valores de SEP y SDBR son relativamente bajos, confirmando la obtención de modelos de calibración con buena capacidad predictiva y repetibilidad. Los resultados para la xantina indican una ligera no-linearidad del sistema (SEP mucho mayor que SDBR). Sin embargo la capacidad predictiva de los modelos es suficientemente buena para cuantificar los dos analitos en orina humana.

5.4. HS-MCR-ALS APLICADO AL ESTUDIO Y MONITORIZACIÓN DE SISTEMAS ENZIMÁTICOS EN FLUIDOS BIOLÓGICOS.

5.4.1. OBJETIVOS

En este trabajo se demuestra el potencial de la resolución multivariable de curvas *Hard-Soft*, HS-MCR-ALS, para modelar sistemas enzimáticos a través de dos aplicaciones reales ligadas a la catálisis de xantina, hipoxantina y ácido úrico usando la xantina oxidasa.

La primera aplicación está relacionada con la elucidación del mecanismo enzimático real cuando uno de los analitos se desvía del mecanismo propuesto. En esta elucidación se calcularán las constantes enzimáticas del sistema.

La segunda aplicación hace referencia al modelado de sistemas enzimáticos en presencia de interferencias absorbentes en el rango espectral de trabajo, como es la orina. En este sentido, se estudiará su efecto en el mecanismo enzimático determinado, a través del cálculo de las constantes enzimáticas del sistema con y sin influencia de la interferencia espectral.

Para este trabajo, se aplicará el algoritmo de HS-MCR-ALS al sistema formado por la hipoxantina, xantina y ácido úrico catalizados por la xantin oxidasa.

5.4.2. EL SISTEMA ENZIMÁTICO: DESCRIPCIÓN Y MONITORIZACIÓN EXPERIMENTAL

5.4.2.1. Descripción detallada del sistema enzimático. Modelo de Michaelis-Menten

Tal y como se ha descrito anteriormente el sistema en estudio es la transformación consecutiva de hipoxantina (H) y xantina (X) a ácido úrico (U) por la acción de la xantina oxidasa (E) (figura 5.1).

Se han desarrollado diferentes mecanismos enzimáticos basados en metodologías que solamente tenían en cuenta uno de los dos pasos **[Mc Whirter 1991, Olson 1974b, Massey 1969]**. Escribano y colaboradores **[Escribano 1988]** propusieron un mecanismo ping-pong incluyendo la inhibición competitiva del producto de las reacciones, el ácido úrico, de acuerdo con el siguiente esquema de reacciones:

$$\begin{aligned} H + E &\Leftrightarrow [HE] \rightarrow X + E \\ X + E &\Leftrightarrow [XE] \rightarrow U + E \\ U + E &\Leftrightarrow [UE] \end{aligned} \tag{5.8}$$

Si asumimos las condiciones de estado estacionario, se pueden obtener las siguientes ecuaciones diferenciales para describir el comportamiento de H, X y U en función del tiempo (ecuaciones 5.9, 5.10 y 5.11, respectivamente):

$$\frac{-d[H]}{dt} = \frac{V_{mx}^{X}[H]}{K_m^X + [H] + \frac{K_m^H[U]}{K_i} + \frac{K_m^H[X]}{K_m^X}} \tag{5.9}$$

$$\frac{d[X]}{dt} = \frac{V_{mx}^{H}[H]K_m^X - V_{mx}^{X}[X]K_m^H}{K_m^X K_m^H + K_m^H[X] + \frac{K_m^X K_m^H[U]}{K_i}} \tag{5.10}$$

$$\frac{d[U]}{dt} = \frac{V_{mx}^{X}[X]}{K_m^X + [X] + \frac{K_m^X[U]}{K_i} + \frac{K_m^X[H]}{K_m^H}} \tag{5.11}$$

En este mecanismo, la variación en la concentración del sustrato (H), intermedio (X) y producto (U) con el tiempo, *t*, está en función de las concentraciones y de los parámetros cinéticos del sistema. K_m^H, K_m^X son las constantes de Michaellis-Menten de H y X, respectivamente. V_{mx}^H, V_{mx}^X son sus velocidades máximas. K_i es la constante de inhibición competitiva.

Aunque este modelo es aceptado como un buen modelo enzimático, algunos autores han postulado que el ácido úrico no es estable en soluciones alcalinas y es oxidado a otros productos secundarios **[Clark 1979, Lindsey 1981]**. Esta degradación del producto puede afectar el comportamiento del ácido úrico, por lo que esta influencia debe ser estudiada.

El modelo postulado se ha determinado en soluciones acuosas, pero la catálisis enzimática se produce mayoritariamente en fluidos biológicos como sangre, plasma u orina **[Tawa 1981]**, por lo que la catálisis enzimática se puede ver afectada por el medio biológico en el que se encuentra. Esta posible influencia del medio de reacción en el proceso enzimático se estudiará también en este trabajo.

Tabla 5.6: Concentraciones iniciales (μmol L^{-1}) para la hipoxantina, xantina, ácido úrico y porcentaje de orina en los experimentos usados para cada conjunto de datos A y D.

		Concentraciones iniciales(μmol L^{-1})			
Conjunto de datos	Muestras	H	X	U	% orina
A	A1	60	0	0	
	A2	0	30	0	
	A3	30.2	15.1	0	
	A4	45.2	0	20.2	
	A5	10.1	22.2	15.1	
	A6	0	15.1	30.3	
D	D1	50.0	0	0	0
	D2	0	50.0	0	0
	D3	50.0	0	0.9	3
	D4	50.0	0	1.8	6
	D5	0	50.0	0.9	3
	D6	0	50.0	1.8	6
	D7	30.0	30.0	0.9	3
	D8	30.0	30.0	1.8	6
	D9	0	30.0	30.9	3
	D10	0	30.0	31.8	6

5.4.3. CONJUNTOS DE DATOS

La tabla 5.6 muestra las concentraciones iniciales de los analitos y de la orina en la celda de reacción (cubeta del espectrofotómetro UV-Vis acoplado al módulo de flujo interrumpido). Los rangos de concentración de la hipoxantina, la xantina y el ácido úrico recogen el rango de niveles de concentración en orina para gente sana y gente con fallos renales (gota, toxaemia, etc) **[Tawa 1981]**.

Cada muestra de la tabla 5.6 proporciona una matriz bidimensional, de dimensiones (número de espectros x número de longitudes de onda), que contiene todos los espectros registrados a lo largo del tiempo durante la reacción enzimática.

El primer conjunto de experimentos, A, contiene 6 mezclas de H, X y U a diferentes niveles de concentración comprendidos entre 10-60 μmol L^{-1}. Este conjunto de experimentos se utilizó para la elucidación del modelo enzimático correcto. El conjunto de datos D se construyó con 10 muestras que contenían H, X y U, en el rango de concentraciones de 30-55 μmol L^{-1}, más diferentes cantidades de orina añadida. El rango de concentración de la orina estuvo comprendido entre el 3% y el 6% de nivel de interferencia, aunque dos de las muestras no contenían orina (D_1 y D_2). La concentración de U de cada muestra tiene en cuenta la pequeña cantidad de ácido úrico que contenía la orina utilizada. Este conjunto de datos se utilizó para modelar el sistema enzimático en presencia de orina, una interferencia de origen biológico absorbente en el rango de longitudes de onda estudiado (230-320 nm).

5.4.4. ALGORITMO HS-MCR-ALS. INTRODUCCIÓN DE LA RESTRICCIÓN DEL MODELO ENZIMÁTICO EN MCR-ALS

Revisando el algoritmo MCR-ALS descrito en el capítulo 3 (sección 3.7) y considerando el modelo enzimático como una restricción adicional, los pasos generales del algoritmo HS-MCR-ALS son los mismos que los descritos en la sección 3.7 excepto en el tercer paso, donde se introduce la restricción utilizando las ecuaciones diferenciales

del modelo cinético (restricción de modelo). Cuando se trata más de una muestra a la vez, las correspondientes matrices de datos se encadenan por columna, tal y como se detalla en el capítulo 3 (sección 3.7) (figura 3.17). Los perfiles de concentración obtenidos en el segundo paso del algoritmo MCR-ALS y las concentraciones iniciales de los analitos involucrados en el sistema enzimático que se quieran ajustar, son usados como datos de entrada para la resolución del sistema de las ecuaciones diferenciales del modelo enzimático (ecuaciones 5.9, 5.10 y 5.11) utilizando un método no lineal. Como se propone en trabajos anteriores **[De Juan 2000, Diework 2003]**, la restricción de modelo actúa directamente en la matriz de perfiles de concentración, ya sea en la global C o en las matrices individuales C_i previamente seleccionadas. A partir de las concentraciones halladas en el ajuste no lineal se calculan unos nuevos perfiles cinéticos que actualizan a los anteriores (obtenidos en el paso 2), obteniendo además, como información adicional todos los parámetros del modelo enzimático (constantes de Michaelis-Menten, velocidades máximas de reacción).

Debido a la flexibilidad intrínseca del algoritmo básico del MCR-ALS, se consiguen ciertas ventajas en el modelado del sistema enzimático. La primera es la posibilidad de tratar sistemas enzimáticos en presencia de interferentes absorbentes que no influyen en la cinética del sistema. En este caso, sólo son ajustados los perfiles de concentración que participan en el modelo, mientras que la contribución de la especie interferente es modelada únicamente con MCR-ALS (de forma *soft*). También es posible aplicar diferentes modelos enzimáticos a diferentes experimentos en su análisis conjunto y seleccionar diferentes especies para ser modeladas utilizando las ecuaciones diferenciales cinéticas (modelado *hard*). Finalmente, el algoritmo es capaz de prever el comportamiento de especies no absorbentes (a veces denominadas “especies silenciosas”) que están incluidas en el modelo enzimático.

5.4.5. RESULTADOS Y DISCUSIÓN

5.4.5.1. Estrategia para la elucidación del modelo enzimático

A modo de ejemplo, la figura 5.7a muestra una matriz espectral obtenida en la monitorización de un experimento que contenía 30 μmol L^{-1} de hipoxantina y xantina. En la figura 5.7b se observa la absorbancia a 290 nm (máximo de absorción del U) en función del tiempo. Esta absorbancia alcanza un máximo a los 35 segundos, aproximadamente y empieza a decrecer a partir de los 40 segundos. Esta caída en la absorbancia apoya la hipótesis de la inestabilidad del U en disoluciones alcalinas (el pH de trabajo era de 8.5) por oxidación a allantoin, principalmente, una especie no absorbente en el rango de longitudes de onda de trabajo **[Clark 1979, Lindsey 1981]**.

Para elucidar el modelo enzimático correcto se puede utilizar el algoritmo HS-MCR-ALS. Se aplica a una matriz individual, o datos de la muestra D2 de la tabla 5.6. Se escogió este experimento debido a que sólo contiene X, por lo que la reacción enzimática implica solamente el caso más sencillo en el que participa el U. Era necesaria una estimación inicial de los espectros de las sustancias presentes (perfiles espectrales) que se pudo obtener realizando previamente un análisis MCR-ALS de la muestra A1 de la tabla 5.6. En ambos casos se impusieron restricciones de nonegatividad y unimodalidad a los perfiles de concentración, mientras que a los perfiles espectrales sólo se impuso la restricción de nonegatividad.

El modelo enzimático usado fue el descrito en las ecuaciones 5.9 a 5.11. Como se ha comentado anteriormente, la restricción del modelo se puede aplicar individualmente a cada una de las especies del sistema. En este caso se ha aplicado únicamente a la X, teniendo en cuenta el perfil de concentraciones de X como datos de entrada para el ajuste no lineal del sistema diferencial de ecuaciones.

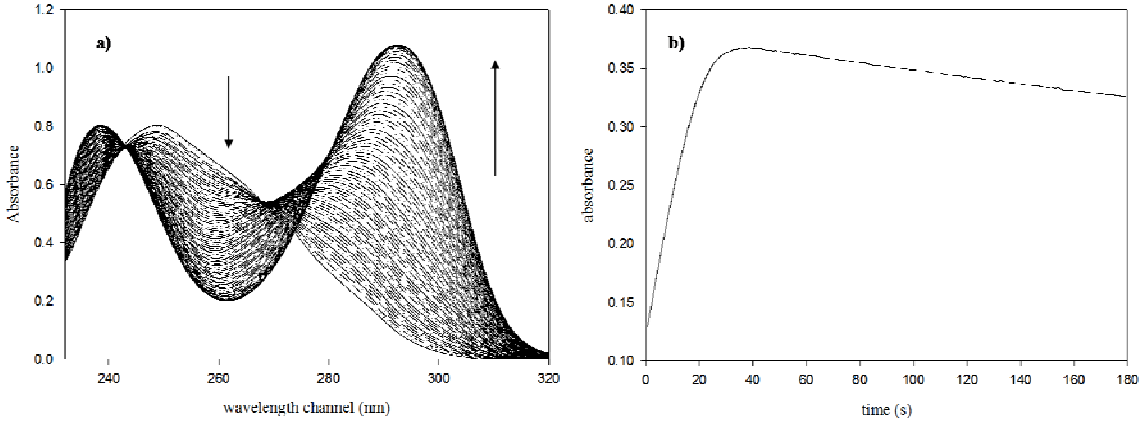


Figura 5.7: a) Monitorización UV-Visible de una muestra que contiene 30 μmol L^{-1} de hipoxantina y xantina, respectivamente (experimento A_3). b) Evolución de la absorbancia en el máximo de absorbancia del ácido úrico (290 nm.).

Los resultados serán los perfiles espectrales puros de X y U (en la matriz $\mathbf{S}^{\mathbf{T}}$), el perfil de concentración de X ajustado al modelo enzimático y, en la matriz **C**, el perfil de concentración de U modelado únicamente con MCR-ALS (de forma *soft*). Además se obtiene un perfil de concentración adicional, que correspondería al comportamiento de U si hubiera sido modelado siguiendo el mecanismo enzimático propuesto; la "especie silenciosa".

Se puede comparar el perfil de concentración del ácido úrico obtenido de forma *soft* con el perfil de concentración correspondiente al comportamiento del ácido úrico postulado por el mecanismo enzimático. Si el modelo asumido es correcto, los dos perfiles de concentración deberían coincidir. Si se observan desviaciones sistemáticas entre ambos perfiles, el modelo debería ser reformulado.

La figura 5.8 muestra los perfiles de concentración y espectrales obtenidos después de la aplicación del HS-MCR-ALS a la muestra A_2. Los perfiles en líneas continuas son los perfiles en **C** y $\mathbf{S}^{\mathbf{T}}$, mientras que los perfiles en líneas punteadas son los perfiles de la evolución de acuerdo con el modelo enzimático. Se puede observar que no existe coincidencia entre el perfil obtenido por ajuste *soft* y el perfil teórico del ácido úrico. La concentración de U decrece visiblemente a partir de los 30 segundos de

reacción. El perfil de la degradación indica que este descenso en la absorbancia podría asociarse a una cinética de degradación de primer orden (ecuación 5.12)

$$\frac{d[U]}{dt} = -K_d[U] \tag{5.12}$$

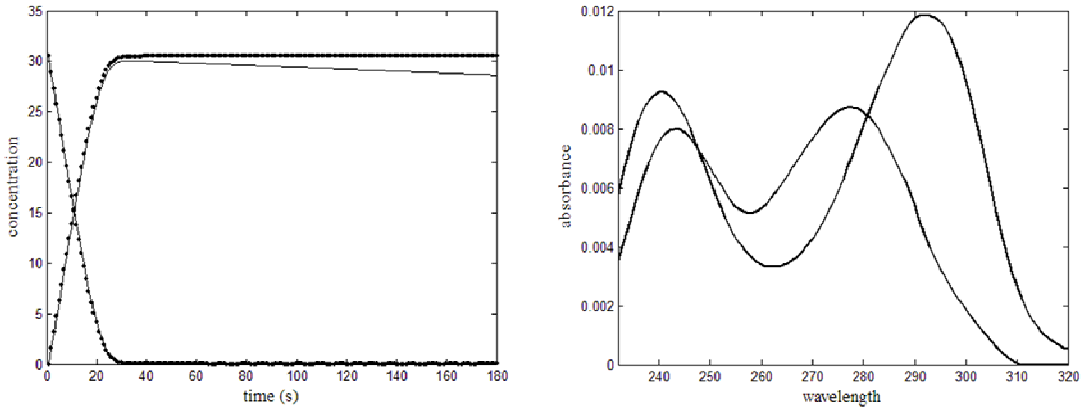


Figura 5.8: Perfiles cinéticos y espectrales obtenidos para el experimento A_3. Las líneas continuas pertenecen a los perfiles obtenidos aplicando HS-MCR-ALS. Las líneas punteadas pertenecen a la evolución cinética esperada según el modelo de Michaellis-Menten.

En esta ecuación, el descenso del U es linealmente dependiente de su concentración y de K_d, la constante de degradación. De esta manera, la expresión global del comportamiento del ácido úrico se puede definir como la unión de las ecuaciones 5.11 y 5.12 de la siguiente manera:

$$\frac{d[U]}{dt} = \frac{V_{mx}^{X}[X]}{K_m^X + [X] + \frac{K_m^X[U]}{K_i} + \frac{K_m^X[H]}{K_m^H}} - K_d[U] \tag{5.13}$$

Teniendo en cuenta este nuevo comportamiento del ácido úrico, se realizó un ajuste no lineal clásico (de *hard-modelling*) para la muestra A_2 con el nuevo modelo enzimático (experimento 4 de la tabla 5.7). La tabla 5.7 muestra la comparación entre el ajuste *hard* realizado siguiendo el modelo de **[Escribano 1988]** (experimento 3) y el ajuste *hard* realizado siguiendo el nuevo modelo. El valor de falta de ajuste es mejor

cuando se aplica el nuevo modelo.

La validación general del nuevo modelo propuesto se comprobó mediante el análisis no lineal clásico (*hard modelling)* del conjunto completo de muestras A (experimento 5). En la figura 5.9 se pueden observar los perfiles de concentración obtenidos en el ajuste *hard* para cada experimento individual y los correspondientes espectros derivados por mínimos cuadrados. El valor de falta de ajuste (tabla 5.7) es satisfactorio y los parámetros cinéticos de la tabla 5.8 concuerdan con los valores de la literatura **[Massey 1969, Escribano 1988]**. El valor de K_D es muy pequeño en comparación con los otros valores de constantes, lo que indica que la degradación del ácido úrico es un proceso lento, únicamente apreciable cuando la mayoría del ácido úrico ya ha sido formado.

Tabla 5.7: Características de los métodos aplicados para elucidar el modelo enzimático (conjunto de datos A) y para modelar el sistema enzimático en presencia de orina (conjunto de datos D).

Experimento	Muestras tratadas	Análisis[a]	Estimaciones iniciales	Restricciones[c]	% Falta de ajuste
1	A1	SM	C (EFA) [b]	[1, 2]	1.10
2	A2	HS-MCR	S^T (a partir del experimento 1)	[1, 2]	3.19
3	A2	HM viejo[e]	Constantes[h]		3.16
4	A2	HM nuevo[f]	Constantes		2.99
5	[A1-A6]	HMnuevo	Constantes		1.33
6	D1	SM	C (EFA)	[1, 2]	2.40
7	[D1 y D3]	HSMCR	S^T (experimento 6)[g]	[1, 2, 3]	2.23
8	[D1-D10]	HSMCR	S^T (experimento 7)	[1, 2, 3]	1.78

[a] HM: Modelado *Hard*, SM. modelado *Soft*, HS-MCR: Aproximación conjunta (*hard + soft*)
[b] C: Perfiles de concentración como estimaciones iniciales; S^T: Perfiles espectrales como estimaciones iniciales Estimaciones obtenidas en el *run* indicado.
[c] Restricciones aplicadas a los perfiles de concentración 1, nonegatividad, 2, unimodalidad, 3, restricción de modelo aplicado a H, X, U de acuerdo al nuevo modelo para U. Los perfiles espectrales son siempre restringidos a ser nonegativos
[d] Falta de ajuste.

$$100\times\sqrt{\left(\sum_{ij} e_{ij}^2\right)\Big/\left(\sum_{ij} d_{ij}^2\right)}$$

donde e_{ij} son los residuales y d_{ij} son los elementos de la matriz original.
[e] Modelado *Hard* de acuerdo con **[Escribano 1988]**
[f] Modelado *Hard* de acuerdo con el modelo enzimático propuesto en las ecuaciones 4.9, 4.10, 4.13
[g] Las estimaciones iniciales son los perfiles espectrales obtenidos en el *run* 6 más el primer espectro de la matriz D3
[h] Las estimaciones iniciales son valores iniciales de las constantes enzimáticas.

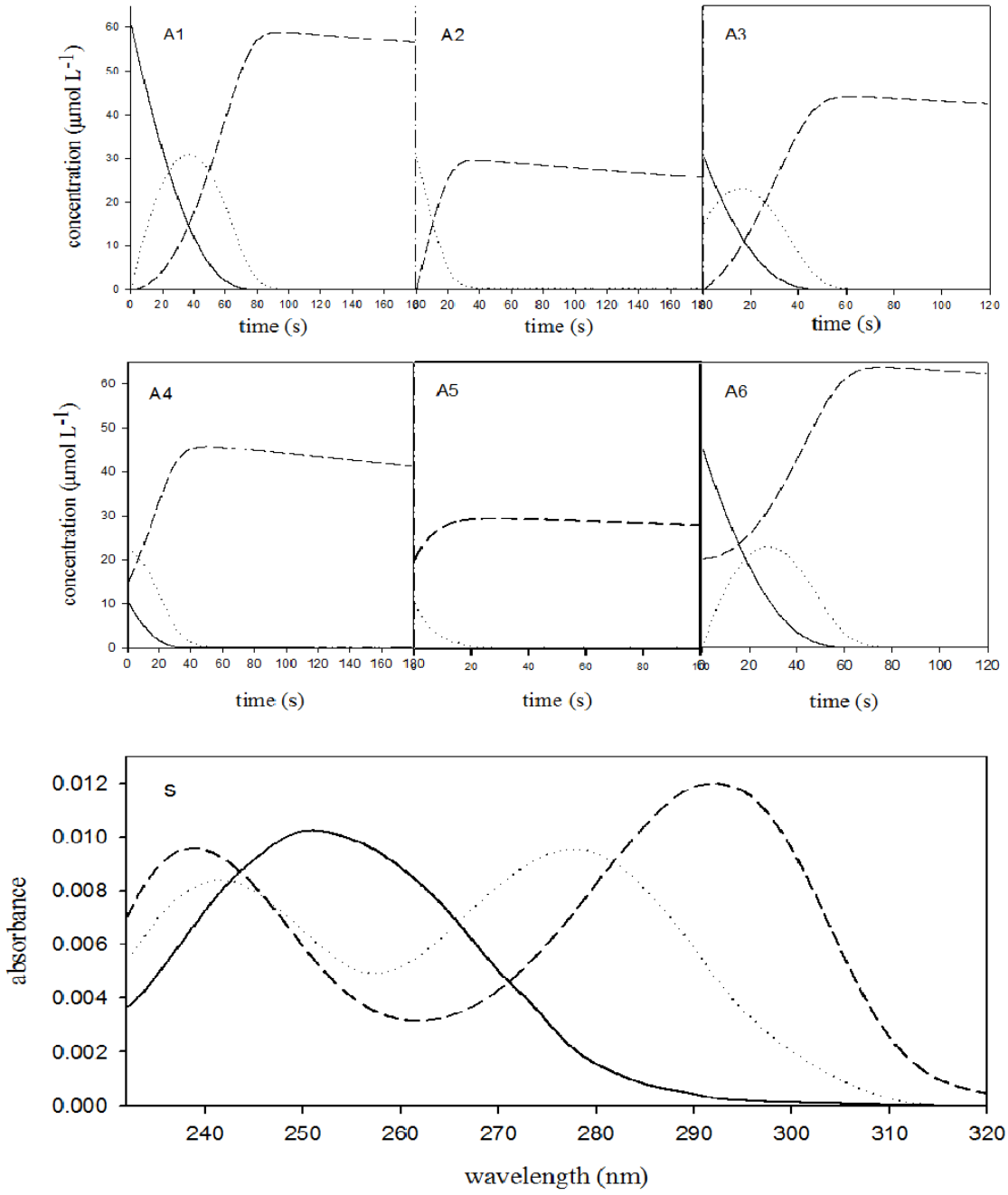


Figura 5.9: Perfiles cinéticos y espectros puros obtenidos en el análisis del conjunto de datos A por medio del análisis *hard* aplicando el nuevo modelo enzimático propuesto (hipoxantina, líneas continuas; xantina, líneas punteadas; ácido úrico, líneas rayadas)

Tabla 5.8: Constantes enzimáticas obtenidas aplicando las diferentes metodologías a los conjuntos de datos A y D. Desviación estándar entre paréntesis. Las constantes obtenidas para el conjunto de datos D son el promedio para las muestras D3 a D10.

Conjunto de datos	Metodología	Modelo enzimático	Constantes enzimáticas					
			V_H	K_H	K_X	K_I	V_X	K_D
A	Modelado *Hard*	[Escribano 1988]	1.9 (0.1)	2.2 (0.3)	4(1)	178.1 (0.4)	1.8 (0.2)	-
	Modelado *Hard*	New model	1.77 (0.08)	2.9 (0.4)	6.8 (0.7)	178.0 (0.3)	2.1 (0.1)	0.0005 (0.0001)
D	HS-MCR	New model	1.2 (0.2)	2.6 (0.4)	7.0 (1.0)	176 (2)	1.1 (0.1)	0.0006 (0.0002)

5.4.5.2. Modelado del proceso en presencia de fluidos biológicos

La segunda aplicación del HS-MCR-ALS está enfocada al modelado de procesos enzimáticos en un fluido biológico absorbente. En esta situación, el proceso tiene lugar en diferentes niveles de interferencia de orina, la cual produce una absorción que permanece constante a lo largo del proceso cinético.

En este tipo de experimentos, la presencia de una especie inerte, cinéticamente hablando, que produce una absorbancia constante a lo largo de la reacción se traduce en un problema de deficiencia de rango. Para mostrar este hecho, se realizó un análisis de rango de los experimentos D3 y D1 (la misma concentración de hipoxantina con y sin orina, respectivamente) (ver tabla 5.6). El número de componentes relacionado con cada experimento se estimó usando el análisis de factores evolutivos, EFA (*Evolving Factor Analysis*) **[Maeder 1987]**. La figura 5.10 muestra el análisis EFA para los experimentos D_1 y D_3. Las líneas por encima del nivel de ruido se refieren a los componentes químicos. Las líneas continuas se refieren a la aparición de cada componente, mientras que las líneas punteadas corresponden a la desaparición. Como se puede observar, ambos análisis indican que existen tres componentes. Este resultado es incorrecto en D3, ya que se sabe que existen cuatro componentes absorbentes. La orina no ha sido detectada como un componente adicional debido a que los dos sistemas evolucionan en paralelo, el primero

debido a las especies involucradas en el proceso enzimático y el segundo debido a las mismas especies con la contribución constante de la orina.

Una manera de resolver este problema de deficiencia de rango es la concatenación de matrices. Algunas de ellas, D_1 y D_2, sin orina (matrices de rango completo) y otras, D_3 a D_{10}, con la interferencia de la orina (matrices deficientes de rango).

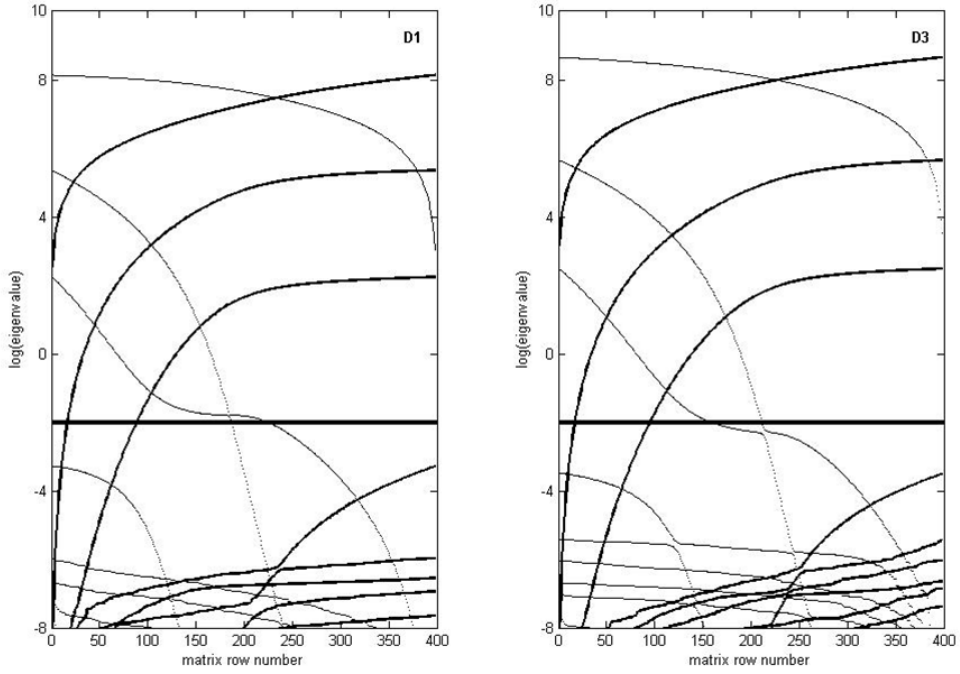


Figura 5.10: Análisis de factores evolutivos, EFA, para los experimentos D_1 y D_3. Las líneas continuas se refieren a la aparición de cada componente, mientras que las líneas punteadas corresponden a la desaparición.

La estrategia para la resolución del problema de la deficiencia de rango pasa por el análisis de la muestra D_1 mediante MCR-ALS para obtener una buena estimación inicial de los perfiles espectrales de H, X y U (tabla 5.7, experimento 6). En el estudio se impusieron restricciones de nonegatividad y unimodalidad a los perfiles de concentración y la restricción de nonegatividad a los perfiles espectrales.

El modelado conjunto de la muestra D_1 (de rango completo) y de la muestra D_3 (deficiente de rango) permitió solventar el problema de la deficiencia de rango en la

muestra D_3 y el correcto modelado del sistema completo, dejando fuera del proceso cinético la interferencia de la orina. Para ello se aplicó HS-MCR-ALS a las muestras D_1 y D_3 (experimento 7 de la tabla 5.7) considerando cuatro contribuciones absorbentes, tres relacionadas con las especies involucradas en el proceso enzimático, sometidas a la restricción del modelo, y una adicional relacionada con la interferencia absorbente, que se modeló únicamente utilizando MCR-ALS (*soft modelling*). Las estimaciones iniciales para la matriz aumentada (concatenada) $[D_1; D_3]$ fueron los perfiles espectrales de H, X y U, obtenidos a partir del experimento 6, más el espectro a tiempo inicial de la muestra D_3 para obtener una estimación del espectro de la orina. Como siempre, los perfiles de concentración de H, X y U fueron forzados a ser nonegativos, unimodales y ajustados al nuevo modelo enzimático; el perfil de concentración de la orina fue forzado a ser nonegativo para la muestra D_3. Los perfiles espectrales de las cuatro especies se restringieron a ser nonegativos. Los buenos resultados de este análisis se pueden observar en las tablas 5.7 y 5.8.

El verdadero potencial de HS-MCR-ALS se pudo demostrar en el análisis de las 10 muestras conjuntas de D. Las restricciones se aplicaron atendiendo a la diferente naturaleza de cada experimento, de la misma manera que se hizo en el experimento 7. Las estimaciones iniciales fueron los perfiles espectrales de las cuatro especies absorbentes obtenidos en el experimento 7. El % de la falta de ajuste aparece en la tabla 5.7 (experimento 8), mientras que la tabla 5.8 muestra el promedio de las constantes cinéticas calculadas para las muestras D_3 a D_{10}, las que tienen interferencia de orina. En la misma tabla también se muestran las constantes promedio obtenidas para un ajuste no lineal clásico de cada una de los experimentos individuales con orina. La pequeña dispersión en los resultados obtenidos confirma la validación del modelo general y la calidad de los datos experimentales.

La figura 5.11 muestra los perfiles de concentración resueltos y los perfiles espectrales para el conjunto de datos D. Se han obtenido perfiles espectrales para H, X y U idénticos a los que se obtuvieron en el conjunto de datos A (sin orina) debido a que la interferencia de la orina pudo ser modelada independientemente.

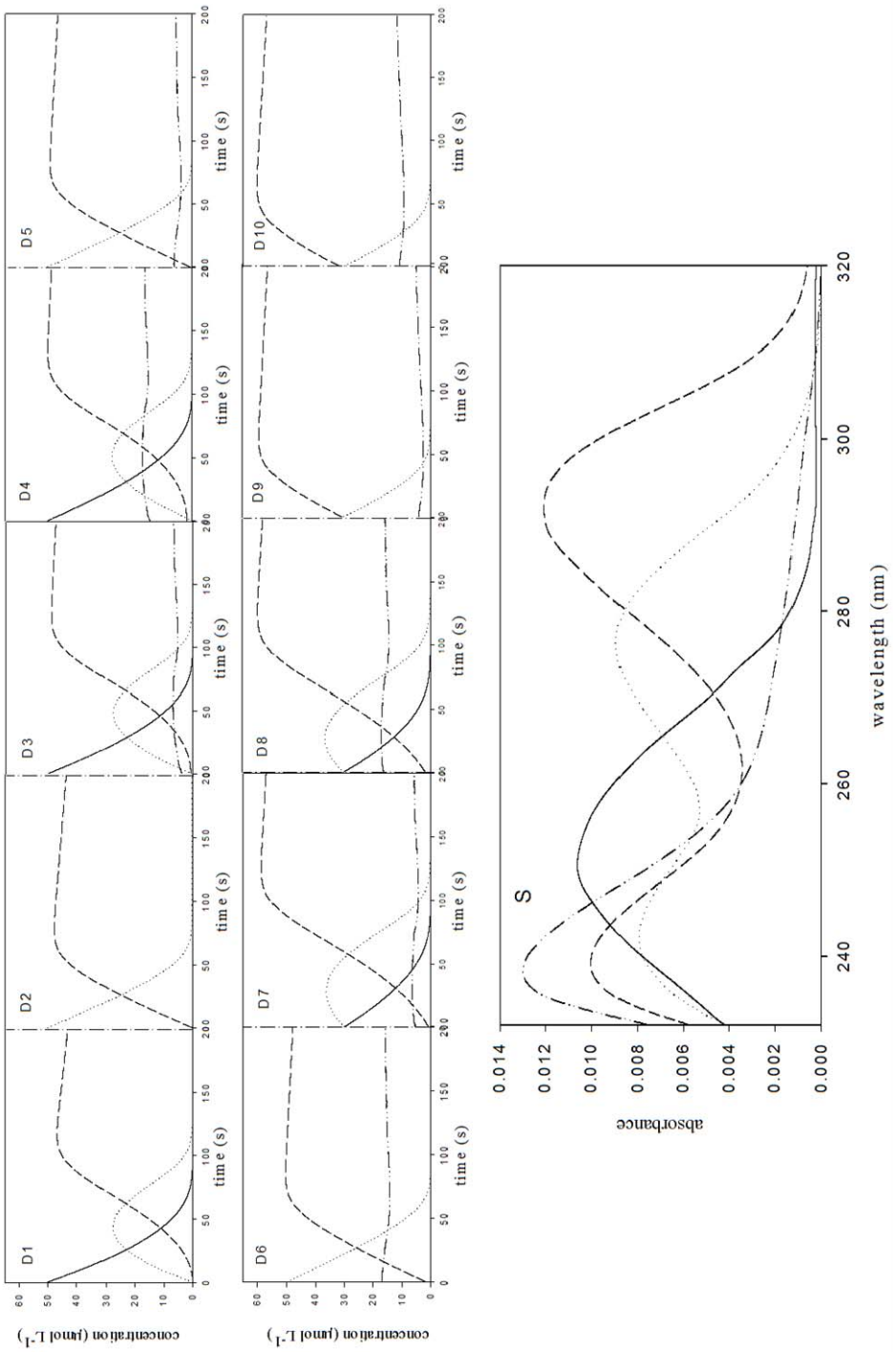


Figura 5.11: Perfiles cinéticos y espectrales obtenidos en el análisis del conjunto de datos D por medio de la aplicación del algoritmo HS-MCR-ALS (hipoxantina, líneas continuas; xantina, líneas punteadas; ácido úrico, líneas rayadas; orina, líneas rayadas-punteadas).

Como se esperaba, el perfil de concentración de la orina aparece como una contribución constante a lo largo del tiempo de reacción. Los dos niveles de interferencia se han modelado correctamente en los diferentes experimentos. Los perfiles de concentración obtenidos para H, X y U corresponden a los perfiles esperados de acuerdo al sistema enzimático propuesto. Las constantes enzimáticas obtenidas en presencia de orina no difieren significativamente de las obtenidas sin orina.

5.4.6. CONCLUSIONES DE ESTE TRABAJO

La combinación del modelado clásico *hard* con el modelado *soft* en el algoritmo HS-MCR-ALS muestra un efecto sinérgico que mejora los resultados obtenidos por cada una de las técnicas individuales para el estudio de procesos enzimáticos. Las diferentes aplicaciones de las restricciones de modelo en el algoritmo HS-MCR-ALS han ayudado a elucidar el mecanismo enzimático real y a tener en cuenta la presencia de una interferencia absorbente originada por un fluido biológico.

El nuevo mecanismo propuesto, que incluye la oxidación del ácido úrico, fue sugerido debido a la diferencia existente entre el perfil de concentración obtenido con MCR-ALS para el ácido úrico y la información encontrada en la literatura. Esta misma estrategia se puede usar en otros ejemplos de procesos para elucidar las evoluciones reales de las especies.

HS-MCR-ALS ha demostrado su utilidad cuando las reacciones enzimáticas tienen lugar en fluidos biológicos. Aplicando la restricción de modelo sólo a la hipoxantina, xantina y ácido úrico y permitiendo a la orina ser modelada de forma *soft*, se han obtenido las constantes de reacción, los espectros puros y los perfiles de concentración para los cuatro componentes en el sistema. Este tipo de análisis de datos permite separar y modelar correctamente las contribuciones de especies inertes. Además, el efecto de estas interferencias puede ser detectado a través de posibles variaciones en las constantes de la reacción detectadas en el modelado de los diferentes experimentos.

5.5. HS-MCR-ALS APLICADO A LA DETERMINACIÓN SIMULTÁNEA DE OXIPURINAS Y ÁCIDO ÚRICO EN ORINA HUMANA

5.5.1. OBJETIVOS

En el trabajo previo, se obtuvo el mecanismo real de catálisis de las oxipurinas y el ácido úrico con la xantina oxidasa y se calcularon sus constantes enzimáticas utilizando el algoritmo HS-MCR-ALS con restricción de modelo.

El objetivo de este tercer trabajo es la utilización del algoritmo desarrollado de HS-MCR-ALS para la obtención de información cuantitativa a través de la resolución cinética y la determinación simultánea de las oxipurinas y el ácido úrico. Se aplicará el algoritmo HS-MCR-ALS a la matriz aumentada (concatenada) formada por solamente dos disoluciones estándar de hipoxantina y xantina, respectivamente, y muestras con orina. La restricción de modelo se aplicará solamente a los tres analitos involucrados en el modelo enzimático en las dos disoluciones estándar, mientras que las muestras de orina se modelarán usando solamente el modelado *soft*.

El segundo objetivo de este trabajo es la comparación de los resultados obtenidos con la nueva metodología HS-MCR-ALS con 3W-PCA y 3W-PLS y otras metodologías clásicas de análisis (HPLC, HPGC, etc).

5.5.2. CONJUNTO DE DATOS

Se prepararon 19 mezclas de H, X y U comprendidas en un rango de concentración de 0-100 μmol L^{-1}. Estas mezclas se doparon con diferentes cantidades de orina regenerada para obtener diferentes niveles de interferencia (tabla 5.9). La orina utilizada en este experimental no contenía H ni X, pero sí que contenía una pequeña cantidad de U. Esta concentración se encuentra certificada por el fabricante, por lo que se tuvo en cuenta a la hora de la cuantificación. Los valores de concentración de la tabla son

los valores de concentración en la celda de reacción a tiempo 0, justo antes de comenzar a monitorizar las reacciones enzimáticas. Todas las disoluciones se prepararon justo antes de ser registradas sus cinéticas.

Tabla 5.9: Composición de los dos estándares (D1 y D2) y de las 19 mezclas (D3-D21) que contenían orina (Expresadas en % de volumen).

	Concentración (μmol L^{-1})			
Muestra	Hipoxantina	Xantina	Ácido úrico	Orina %
D1	50	0	0	0
D2	0	50	0	0
D3	20	0	0.23	0.8
D4	20	0	0.69	2.4
D5	20	0	1.85	6.4
D6	0	50	1.85	6.4
D7	0	20	0.69	2.4
D8	0	20	0.23	0.8
D9	0	20	0.69	2.4
D10	0	20	1.85	6.4
D11	0	50	0.69	2.4
D12	50	0	0.69	2.4
D13	0	0	51.85	6.4
D14	30	30	0.69	2.4
D15	30	30	1.85	6.4
D16	30	0	30.69	2.4
D17	30	0	31.85	6.4
D18	0	30	30.69	2.4
D19	0	30	31.85	6.4
D20	30	30	30.69	2.4
D21	30	30	31.85	6.4

5.5.3. TRATAMIENTO DE LOS DATOS

5.5.3.1. Aplicación de HS-MCR-ALS a la cuantificación de los analitos

Para conseguir obtener información cuantitativa con HS-MCR-ALS es absolutamente necesario el encadenamiento de las matrices obtenidas con las diferentes muestras. Se encadenan las matrices estándar, de las que se dispone toda la información necesaria, con las muestras de orina, de las que no se conocen las concentraciones de los analitos ni el nivel de interferencia. Estas matrices se encadenan por columna tal y como se muestra en la figura 5.12.

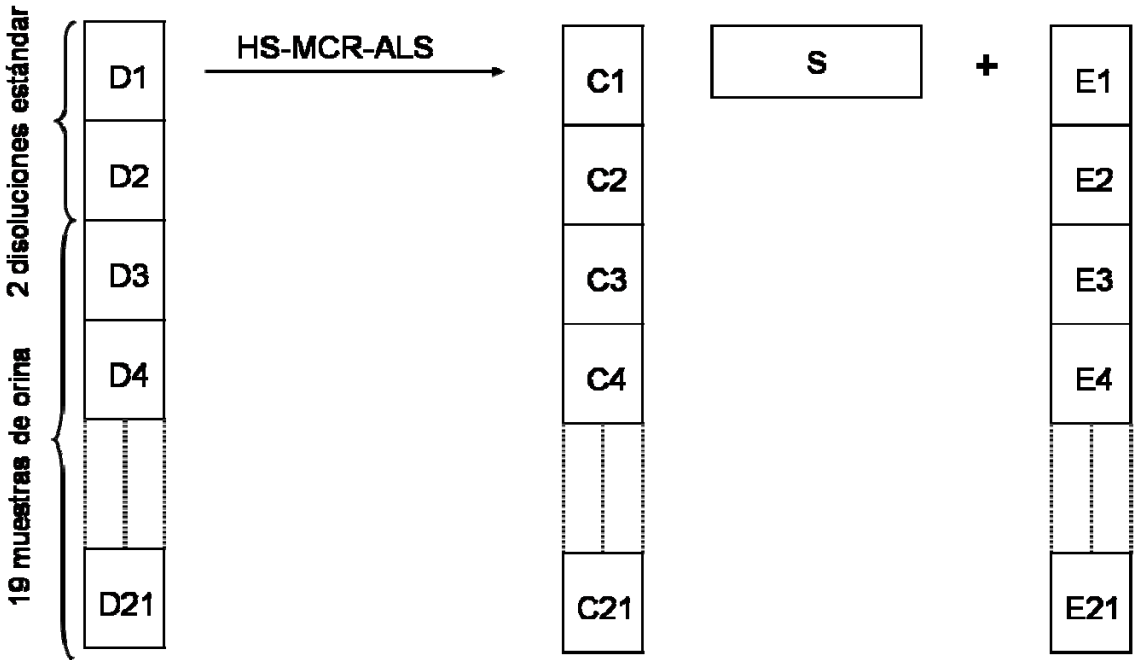


Figura 5.12: Concatenación en columna de las dos muestras estándar y las 19 muestras de orina para la aplicación de HS-MCR-ALS.

En la resolución del sistema se obtendrá una matriz $\mathbf{S}^T$ con los perfiles espectrales de los tres analitos y la orina, y tantas submatrices $\mathbf{C}_i$ como muestras haya presentes en el análisis. La información cuantitativa obtenida, esto es la concentración inicial de los analitos (H, X y U) en las muestras con orina, se obtendrá en la primera fila de cada una de las matrices $\mathbf{C}_i$ (figura 5.13).

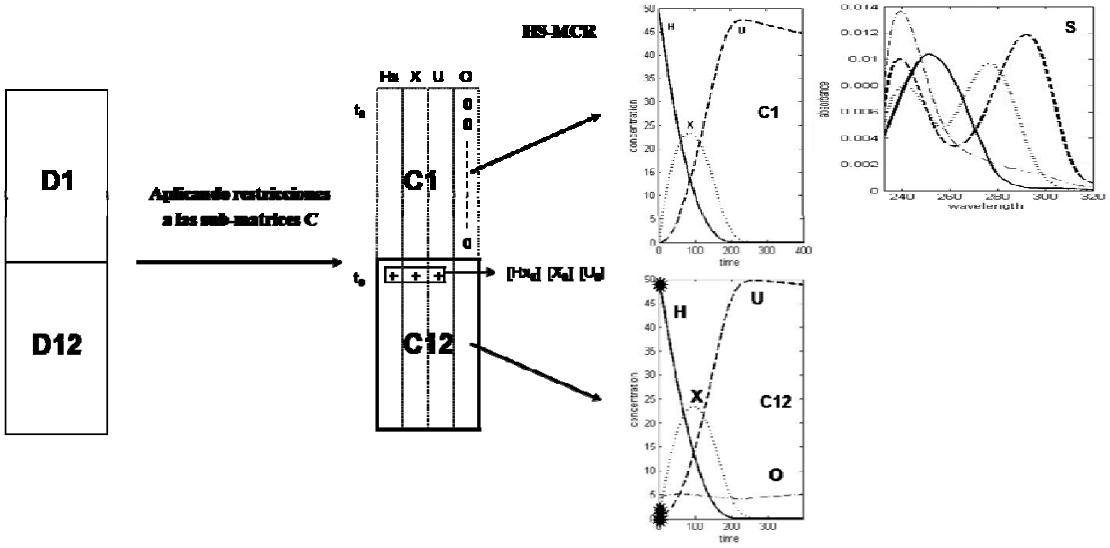


Figura 5.13: Esquema gráfico para la resolución HS-MCR-ALS de una matriz aumentada [D_1;D_{12}] que contiene un experimento sin orina (D_1) y otro con orina (D_{12}). La matriz D_1 se ajusta al modelo enzimático, mientras que D_{12} es modelada de forma *soft*. Se obtienen las concentraciones iniciales de D_{12} (*).

A parte de la restricción de modelo, se pueden imponer ciertas restricciones para mejorar la calidad de los resultados cuantitativos. Una de ellas es la llamada "correspondencia entre especies" **[Tauler 1993]**. Esta restricción usa la información relacionada con la identificación de las especies en las diferentes mezclas, si se conoce, y fuerza a las especies ausentes a tener perfiles de concentración nulos. Esta información es codificada en una matriz denominada matriz *isp*. Esta matriz tiene el mismo número de filas que el número de muestras encadenadas y tienen un número de columnas igual al número de especies totales que forman el sistema. La presencia o ausencia de una especie en particular en la matriz *isp* se codifica como 1 o 0, respectivamente.

En nuestro sistema, esta información está únicamente disponible en las dos muestras estándar porque en las muestras, tal y como pasa en la realidad, se asume que no se conoce la composición. Es importante destacar que, en las muestras estándar, se tienen que tener en cuenta tanto las especies iniciales como las que se forman debido a la reacción enzimática a la hora de construir la matriz *isp*.

Una matriz isp relacionada con una matriz encadenada que contenga los dos estándares de H (**D_1**) y X (**D_2**) y una muestra de orina (**D_3**) se construiría de la siguiente manera:

$$\begin{array}{c} \\ D_1 \\ D_2 \\ D_3 \end{array} \begin{array}{c} \begin{array}{cccc} H & X & U & O \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \end{array}$$

"O" se refiere a la orina, la cuarta especie absorbente en el sistema. Se sabe que los estándares están preparados en solución acuosa, por lo que no habrá contribución de la orina (0 en las columnas) y que en la mezcla de H se formará X y U por el propio proceso enzimático. Para la X se formará U. En el caso de las muestras de orina, sabemos que siempre tendremos la contribución espectral de la propia orina (1 en la columna) y que las otras tres especies (los analitos) pueden estar presentes, a pesar de que se desconoce su descomposición.

La aplicación de la restricción de modelo a los perfiles de concentración en las matrices estándar juega un importante papel en la calidad de la información cuantitativa obtenida. El efecto positivo de esta restricción responde, principalmente, a dos razones:

a) La minimización de las ambigüedades en los perfiles de concentración obtenidos debido al cumplimiento del modelo enzimático. Los perfiles espectrales de concentración se definen mucho mejor al aplicar HS-MCR-ALS que si se aplicasen solamente MCR-ALS

b) Para aplicar el ajuste del modelo enzimático son necesarias dos cosas: las leyes por las que se rige el sistema y la concentración inicial de cada especie implícita en el sistema. Esta información está disponible en los estándares, de los que se conoce su composición inicial. El introducir esta información elimina la típica ambigüedad de intensidad ligada a la resolución MCR-ALS, la cual conlleva obtener información cuantitativa relativa.

5.5.3.2. Validación de los resultados cuantitativos

Como en cualquier método analítico cuantitativo, se necesitan parámetros estadísticos que confirmen la fiabilidad de los resultados obtenidos en cuanto a precisión y exactitud en las concentraciones de los tres analitos. En este trabajo se utilizaron los siguientes: el error cuadrático medio de predicción, RMSEP (*Root Mean Square Error of Prediction*) (ecuación 5.14) y el error estándar de predicción, SEP (*Standard Error of Prediction*) (ecuación 5.5). Estos dos parámetros se calcularon teniendo en cuenta el *bias* del modelo (ecuación 5.6).

$$\mathbf{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{\mathbf{c}}_i - \mathbf{c}_i)^2}{\mathbf{n}}} \qquad (5.14)$$

En este trabajo no se calculó ningún parámetro estadístico referente al calibrado, ya que no se realiza un calibrado propiamente dicho. Toda la información referente a las dos muestras estándar se utiliza en las iteraciones HS-MCR-ALS, por lo que la concentración inicial de los analitos ya está fijada. A la hora de ajustar los perfiles de concentración de estas muestras al modelo enzimático, los valores de inicio son las concentraciones iniciales de cada analito presente.

5.5.4. RESULTADOS Y DISCUSIÓN

En primer lugar se realizó un análisis MCR-ALS a la muestra D_1, la cual contenía el estándar de H, con el fin de obtener estimaciones iniciales de los perfiles espectrales de H, X, U (tabla 5.10, experimento 0). Se impusieron las restricciones de nonegatividad y unimodalidad a los perfiles cinéticos, mientras que para los perfiles espectrales se impuso solamente la restricción de nonegatividad.

Tabla 5.10: Aplicación de las metodologías propuestas para el análisis del sistema enzimático.

Experimento	Muestras	Método de análisis[a]	Estimaciones iniciales[b]	Restricciones[c] Estándares	Muestras	% Falta de ajuste
0	D1	SM	C (EFA)	[1, 2]	-	1.24
1	[D1 y D12]	HS-MCR	S^T (experimento 0)[d]	[1, 2, 3, 4]	[1, 2, 4]	1.40
2	[D1 y D2; D3-D21]	HS-MCR	S^T (experimento 1)	[1, 2, 3, 4]	[1, 2, 4]	1.35
3	[D1 y D2; D3-D21]	HS-MCR	S^T (experimento 1)	[1, 2, 3, 4]	[1, 2, 4]	1.80

[a] SM: Modelado *Soft* (MCR-ALS). HS-MCR: Modelado conjunto *Hard-Soft*
[b] C: Perfiles de concentración. S^T: Perfiles espectrales obtenidos en el *run* indicado
[c] Para los perfiles de concentración ;1, nonegatividad; 2, unimodalidad; 3, restricción de modelo. Los perfiles espectrales fueron restringidos a ser nonegativos.
[d] Estimaciones iniciales del *run* 0 más el primer espectro obtenido en el experimento D12

Para realizar cualquier cálculo es necesario obtener una buena estimación espectral de la orina ya que, como se ha visto en el trabajo anterior, ésta es una interferencia constante a lo largo de la reacción. Para ello se puede resolver un sistema sencillo formado por dos matrices encadenadas; una de ellas matrices es el estándar de H, muestra D_1, y la otra es una de las muestras de orina, muestra D_{12} (tabla 5.10, experimento 1). Como estimaciones iniciales se usaron los perfiles espectrales de H, X y U obtenidos en el experimento 0 y el primer espectro de D_{12} como estimación inicial del espectro de la orina.

Se aplicó la restricción del modelo cinético a la muestra D_1. Así se minimizan las posibles ambigüedades rotacionales y de intensidad en dichos perfiles. Por otra parte, los perfiles cinéticos de la muestra D_{12} fueron únicamente modelados con MCR-ALS (*soft modelling*). A ambas muestras se aplicaron las restricciones de nonegatividad a los perfiles cinéticos y espectrales, mientras que la restricción de unimodalidad se aplicó a los perfiles de concentración. Se sabe de antemano que D1 no contiene orina. Además, H evoluciona catalíticamente a X y a U, dos especies absorbentes. Por el contrario, de la muestra D12 sólo se conoce con seguridad que contiene orina. Por esta razón, la fila perteneciente a D12 en la matriz *isp* fue codificada con 1, asumiendo que la orina podría contener cualquiera de los analitos. La matriz *isp* usada en este análisis fue:

$$isp = \begin{array}{c} \begin{array}{cccc} H & X & U & O \\ \downarrow & \downarrow & \downarrow & \downarrow \end{array} \\ \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \end{array} \begin{array}{l} \leftarrow D1 \\ \leftarrow D12 \end{array}$$

En la figura 5.13 se puede observar el resultado de este análisis HS-MCR-ALS. Se han obtenido buenos perfiles cinéticos para las dos muestras (C1 y C12) y buenos perfiles espectrales para H, X y U. También se ha obtenido una estimación de los parámetros cinéticos del modelo propuesto para la muestra D1 y, por supuesto, la concentración inicial de los tres analitos de la muestra D12. Esta información se encuentra en la primera fila de la sub-matriz C12. La interferencia de la orina fue modelada de forma *soft* obteniéndose un perfil cinético constante (ya que se trata de una interferencia con un valor constante de absorbancia) y un perfil espectral como el de la figura 5.13. Este perfil espectral se usará en sucesivos experimentos como estimación inicial.

Una vez obtenida la información necesaria, se lleva a cabo el análisis conjunto de las 21 muestras presentadas anteriormente. La matriz concatenada utilizada (presentada como experimento 2 en la tabla 5.10) incluye los dos estándar, D1 y D2, y el resto de experimentos con orina (D3-D21). Para este análisis se usaron los perfiles espectrales obtenidos en el experimento 1 como estimaciones iniciales. La nueva matriz aumentada $[D_1;D_{21}]$ se modeló en un proceso análogo al experimento 1. En este análisis, la restricción de modelo se aplicó únicamente a los perfiles cinéticos de los estándar D1 y D2, y el resto de experimentos fueron modelados de forma *soft*. De eta manera, la matriz *isp* usada en este experimento fue:

$$isp = \begin{array}{c} \begin{matrix} Hx & X & U & O \\ \downarrow & \downarrow & \downarrow & \downarrow \end{matrix} \\ \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & 1 \end{pmatrix} \end{array} \begin{matrix} \leftarrow D1 \\ D2 \\ D3 \\ \\ \\ D21 \end{matrix}$$

Las dos primeras filas corresponden a los dos estándar, mientras que el resto de filas son "vectores unidad" correspondiente a las muestras de orina.

Los resultados obtenidos para la descomposición HS-MCR-ALS de la matriz aumentada se muestran en la figura 5.14. Los perfiles cinéticos y espectrales obtenidos para los analitos y la interferencia de la orina coinciden con el modelo enzimático propuesto, obteniéndose valores bajos de falta de ajuste (tabla 5.10).

Las concentraciones obtenidas para los diferentes analitos se compararon con los valores de referencia. Los parámetros estadísticos y la correlación lineal entre los valores de concentración obtenidos y los valores de referencia se pueden observar en la tabla 5.11 y en la figura 5.15a. Los valores de SEP y RMSEP obtenidos son bajos, mostrando que la capacidad predictiva del método propuesto es buena, con valores de *bias* cercanos a cero. Es importante señalar que sólo se han utilizado dos muestras estándar para obtener la información cuantitativa de las 19 muestras de orina.

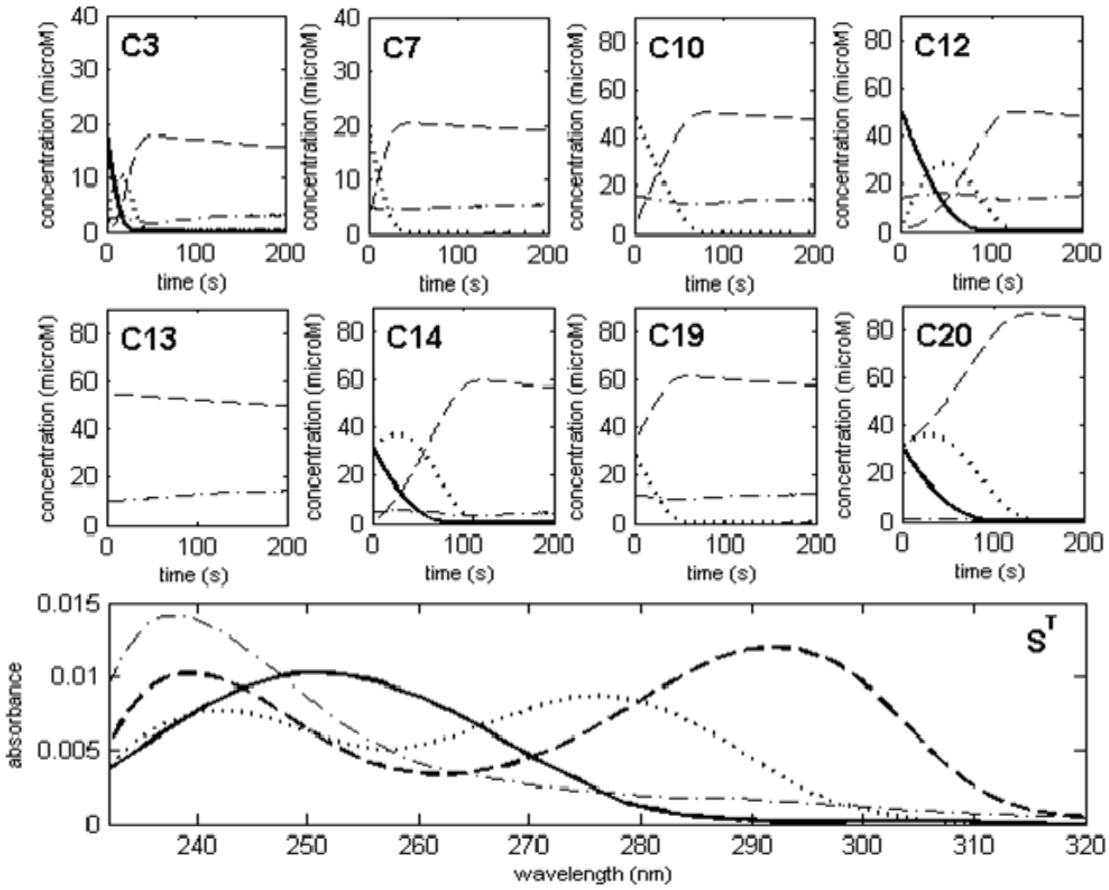


Figura 5.14: Perfiles de concentración (C3-C20) y espectros puros (S^T) obtenidos en el *experimento* 2 para algunas de las muestras con orina (hipoxantina, líneas continuas; xantina, líneas punteadas; ácido úrico, líneas rayadas; orina, líneas rayadas-punteadas).

Tabla 5.11: Comparación de los parámetros estadísticos obtenidos para las diferentes maneras de aplicación de HS-MCR-ALS en la cuantificación de H, X, y U. Desviación estándar entre paréntesis. RMSEP, SEP y *bias* están expresadas en μmol L^{-1}.

hipoxantina

Experimento	Pendiente	Ordenada en el origen	Correlación	RMSEP[b]	SEP	Bias
2	1.02 (0.02)[a]	-1.7 (0.7)	0.986	2.393	2.057	1.311
3	1.01 (0.02)	-0.4 (0.5)	0.993	1.488	1.527	-0.059

Xantina

Experimento	Pendiente	Ordenada en el origen	Correlación	RMSEP	SEP	Bias
2	1.03 (0.02)	-2.1 (0.5)	0.994	1.964	1.465	1.350
3	0.96 (0.01)	0.9 (0.3)	0.997	1.143	1.120	0.341

Ácido úrico

Experimento	Pendiente	Ordenada en el origen	Correlación	RMSEP	SEP	Bias
2	0.97 (0.02)	-0.8 (0.3)	0.995	1.666	1.226	1.163
3	1.02 (0.01)	0.5 (0.3)	0.997	1.412	1.085	0.937

De los resultados obtenidos se puede observar (figura 5.15a) que la contribución de la xantina y la hipoxantina en algunas de las muestras es extremadamente pequeña. Se puede sugerir que estos analitos no están presentes en la muestra.

El último experimento (experimento 3) se realizó usando la misma matriz que en el experimento 2 (2 estándares más 19 muestras de orina) con las mismas restricciones, pero modificando la matriz *isp* para tener en cuenta la información acerca de la ausencia de ciertos analitos en ciertas muestras de orina. Ahora, a las filas correspondientes a las muestras D6-D10 se las codificó con un "0" en la columna correspondiente a H y la fila de la matriz D13 se codificó con un "0" en las columnas de H y X.

La similitud en las concentraciones obtenidas para cada analito en las muestras, la similitud en los perfiles de concentración y espectrales obtenidos y el valor aceptable de

falta de ajuste en este análisis confirman la validez de la nueva información de la ausencia de especies incluidas en el experimento 3. Como consecuencia de la información incorporada, los valores de SEP y RMSEP y los parámetros estadísticos (tabla 5.11, figura 5.15b) para los tres analitos, son mejores que las obtenidas en el experimento 2.

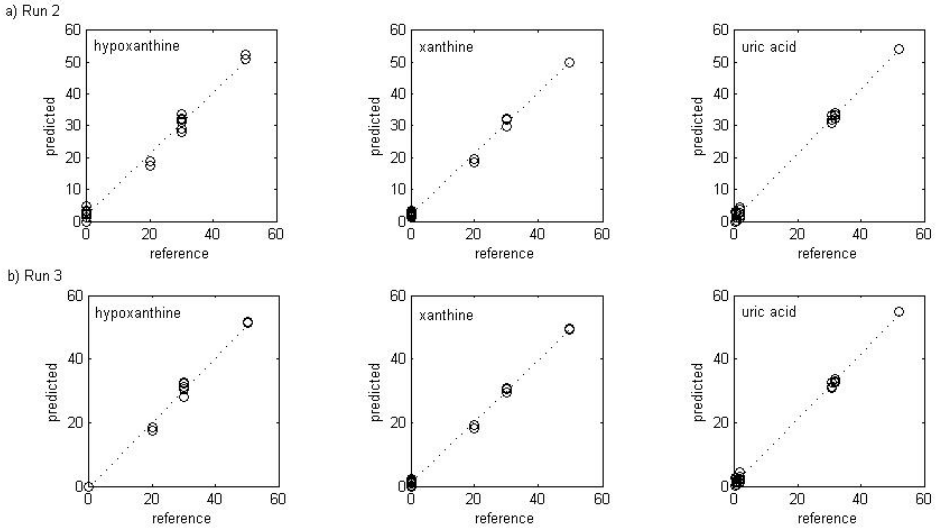


Figura 5.15: Rectas de correlación entre los valores predichos con HS-MCR-ALS y los valores de referencia para los tres analitos. a) experimento 2, b) experimento 3. Concentración en μmol L^{-1}.

5.5.5. COMPARACIÓN DE LOS RESULTADOS OBTENIDOS CON DISTINTAS METODOLOGÍAS

La tabla 5.12 muestra que los resultados obtenidos por HS-MCR-ALS (tablas 5.10 y 5.11) son comparables en calidad a otros resultados obtenidos por técnicas de separación, como HPLC **[Putterman 1979]** y por otras metodologías quimiométricas como *unfold*-PLS y 3W-PLS **[Amigo 2005]**. La ventaja de la nueva metodología presentada frente a las técnicas de separación es que con ésta no son necesarios ni pretratamientos previos de la muestra ni la eliminación de interferencias. Comparando con otras metodologías multivariable de análisis, el número de disoluciones estándar

necesarias para el HS-MCR-ALS es mucho más pequeño y la información obtenida ofrece una visión cualitativa más veraz sobre la evolución de la reacción enzimática.

Tabla 5.12: Comparación de los resultados obtenidos con otros trabajos de determinación de hipoxantina, xantina y ácido úrico. Todas las concentraciones se refieren a la concnetración en orina. HPLC: High Performance Liquid Chromatography. GC: Gas Chromatography. NSPEs: *Preanodized nontroite-coated screen-printed carbon electrodes*.

	Muestra	**Método**	**analito**	**Rango de concentración**	**r^2**
[Putterman 1979]	Orina humana	HPLC and GC	Hipoxantina	0-1300 µmol/L	0.95
			Xantina	0-2000 µmol/L	0.98
[Zen 2002]	Soluciones dopadas	NSPEs	Xantina	2-40 µmol/L	---
[Amigo 2005]	Orina humana	3W-PLS2	Hipoxantina	0-3000 µmol/L	0.998
			Xantina	0-1500 µmol/L	0.982
Nuevo método	Orina humana	HS-MCR	Hipoxantina	0-5000 µmol/L	0.993
			Xantina	0-5000 µmol/L	0.997
			Ácido úrico	0-3000 µmol/L	0.997

5.5.6. CONCLUSIONES DE ESTE TRABAJO

La aplicación del HS-MCR-ALS ha demostrado ser una buena metodología de trabajo para la descripción cinética completa de muestras de orina humana que contienen hipoxantina, xantina y ácido úrico. Se puede utilizar HS-MCR-ALS para la obtención de información cuantitativa en sistemas multicomponente, donde los analitos evolucionan de acuerdo a un modelo enzimático conocido con la presencia de interferentes absorbentes.

En el sistema enzimático propuesto, el HS-MCR-ALS se aplicó a un conjunto de muestras encadenadas por columna formado por procesos enzimáticos monitorizados en solución acuosa (muestras estándar) y por otros procesos monitorizados en orina humana (muestras de composición desconocida). La restricción de modelo se aplicó a los perfiles cinéticos de los estándares. Esta restricción minimiza de forma drástica la ambigüedad

rotacional de los resultados de la resolución e incorpora la información cuantitativa absoluta de los estándares como parámetros necesarios en el ajuste enzimático.

Como método cuantitativo, la metodología experimental es sencilla y rápida. Además, no es necesario ningún tratamiento previo de la muestra para eliminar las interferencias. El algoritmo HS-MCR-ALS muestra claras ventajas sobre otros métodos multivariable de calibración, tales como la pequeña cantidad de muestras estándar necesarias (uno o dos estándares son suficientes) y el hecho de que estos estándares se preparan en solución acuosa, sin necesidad de conocer o incluir las interferencias presentes en las muestras problema.

PARTE II:

CONTROL EN TIEMPO REAL DE PROCESOS DE PRODUCCIÓN DE ENZIMAS. PRODUCCIÓN DE LIPASAS

5.6. INTRODUCCIÓN

Este segundo bloque de la memoria está dedicado a la presentación de un trabajo de monitorización y control a tiempo real del proceso industrial de generación de lipasas a través de la utilización de sondas de espectrofluorimetría multidimensional y el uso de algoritmos tridimensionales. El trabajo se plantea como una alternativa a la metodología clásica de control de procesos de fermentación y producción de enzimas a escala industrial.

El *Pichia pastoris* es un microorganismo que se está convirtiendo en uno de los más utilizados para la producción de una gran variedad de proteínas, tanto a nivel académico como industrial, a partir de fuentes de carbono sencillas (Metanol y glicerol, por ejemplo) La aplicación de medidas fluorimétricas a la monitorización de cultivos de *Pichia pastoris* se encuentra todavía en un estado preliminar.

En este aspecto se realizó un trabajo previo donde se observaba la aparición de ciertos analitos fluoróforos a lo largo del proceso, como el triptófano, el NAD(P)H y la riboflavina. Se observó también que la aparición de estos fluoróforos estaba íntimamente ligada a la generación de biomasa y de enzimas. Consecuentemente, estas señales podían ser utilizadas para la estimación indirecta de biomasa y proteína **[Surribas 2006]**. La monitorización del proceso se realizó con una sonda multicanal que permitió la obtención

del espectro de excitación y cada uno de su correspondiente espectro de emisión en cada unidad de tiempo. Se utilizó PARAFAC para la obtención de una visión cualitativa del sistema y PLS para la obtención de la concentración de biomasa y glicerol a lo largo del proceso.

5.7. OBJETIVOS

El objetivo fundamental de este trabajo es el desarrollo de una metodología de actuación en el control a tiempo real de bioprocesos mediante la unión del algoritmo tridimensional PARAFAC y las metodologías de control de procesos mediante estadística multivariable, MSPC (*Multivariate Statistic Process Control*). De esta manera, se estudiará la bioquímica del proceso para utilizarla de una manera práctica. Se estudiará la relación intrínseca existente entre los fluoróforos y el substrato, productos y posibles desviaciones físico-químicas del bioproceso. Es importante definir las diferentes etapas del proceso, estudiando el comportamiento de cada uno de los fluoróforos con el fin de establecer parámetros de control adecuados que ayuden a tomar decisiones acerca del control del proceso. Por ejemplo, la predicción del punto final del bioproceso (desaparición total de substrato) es de vital importancia cuando se quiere realimentar o para el biorreactor.

5.8. RESULTADOS Y DISCUSIÓN

5.8.1. MODELO PARAFAC DE CALIBRACIÓN

Para la construcción del modelo PARAFAC de calibración se usaron los datos recogidos para el bioproceso **M**, en el cual se aseguraron condiciones normales de operación (NOC). El análisis de la varianza explicada y la consistencia del "*core*" con el número de componentes indicaron que el número óptimo de componentes era de tres (más del 99.7 % de varianza explicada para el modelo y un 84% de consistencia del "*core*"). Los perfiles espectrales de excitación y emisión para los tres componentes se muestran en la figura 5.16, respectivamente. Estos perfiles corresponden a los perfiles

espectrales de excitación y emisión de los fluoróforos que hacen variar la señal analítica a lo largo del proceso, los cuales se identificaron como triptófano, NADH y riboflavina.

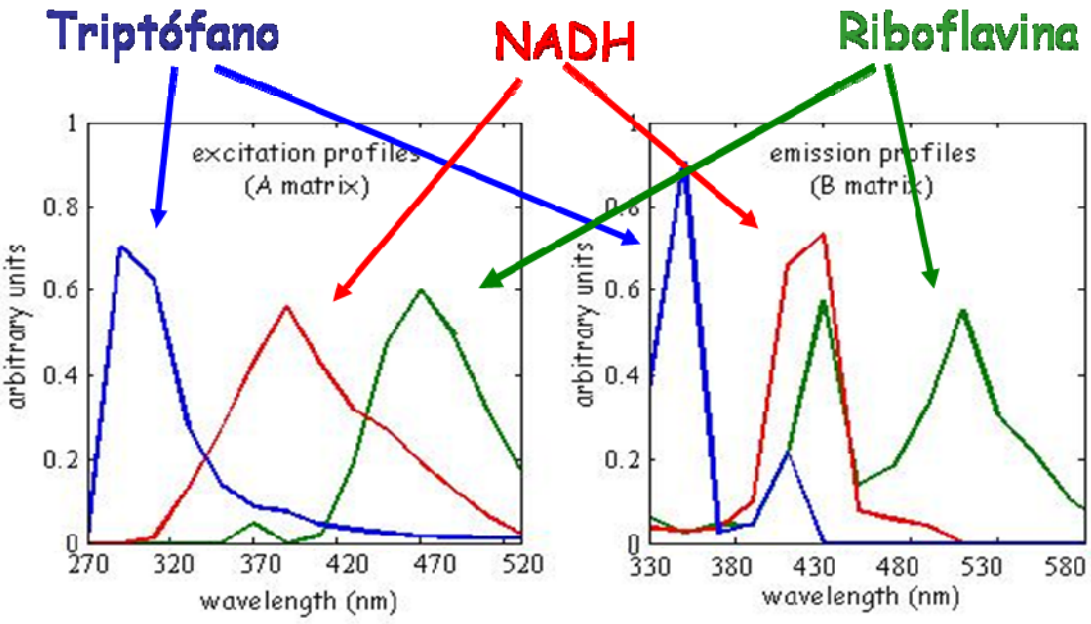


Figura 5.16: *Loadings* espectrales de excitación y emisión obtenidos para el modelo de calibración.

La figura 5.17 muestra los *loadings* de concentración obtenidos para cada componente. Estos perfiles indican la evolución del bioproceso a lo largo del tiempo. Teniendo en cuenta su forma, se pueden diferenciar tres zonas:

1) **Periodo latente:** Durante las primeras 4 horas, los perfiles de concentración de los fluoróforos permanecen constantes. El microorganismo se está adaptando a las condiciones del medio de cultivo en el bioreactor.
2) **Fase exponencial:** Después del periodo de adaptación, las intensidades de los perfiles de concentración empiezan a crecer. En este periodo, la levadura está metabolizando exponencialmente al metanol para generar biomasa, energía y otros productos secundarios. La relación señal-ruido aumenta debido a la formación de los fluoróforos relacionados con la generación de biomasa.
3) **Fase estacionaria:** La total desaparición del metanol y el máximo en la producción de lipasa coinciden con el comienzo de la fase estacionaria. Una

vez que el metanol se metaboliza totalmente del medio, se produce un descenso repentino de la señal del NADH y, consecuentemente, los perfiles de concentración del los tres fluoróforos permanecen constantes a lo largo del tiempo.

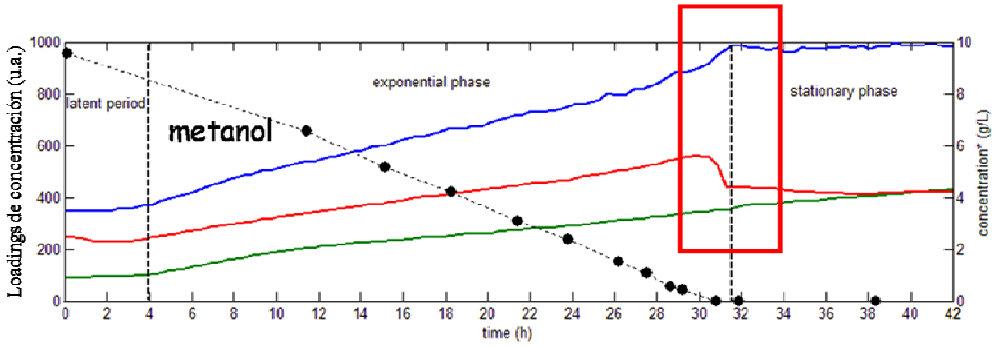


Figura 5.17: Loadings de concentración obtenidos para el modelo de calibración. Triptófano, azul. NADH, rojo. Riboflavina, verde. La línea discontinua de puntos negros representa la evolución de la concentración off-line de metanol. El eje de la izquierda corresponde a los valores obtenidos de los *loadings* de concentración. El eje de la derecha corresponde a la concentración de metanol.

Esta evolución de los perfiles de concentración se comparó con las medidas off-line de biomasa, metanol y producción de lipasa para verificar el correcto desarrollo del bioproceso y la veracidad del modelo PARAFAC. La información obtenida con los fluoróforos se corroboró con las medidas off-line de metanol (figura 5.17).

5.8.2. MONITORIZACIÓN EN TIEMPO REAL DE UN BIOPROCESO

Tal y como se ha detallado, los valores de los residuales del modelo PARAFAC de calibración se han usado para establecer los dos límites de control del nuevo bioproceso muestra a muestra. En la figura 5.18 se muestra la evolución para el nuevo proceso **P**. En la parte superior se muestran los perfiles de concentración obtenidos junto a medidas off-line obtenidos para comparar y contrastar las conclusiones. La parte inferior muestra los valores de *Qn* obtenidos en la monitorización del nuevo proceso. Como se puede observar, el valor *Qn* al inicio del proceso es ligeramente superior al primer límite de control (95%). Esta observación podría indicar que algún hecho

inesperado está sucediendo en el bioproceso. Sin embargo, se debe señalar que el proceso acaba de empezar, por lo que la relación señal-ruido es todavía baja. El periodo latente depende de muchos factores, tales como la inoculación del cultivo o la concentración de biomasa. Estos factores pueden hacer variar la etapa latente entre los diferentes cultivos, por lo que el poder de predicción en esta etapa puede ser bajo. El periodo latente permanece durante 4 horas, aproximadamente.

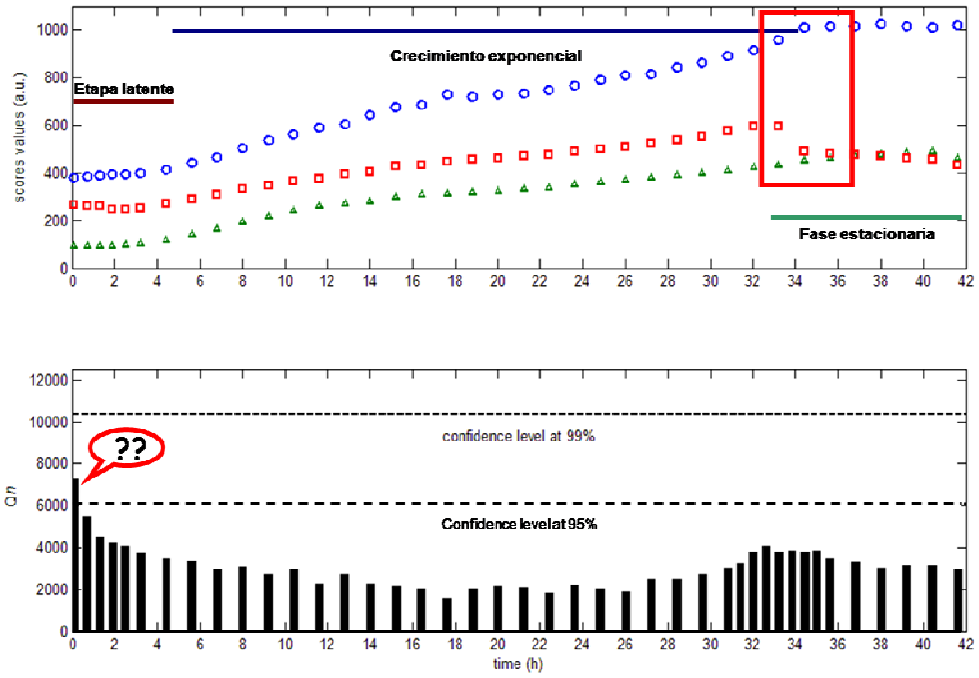


Figura 5.18: *Loadings* de concentración y valores de *Qn* obtenidos con PARAFAC en el control a tiempo real del Segundo bioproceso. (Triptófano, azul; NADH, rojo; riboflavina, verde)

Durante la fase de crecimiento exponencial, los valores de *Qn* permanecen por debajo del primer límite de control. Este hecho indica que el bioproceso se encuentra bajo condiciones de control.

La muestra recogida después de 33 horas (figura 5.18) muestra un valor de perfil de concentración para el NADH claramente inferior que el anterior, mientras que el valor para el triptófano llega a un valor constante. Este hecho puede indicar que el bioproceso se está acabando, coincidiendo con la desaparición del substrato. El tiempo de

adquisición de muestras se acortó. La siguiente medida certifica que la señal del NADH decrece mientras que las señales del triptófanos y la riboflavina permanecen constantes. Observando los valores de Q*n* para esta muestra, se puede considerar que el cultivo está bajo condiciones controladas. Esto indica que el bioproceso está terminando. En este punto, las posibilidades son realimentar el reactor o finalizar el bioproceso.

Si se comparan los resultados obtenidos en la monitorización a tiempo real con PARAFAC con las medidas *off-line* realizadas (figura 5.19) se observa la total coincidencia de la desaparición del substrato y el máximo de la producción de lipasa con el decrecimiento en la señal del NADH y el inicio de señal constante del triptófano. Este hecho indica la fiabilidad del uso de los *loadings* de concentración obtenidos con PARAFAC para el control a tiempo real del final de bioprocesos, sin necesidad de realizar medidas *off-line*.

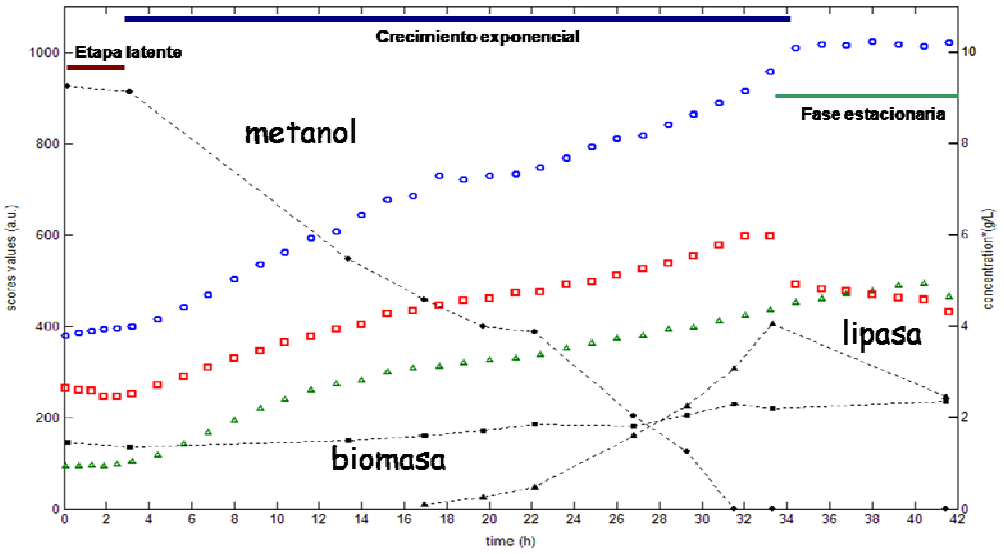


Figura 5.19: Comparación de los loadings de concentración obtenidos en el bioproceso controlado a tiempo real con las medidas off-line (Triptófano, azul; NADH, rojo; riboflavina, verde). Las líneas punteadas corresponden a las medidas *off-line* (Metanol, círculo; biomasa, cuadrado; lipasa, triángulo). El eje de la izquierda corresponde a los *loadings* de concentración obtenidos con PARAFAC. El eje de la derecha corresponde a las concentraciones *off-line*.

5.9. CONCLUSIONES DE ESTE TRABAJO

En este trabajo se presenta una nueva metodología de trabajo usando PARAFAC aplicado al control y diagnóstico de fallos de bioprocesos. La mayor novedad de esta metodología radica en la nueva consideración de los datos recogidos usando la fluorescencia multicanal, desde el punto de vista que la matriz bidimensional obtenida para cada tiempo de muestreo ($\lambda_{ex} \times \lambda_{em}$) se puede visualizar como una matriz tridimensional donde la primera dimensión es el tiempo ($1 \times \lambda_{ex} \times \lambda_{em}$).

Se ha demostrado que PARAFAC es una herramienta fácil de usar y adecuada para extraer la información relevante de sistemas tridimensionales de fluorescencia. Este algoritmo ha permitido la monitorización y el control a tiempo real de bioprocesos. La descomposición PARAFAC del bioproceso de calibración, **M**, ha ayudado a la identificación de los fluoróforos relacionados con el crecimiento del microorganismo. Se han identificado las tres fases de la evolución del bioproceso. La información espectral obtenida se ha utilizado para construir un modelo adecuado para la monitorización a tiempo real de diferentes bioprocesos sin necesidad de disponer de concentraciones de referencia de biomasa, metanol y crecimiento de lipasa.

Con esta metodología se ha establecido un parámetro de decisión. El descenso en la señal del NADH se ha relacionado con la metabolización total del substrato. Con este parámetro, se ha podido decidir si se realimenta el reactor para seguir con el proceso o se para la reacción.

La determinación *on-line* del parámetro estadístico *Qn* ofrece la ventaja de certificar que el bioproceso se encuentra en todo momento bajo condiciones normales de operación (NOC).

Un hecho que se ha de destacar es que el tiempo requerido para realizar las medidas y la obtención del valor de los perfiles de concentración y su correspondiente

valor de *Qn* no excede de los 5 minutos. Por lo que se obtiene una metodología de trabajo que permite la actuación directa a tiempo real.

También se ha demostrado que los modelos de calibración PARAFAC se pueden construir solamente con un bioproceso completo registrado en condiciones NOC. Pero, para asegurar la robustez del modelo, la metodología de trabajo se puede adaptar para usar varios procesos NOC en la construcción del modelo de calibración.

5.10. MONITORIZACIÓN USANDO GLICEROL COMO SUSTRATO

En el trabajo presentado en el Anexo 1 esta metodología de trabajo se aplica a bioprocesos donde se utiliza glicerol como sustrato. En este caso se utilizan dos bioprocesos para crear el modelo de calibración PARAFAC y establecer los límites de control. Los resultados obtenidos son similares a los obtenidos en la aplicación de la metodología a bioprocesos usando metanol como sustrato, corroborando la gran aplicabilidad de la metodología presentada para bioprocesos realizados en laboratorio. En procesos industriales se haría necesario la realización de un modelo de calibración más robusto. Esto es, utilizando un mayor número de bioprocesos que reflejen la mayor variabilidad existente.

PARTE III:

INTRODUCCIÓN A LA RESOLUCIÓN MULTIVARIABLE DE CURVAS: ESTUDIO ESPECTROFOTOMÉTRICO DE LOS EQUILIBRIOS ÁCIDO BASE DEL ÁCIDO 8-HIDROXIQUINOLINA-5-SULFÓNICO

5.11. INTRODUCCIÓN

El tratamiento de los datos el Química Analítica ha experimentado una rápida evolución en los últimos 15 años. Estos avances se han utilizado en laboratorios de investigación y en la industria, donde la Quimiometria ha empezado a establecerse como una herramienta de análisis de datos necesaria. Es por esto que la introducción de la Quimiometria en los planes de estudio universitarios es cada vez más común. Los estudiantes de Química y, particularmente, aquellos que se especialicen en Química Analítica, deben adquirir conocimientos básicos de diseño experimental, PCA, calibración multivariable y reconocimiento de pautas **[Kellner 2004, Howery 1983, Delaney 1981]**.

La monitorización de un proceso en un reactor industrial no es tan diferente a la monitorización de un proceso de laboratorio donde, fijadas unas condiciones iniciales, el sistema evoluciona en función de una variable continua (tiempo, pH, temperatura, etc.). El desarrollo de metodologías quimiométricas para el tratamiento de este tipo de señales evolutivas es uno de los aspectos principales en la investigación quimiométrica actual **[Tauler 1995, Amigo 2006]**. Muchas de las metodologías propuestas derivan de la resolución Multivariable de Curvas-Mínimos Cuadrados Alternos, MCR-ALS

(*Multivariate Curve Resolution-Alternating Least Squares*) **[Tauler 1995b]**. Tal y como se ha visto en capítulos anteriores en esta memoria (capítulo 3, sección 3.7) ésta es una técnica de descomposición matemática cuya finalidad es la obtención los perfiles espectrales y de concentración de las sustancias absorbentes presentes en un determinado sistema evolutivo. La posibilidad de obtener esta información de un sistema evolutivo la convierte en una metodología interesante y con un gran potencial en el campo de la industria de procesos y en investigación.

5.12. OBJETIVOS

En este contexto, el objetivo de este trabajo es ilustrar e introducir la metodología MCR-ALS a estudiantes de cursos superiores de la licenciatura de Química. Para ello, se monitorizó la evolución del espectro UV-Vis del ácido 8-hidroxiquinolina-5-sulfónico (8HQS) con el pH (figura 5.20). Se escogió el 8HQS debido a que es un sistema ácido-base bien conocido compuesto por dos tres especies ácido-base (dos constantes de acidez) las cuales absorben en un rango de longitudes de onda similar.

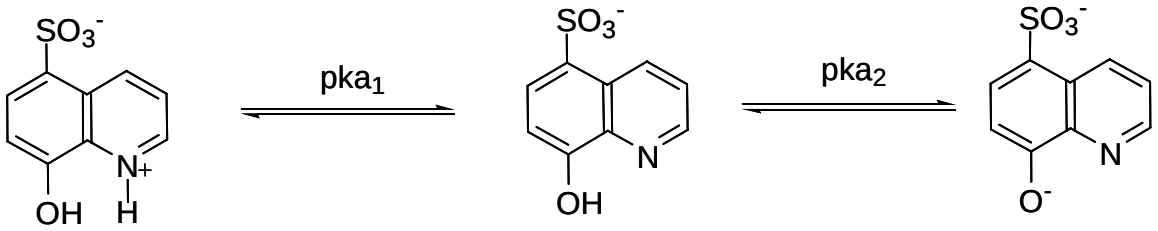


Figura 5.20: Equilibrio ácido-base del 8HQS. $pKa_1 = 3.80$ y $pKa_2 = 8.20$ **[Smith 1975]**.

Las constantes de acidez del sistema (protonación del nitrógeno y desprotonación del grupo fenol) se encuentran bien establecidas en la literatura **[Smith 1975, Beltrán 1993]**. Además, sus valores se encuentran lo suficientemente espaciados como para obtener el perfil espectral puro de la especie neutra (intermedio del proceso) y poder compararlo con el perfil espectral obtenido mediante MCR-ALS.

Otro de los objetivos de este trabajo es la introducción de los estudiantes al trabajo con un programa como MatLab **[Mat]** a través de la generación de una completa *demo* donde se explica, paso a paso, el tratamiento de matrices y la aplicación del MCR-ALS **[O'Haver 1989, Chau 1995]**.

5.13. RESULTADOS Y DISCUSIÓN

5.13.1. APLICACIÓN DE MCR-ALS

Los espectros registrados para la valoración se convirtieron en formato ASCII y se agruparon en una matriz **M** (J x K). En esta matriz, los j espectros registrados a k longitudes de onda se colocaron en filas en función del pH (figura 3.8).

5.13.2. RESULTADOS Y DISCUSIÓN

El algoritmo MCR-ALS produjo una estimación de los perfiles espectrales y de concentración de todas las especies absorbentes del sistema (figura 5.21).

Se ha de demostrar la veracidad de estos resultados para convencer a los estudiantes de la potenciabilidad de esta metodología cuando se estudian procesos cuya evolución es desconocida. A parte de la simple comparación visual de los resultados, se calculó el coeficiente de correlación entre el espectro puro y el perfil espectral obtenido mediante MCR-ALS para el intermedio de la valoración (especie neutra del 8HQS). También se calcularon cualitativamente los valores de las constantes de acidez y se compararon con los valores que se pueden encontrar en la literatura. De esta manera, se pudieron obtener los suficientes parámetros cualitativos para determinar si la descomposición MCR-ALS estaba ofreciendo un resultado fiable y realista de la evolución del sistema (tabla 5.13).

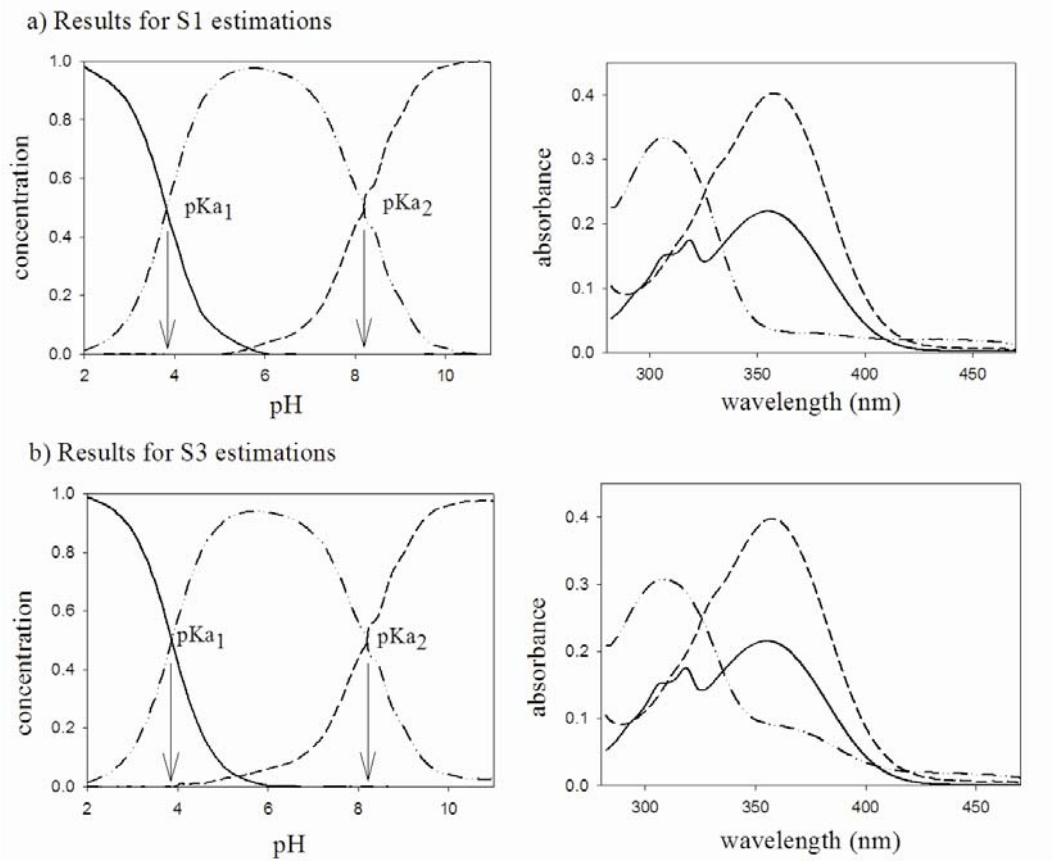


Figura 5.21: Perfiles de concentración y espectrales obtenidos mediante MCR-ALS. a) Estimaciones iniciales S1 y b) Estimaciones iniciales S3. Las flechas muestran las constantes de acidez cualitativas obtenidas.

Tabla 5.13: Resultados obtenidos en la aplicación del MCR-ALS a la valoración ácido base del 8HQS.

MCR-ALS	Estimaciones iniciales	valores de pH[a]			pKa1	pKa2	correlación[b]
	S1	1.95	5.98	10.96	3.80	8.22	0.9998
	S2	2.54	6.42	9.04	3.79	8.22	0.9983
	S3	1.95	3.10	10.96	3.79	8.22	0.9917
	S4	1.95	8.61	10.96	3.78	8.22	0.9996
[Beltrán 1993]					3.80	8.35	
[Smith 1975]					3.80	8.20	

[a] pH de los espectros usados como estimaciones iniciales.
[b] Coeficientes de correlación correspondientes al intermedio de la valoración.

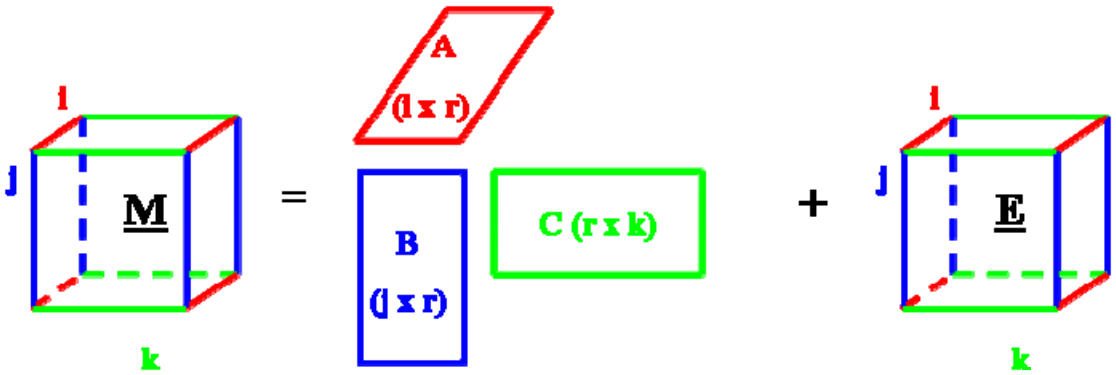
Se usaron espectros experimentales como estimaciones iniciales como punto de inicio de la descomposición matemática. El resultado del MCR-ALS se puede ver

afectado por esta selección, por lo que se usaron diferentes combinaciones de espectros como estimaciones iniciales (S1-S4 en tabla 5.13). Las estimaciones S1 corresponden a las estimaciones iniciales "lógicas" realizadas con un estudio visual de los datos espectrofotométricos y de pH: el primer y último espectro (las mejores estimaciones de la primera y la tercera especie) y un espectro central para la especie intermedia. Los resultados obtenidos para las estimaciones S2-S4 muestran el efecto de utilizar estimaciones iniciales diferentes a las más adecuadas. En la tabla 5.13 se puede observar que en todos los casos se obtuvieron buenas correlaciones y buenos valores de las constantes de acidez. En la figura 5.21 se muestran los resultados para el mejor (S1) y para el peor (S3) ajuste.

5.14. CONCLUSIONES A ESTE TRABAJO

En este trabajo se ha visto que el MCR-ALS es una herramienta de la que se puede obtener una gran cantidad de información de un sistema químico evolutivo sin mucho conocimiento previo sobre la evolución del sistema. En este caso, los estudiantes sólo conocen inicialmente que el sistema es un equilibrio ácido-base. Ellos mismos tienen que investigar sobre el número de especies absorbentes que componen el sistema, tienen que escoger las estimaciones iniciales y tienen que imponer restricciones adecuadas. De una manera sencilla obtendrán la evolución en la concentración de cada una de las especies, así como una estimación de sus perfiles espectrales. Además, mediante un tratamiento gráfico adecuado (por ejemplo, derivadas de los perfiles de concentración) obtendrán una estimación cualitativa de las constantes de acidez del sistema.

Es evidente que cuanto mayor sea el conocimiento previo del que se disponga de un sistema, más fácil resultará su interpretación. Sin embargo, los estudiantes aprenderán que es posible obtener información de calidad aplicando MCR-ALS incluso cuando se dispone de escasa información acerca del sistema evolutivo.



Capítulo 6

CONCLUSIONES GENERALES

En esta Memoria se ha demostrado la utilidad de diversas metodologías quimiométricas multidimensionales a la hora del estudio de sistemas enzimáticos. Las conclusiones generales más destacables a las que se ha llegado son:

MCR-ALS ha demostrado ser un algoritmo muy útil para estimar los perfiles espectrales y la evolución con el tiempo de la concentración de las diferentes especies reaccionantes, sustratos o productos. La introducción en este algoritmo de la restricción del modelo cinético del sistema estudiado ha dado lugar a una nueva metodología de trabajo, denominada por su acrónimo en inglés HS-MCR-ALS. Esta nueva metodología de trabajo, aplicada al seguimiento de las reacciones sucesivas de oxidación de hipoxantina, xantina y ácido úrico a través de su catálisis con la xantina oxidasa ha hecho posible:

· El estudio del sistema enzimático en presencia de interferentes espectrales.

· La correcta elucidación del modelo cinético, ya que el anteriormente descrito en la bibliografía no contemplaba la degradación del ácido úrico.

· La determinación de las constantes cinéticas del modelo postulado, tanto en un medio tamponado de composición conocida como en orina.

Se han aplicado y comparado diferentes métodos, bi o tridimensionales, de análisis multivariable cuantitativo para la determinación de xantina, hipoxantina y ácido úrico en mezclas sintéticas y en orina humana dopada. Los resultados obtenidos en la aplicación cuantitativa del HS-MCR-ALS son comparables a los resultados obtenidos

mediante otros algoritmos tridimensionales, como son el 3W-PLS1 y 3W-PLS2, y a los algoritmos bidimensionales clásicos, PLS1 y PLS2. La gran ventaja de la aplicación de HS-MCR-ALS respecto a las otros métodos radica en la mínima cantidad de muestras patrón necesarias para cuantificar los diferentes analitos en muestras problema (sólo hicieron falta dos patrones), en que no es necesario conocer o incluir en el calibrado las interferencias y a la gran cantidad de información adicional que aporta (perfiles espectrales, de concentración y constantes enzimáticas).

Se ha establecido una nueva metodología de trabajo para el control en tiempo real de procesos de generación de enzimas. Se basa en el seguimiento no invasivo de la reacción mediante una sonda de fluorescencia multidimensional y en el tratamiento de la señal registrada mediante el algoritmo PARAFAC. Aunque se ha aplicado al caso concreto de la generación de enzimas tipo lipasa en cultivos de *Pichia pastoris*, esta metodología de trabajo se puede aplicar a un mayor número de reacciones. La fluorescencia multidimensional aporta una gran cantidad de información en un breve espacio de tiempo, por lo que toda la variabilidad relacionada con los fluoróforos queda recogida. A partir del modelo PARAFAC, aplicado a unos bioprocesos desarrollados en condiciones normales, se obtiene una estimación de los perfiles espectrales de los fluoróforos y de la evolución de su señal. El análisis de residuales de este modelo permite establecer unos límites de control (se ha utilizado el criterio Q). A partir de esa información, se monitorizó y controló en tiempo real un nuevo lote del bioproceso, siendo capaces de determinar el punto final de la producción de lipasas y metabolización total de sustrato sin necesidad de medidas *off-line* de los analitos de interés.

Paralelamente, se ha puesto a punto una práctica de laboratorio para la introducción del algoritmo MCR-ALS a los estudiantes de cursos superiores de la licenciatura de Ciencias Químicas. El sistema estudiado ha sido la evolución espectrofotométrica del ácido 8-hidroxiquinolina-5-sulfónico con el pH. Esta práctica contiene instrucciones de cómo utilizar los datos obtenidos en la valoración para poder aplicar MCR-ALS de una manera sencilla y didáctica. El carácter didáctico de esta práctica trasciende mucho más de lo puramente quimiométrico, ya que el alumno aprende

a realizar valoraciones ácido-base para la determinación de constantes de acidez, así como también aprende el funcionamiento básico del programa MatLab®, introduciéndolo en el tratamiento y manejo de datos. La práctica se compone de una introducción, instrucciones de cómo realizar la valoración, instrucciones de cómo aplicar MCR-ALS y un conjunto de datos para que el alumno pueda aplicar MCR-ALS sin necesidad de realizar la valoración. Todo este material se encuentra disponible en la Web **[Web 4.2]**, o a través de los autores **[Rodríguez-Rodríguez 2007]**.

Chapter 6

CONCLUSSIONS

In this memory, the usefulness of several multidimensional chemometric methodologies for the study of enzymatic systems has been demonstrated. The achievements of this thesis can be summarized as follow:

MCR-ALS has demonstrated to be a useful algorithm to obtain spectral profiles and the evolution with time of the concentrations of the different species involved in the reactions, either products or substrates. The introduction in the algorithm of a restriction related with the kinetic model of the system gives raise to a new method, combining hard-modelling and soft-modelling in the HS-MCR-ALS algorithm. This new methodology has been applied to the consecutive reactions of hypoxanthine, xanthine and uric acid being catalyzed by xanthine oxidase and the following results have been obtained:

· The study of the enzymatic mechanism in presence of a spectral interference.

· The correct elucidation of the kinetic model, since the previously described in the bibliography didn't contemplate the degradation of the uric acid.

· The determination of the enzymatic constants of the proposed kinetic model in presence and absence of the interference of urine.

Several methods of quantitative multivariable analysis, bi or tridimensionals, have been applied to the quantitation of hypoxanthine, xanthine and uric acid in synthetic samples and urine and their results compared. The application of mixed hard- and soft-modelling resolution methodologies has proven to be a useful tool for a complete

quantitative and kinetic description of samples of human urine containing hypoxanthine, xanthine and uric acid. This example shows that HS-MCR-ALS can work for quantitative purposes in any multicomponent system, where the analytes evolve according to a known enzymatic or kinetic process in the presence of absorbing interferences. The obtained results were compared with those obtained by other three-way methodologies (3W-PLS1 and 3W-PLS2) and by the classical two-way methodologies (PLS1 and PLS2). The HS-MCR-ALS algorithm shows clear advantages, such as the small amount of standards needed (one or two standard samples are enough) and the fact that they can be prepared in aqueous solution, with no need to know or include the interferences present in the samples. Furthermore, the kinetic profiles and the spectra of the compounds involved are obtained.

A new working methodology applied to the control and fault diagnosis of fermentation processes in real time is presented. This new methodology is based on the non-invasive monitoring of the reaction by means of a multidimensional fluorescence probe and the treatment of the recorded information with PARAFAC algorithm. Although it has been applied to the formation of lipase enzymes by means of the growth of *Pichia pastoris* micro-organism, the method may be applied to many reactions monitorized by fluorescence. Multidimensional fluorescence collects a great amount of information in a short period of time, so the variability related with the fluorophores is recorded. From the PARAFAC model, built with bioprocesses in normal operating conditions, an estimation of the spectral profiles of the fluorophores and the evolution of their signal is obtained. Residual analysis of the model allows establishing control limits (using Q criterion). From this information, a new batch of the culture was monitored in real-time without needing reference off-line measurements.

A laboratory experiment has been proposed to introduce the algorithm MCR-ALS to the upper-level students of the Chemistry Degree. The experiment includes a laboratory session were the spectrophotometric evolution of 8-hydroxyquinolin-5-sulfonic acid with pH is recorded, and other session were the students apply the MCR-ALS algorithm to the recorded data. This experiment contains instructions in how to use

the obtained data using MCR-ALS in a simple and didactic manner. The scope of the practice surpasses chemometrics and the students learn how to use acid-base titrations to calculate protonation constants and the basis of Matlab® software. A complete demonstration and instructions for the students are available **[Web 4.2, Rodríguez-Rodríguez 2007]**.

Capítulo 7

REFERENCIAS / REFERENCES

[Amigo 2005] J.M. Amigo, J. Coello, S. Maspoch. *Anal. Bioanal. Chem*. 382 (2005) 1380-1388.

[Amigo 2006] J.M. Amigo, A. de Juan, J. Coello, S. Maspoch. *Anal. Chim. Acta* 567 (2006) 245-254

[Amigo 2006b] J.M. Amigo, A. de Juan, J. Coello, S. Maspoch. *Anal. Chim. Acta* 567 (2006) 236-244

[Amrhein 1996] M. Amrhein, B. Srinivasan, D. Bonvin, M.M. Schumacher. *Chemom. Intell. Lab. Syst.* 33 (1996) 17-33.

[Andersen 2003] C.M. Andersen, R. Bro. *J. Chemom.* (2003) 17, 200-215

[Beltrán 1993] J.L. Beltrán, R. Codony, M.D. Prat, *Anal. Chim. Acta*. 276:2 (1993) 441.

[Blanco 1999] M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, M. Porcel, Anal. Chim. Acta 398 (1999) 83-92

[Boiani 1986] J.A. Boiani, *J. Chem. Educ.* 63:8 (1986) 724.

[Bro 1996] R. Bro. *J. Chemom* (1996) 10:47-61

[Bro 1997] R. Bro. *Appl. Spectrosc. Reviews* (1997) 32-3:237-261

[Bro 1997b] R. Bro. *Chemom Intell Lab Syst* 38 (1997) 149-171

[Bro 1998] R. Bro. "Multi-way Analysis in the Food industry. Models, Algorithms and Applications" PhD thesis dissertation (1998)

[Bro 1998b] R. Bro, N.D. Siidiripoulos. *J. Chemom.* 12 (1998) 223-247

[Booksh 1994] K. S. Booksh y B. R. Kowalski. *Anal. Chem*. 782A (1994) 66

[Carrol 1970] J.D. Carrol, J. Chang. *Psychometrika* 35 (1970) 283.

[Carsol 1998] M-A. Carsol, M. Mascini, *Talanta* 47 (1998) 335-342

[Chau 1995] F.T. Chau, W.H. Chung, *J. Chem. Educ.* 72:4 (1995) 78.

[Chen 2002] G. Chen, Q. Chu, L. Zhang, J. Ye, *Anal. Chim. Acta* 457 (2002) 225-233

[Clark 1979] P.M.S. Clark, L.J. Kricka, A. Patel, *Anal. Lett.* 12:B15 (1979) 1537.

[Coello 2000] J. Coello, S. Maspoch, N. Villegas, *Talanta* 53 (2000) 627-637

[Copeland 1996] Robert A. Copeland, "Enzymes. A Practical Introduction to Structure, Mechanism and Data Análisis", VCH, 1996

[Crouch 2000] S.R. Crouch, J. Coello, S. Maspoch, M. Porcel. *Anal. Chim. Acta* 424 (2000) 115-126

[Cuesta 1994] F. Cuesta, D.L. Massart. *Anal. Chimi. Acta* (1994) 298, 331-339

[Czauderna 1997] M. Czauderna, J. Kowalczyk, *J. Chromatogr. B* 704 (1997) 89-98

[De Jong 1998] S. de Jong. *J. Chemom.* 12 (1998) 77-81

[De Jong 1993] S. de Jong. *Chemom. Intell. Lab. Syst.* 18 (1993) 251-263.

[De Juan 2000] A. de Juan, M. Maeder, M. Martinez, R. Tauler. *Chem. Intell. Lab. Syst.* 54 (2000) 123-141

Referencias / References

[De Juan 2000b] A. de Juan, E. Casassas, R. Tauler. "Enciclopedia of Analytical Chemistry: instrumentation and applications, in Soft-Modelling of Analytical Data" Wiley, New York (2000)

[De Juan 2003] A. de Juan, R. Tauler. *Anal. Chim. Acta*. 500:1-2 (**2003**) 195-210.

[Delaney 1981] F. Delaney, Jr. Warren, F. Vicent, *J. Chem. Educ.* 58 (1981) 646.

[Di Pietro 2001] M.C. Di Pietro, D. Vannoni, R. Leoncini, G. Liso, R. Guerranti, E. Marinello, *J. Chromatogr. B* 751 (2001) 87-92

[Diework 2003] J. Diework, A. de Juan, M. Maeder, R. Tauler, B. Lendl. *Anal. Chem.* 75 (2003) 641-647

[Escribano 1988] J. Escribano, F. Garcia-Canovas, F. Garcia-carmona, *J. Biochem.* 254 (1988) 829-833

[Fersht 1980] Alan Fersht, "Estructura y Mecanismo de los Enzimas", Ed. Reverté, 1980

[Foppoli 1997] C. Foppoli, R. Coccia, C. Cini, M.A. Rosei, *Biochim. Biophys. Acta* 1334 (1997) 200-206

[Gacesa 1990] Meter Gacesa, John Hubble, "Tecnología de los enzimas", Ed. Acrilia, Zaragoza, 1990

[Geladi 1988] P. Geladi. *J. Chemom.* 2 (1988) 231-246.

[Gella 2003] F-Javier Gella, "Enzimología clínica", ByoSystems, 2003

[Gemperline 1986] P.J. Gemperline. *Anal. Chem.* 58 (1986) 2656-2663

[Gemperline 1989] P.J. Gemperline, C. Hammilton. *J. Chemom.* 3 (1989) 455-461

[Gemperline 2003] P.J. Gemperline, E. Cash. *Anal. Chem.* 75:16 (2003) 4236 – 4243

[Gran 1952] G. Gran, *Analyst*. 77 (**1952)** 661.

[Gray 1971] C. J. Gray, "Enzyme-Catalysed Reactions", V. N. Reinhold Company, London, 1971

[Haario 1998] H. Haario, V.M. Taavitsainen. *Chem. Intell. Lab. Syst.* 44 (1998) 77-98

[Harshman 1970] R.A. Harshman "Foundations of the PARAFAC procedure: Model and conditions for an 'explanatory' multi-mode factor analysis" *UCLA Working papers in phonetics* 16 (1970) 1

[Harshman 1984] R.A. Harshman, M.E. Lundy "Research methods for multimode data analysis" Praeger, New York (1984)

[Harshman 1994] R.A. Harshman, M.E. Lundy. *Comp. Stat. Data Anal.* 18 (1994) 39

[Helland 1988] I.S. Helland "On the structure of Partial Least-Squares Regression" *Communications in Statistics-Simulation and Computation* 17 (1988) 581-607.

[Hisinger 2005] Hisiger, S., Jolicoeur, M., *J. Biotechnol.* 117 (2005) 325-326.

[Höskuldsson 1988] A. Höskuldsson. *J. Chemom.* 2 (1988) 211-228

[Howery 1983] G. Howery, F. Roland, *J. Chem. Educ.*60 (1983) 656.

[Jackson 1980] J.E. Jackson. *J. of Quality Tech.* 12 (1980) 201-213

[Jaumot 2005] J. Jaumot, R. Gargallo, A. de Juan, R. Tauler. *Chem. Intell. Lab. Syst.* 76:1 (2005) 101.

[Jeżewska 1973] M.M. Jeżewska, *Eur. J. Biochem.* 36 (1973) 385-390

[Johnson 2002] G.W. Johnson, R. Ehrlich. *Environm. Forens.* 3 (2002) 59-79

[Kermis 2002] H.R. Kermis, Y. Kostov, P. Harms, K. Rao, *Biotechnol. Prog.* 18 (2002) 1047.

[Kellner 2004] R. Kellner, J.M. Mermet, M. Otto, M. Varcárcel, H.M Widmer, "Analytical Chemistry", Wiley-VCH; Weinheim, 2004.

[Kiers 1991] H.A.L. Kiers HAL *Psychometrica* 56 (1991) 449-470

[Kruskal 1984] J.B. Kruskal "Methods for Multimode Analysis", H.G. Law, C.W. Snynder, J.A. Hattie, R.P. McDonals (Editores), Praeger, New York (1984)

[Kruskal 1978] J.B. Kruskal "Factor analysis and principal components" International Encyclopedia of Statistics, Ed. The Free Press, New York (1978)

[Lindemann 1998] C. Lindemann, S. Marose, HO. Nielsen, T. Scheper, *Sens Actuators B: Chem.* 51 (1998) 273-277

[Lindsey 1981] A.S. Lindsey, R.K. Sharma, *Anal. Lett.* 14:B10 (1981) 799.

[Louwerse 2000] DJ. Louwerse, AK. Smilde, *Chem. Eng. Sci.* 55 (2000) 1225-1235.

[Maeder 1987] M. Maeder. *Anal. Chem.* 59 (1987) 527-530

[Maeder 1990] M. Maeder, A.D. Zuberbühler. *Anal. Chem.* 62 (1990) 2220-2224

[Marquardt 1963] D.W. Marquardt, *J. Soc. Ind. Appl. Math.* 11 (1963) 431.

[Massey 1969] V. Massey, P.E. Brumby, H. Komai, G. Palmer, *J. Biol. Chem.* 244-7 (1969) 1682-1691

[Massart et al. 1997] D.L. Massart, B.G.M Vandeginste, L.C.M. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke. Handbook of Chemometrics and Qualimetrics: Part A. 1st edition, ed. Elsevier (1997)

[Mat] MatLab® versions 6.0, 6.5, 7.0, The MatWorks, Inc. Massachusetts

[McWhirter 1991] R.B. McWhirter, R.J. Hille, *Biol. Chem.* 266-35 (1991) 23724.

[Menrion 1994] R. Henrion. *Chemom. Intell. Lab. Syst* 25 (1994) 1-23

[Minning 2001] S. Minning, A. Serrano, P. Ferrer, C. Solà, R.D. Schmid, F. Valero, *J Biotechnol.* 86 (2001) 59.

[Msimanga 2005] Z. Msimanga, Z. Elkins, K. Segmia, R. Smith, *J. Chem. Educ.* 82 (2005) 415.

[Nelson 2004] David L. Nelson, Michael M. Cox, "Lehninger Principles of Biochemistry", 4th edition, Freeman & Co., 2004

[Ni 2004] Y. Ni, C. Huang, S. Kokot, *Chemom. Intell. Lab. Syst.* 71 (2004) 177-193

[Núñez 2001] Ignacio Núñez de Castro, "Enzimología", Ed. Pirámide Madrid, 2001

[O'Haver 1989] T.C. O'Haver, *Chemom. Intell. Lab. Sys.* 6 (1989) 95.

[Olson 1974] J.S. Olson, D.P. Ballou, G. Palmer, V. Massey, *J. Biol. Chem.* 249-14 (1974) 4363.

[Olson 1974b] J.S. Olson, D.P. Ballou, G. Palmer, V. Massey, *J. Biol. Chem.* 249-14 (1974) 4350.

[Otto 1999] M. Otto "Chemometris. Statistics and Computer Application in Analytical Chemistry" Ed. Wiley-VCH, New-York (1999)

Referencias / References

[Palmer 1985] Trevor Palmer, "Understanding Enzymes", 2nd Ed., John Wiley & Sons Inc., 1985

[Pei 2000] J. Pei, XY. Li, *Anal. Chim. Acta* 414 (2000) 205-213

[Pettersson 1997] Å. Pettersson, B. Karlberg, *Anal. Chim. Acta* 354 (1997) 241-248

[Phatak 1997] A. Phatak, S.J. De Jong. *J. Chemom.* 11 (1997) 311-338

[PLSTOOL] PLS-Toolbox, versiones 3.0 y 3.5, Eigenvector Research, WA, USA

[Putterman 1979] G.J. Putterman, B. Shaikh, M.R. Hallmark, C.G. Sawyer, C.V. Hixson, F. Perini, *Anal. Biochem.* 98 (1979) 18-26

[Rodríguez-Rodriguez 2007] C. Rodríguez-Rodríguez, J.M. Amigo, J. Coello, S. Maspoch "An introduction to Multivariate Curve Resolution-Alternating Least Squares. Spectrophotometric study of the acid-base equilibria of 8-hydroxyquinoline-5-sulfonic acid" Aceptada su publicación en *Journal of Chemical Education*, 2007

[Saurina 1998] J. Saurina, S. Hernández-Cassou, R. Tauler, A. Izquierdo-Ridorsa. *J. Chemom.* 12 (1998) 183-203

[Segel 1993] Tawin H. Segel, "Enzyme Kinetics. Behaviour & Análisis of Rapad Equilibrium and Steady-State Enzyme Systems", John Wiley & Sons Inc,, 1993

[Sharaf 1986] M.A. Sharaf, D.L. Illman, B.R. Kowalski "Chemometrics" Ed. John Wiley & Sons, New York (1986)

[Skibsted 2001] E. Skibsted, C. Lindemann, C. Roca, L. Olsson, J. Biotechnol. 88:1 (2001) 47.

[Smilde 1997] A.K. Smilde. *J. Chemom.* 11 (1997) 367-377

[Smilde 2004] A.K. Smilde, R. Bro, P. Geladi "Multi-way Analysis. Applications in the Chemical Sciences" Ed. John Wiley & Sons, Ltd. England (2004)

[Smith 1975] R.M. Smith, A.E. Martell, "Critical Stability Constants". Vol. 2, Plenum, New York, 1975.

[Stanimirova 2004] I. Stanimirova, B. Walczak, D.L. Massart, V. Simeonov. *Chemom. Intell. Lab. Syst.* 71 (2004) 83– 95

[Surribas 2006] A. Surribas, JM. Amigo, J. Coello, JL. Montesinos, F. Valero, S. Maspoch, *Anal. Bioanal. Chem.* 385 (2006) 1618-2642.

[Surribas 2006b] A. Surribas, G. Geissler, A. Gierse, T. Scheper, B. Hitzmann, J.L. Montesinos, F. Valero, *J. Biotechnol.* 124:2 (2006) 412-419

[Surribas 2006c] A. Surribas, J.L. Montesinos, F. Valero, *J. Chem. Technol. Biotechnol.* 81 (2006) 23.

[Tauler 1993] R. Tauler, D. Barceló, *Trends Anal. Chem.* 12 (1993) 319

[Tauler 1995] R. Tauler, A.K. Smilde, B.J. Kowalski. *J. Chemom*. 9 (1995) 31-58

[Tauler 1995b] R. Tauler. *Chemom. Intell. Lab. Syst.* 30 (1995) 133-146

[Tauler 1997] R. Tauler. Anàlisi de Mescles Mitjançant Resolució Multivariant de Corbes. Ed. Institut d'Estudis Catalans (1997)

[Tauler 2001] R. Tauler. *J. Chemom.* 15 (2001) 627-646.

[Tawa 1981] R. Tawa, M. Kito, S. Hirose, *Chem. Lett.* (1981) 745-748

[Tu 1992] X.M. Tu, D.S. Burdick. *Stat. Sinica* 2 (1992) 507

[UNS] The Unscrambler®, Camo, Norway

[Vandeginste 1985] B.G.M. Vandeginste, R. Esserst, T. Bosman, J. Reijnen, G. Kateman. *Anal. Chem.* 57 (1985) 971-985

[Winding 1988] W. Windig. *Chemom. Intell. Lab. Syst.* 4 (1988) 201-213.

[Winding 1992] W. Winding W. *Chemom. Intell. Lab. Syst.* 16:1 (1992) 1-16

[Wise 1996] B.M. Wise, N.B. Gallagher, *J. Proc. Cont.* 6 (1996) 329-348

[Wold 1972] S. Wold. *Kemisk Tidskrift* 84:3 (1972) 34-37

[Wold 1987] S. Wold, P. Geladi, K. Esbensen, J. Öhman. *J. Chemom* 1 (1987) 41-56

[Wold 1987b] S. Wold, K.H. Esbensen, P. Geladi. *Chemom. Intell. Lab. Syst*. 2 (1987) 37-52

[Wold 1995] S. Wold. *Chem. Intell. Lab. Syst.* 30:1 (1995) 109-115

[WoldH 1966] H. Wold. "Multivariate Analysis" Ed. Krishnaiah, P.R. Academic Press, New York (1996)

[Zen 2002] J-M. Zen, Y-Y. Lai, H-H. Yang, A. Senthil Kumar, *Sens. Actuators B* 84 (2002) 237-244

Enlaces web: (Todos los enlaces Web han sido visitados por última vez en mayo de 2007)

[Web 1.1] http://www.novozymes.com

[Web 1.2] http://www.genencor.com

[Web 1.3] http://www.tkk.fi/Units/BioprocessEngineering

[Web 1.4] http://en.wikipedia.org

[Web 1.5] http://rae.es

[Web 1.6] http://www.chem.qmul.ac.uk/iubmb/enzyme/

[Web 3.1] http://www.ub.es/gesq/mcr/mcr/.htm

[Web 3.2] http://personal.ecu.edu/gemperlinep/Research/gemper_software.html

[Web 3.3] http://www.models.kvl.dk

[Web 4.1] http://www.berkeleymadonna.com/

[Web 4.2] http://jchemed.chem.wisc.edu/

Referencias / References

ANEXOS / ANNEXES

Anexos / Annexes

ANEXO 1 / ANNEX 1

Artículos producidos en la tesis

Papers

Anexos / Annexes

Three-way Partial Least-Squares regression for the simultaneous kinetic-enzymatic determination of xanthine and hypoxanthine in human urine.

José Manuel Amigo, Jordi Coello, Santiago Maspoch.

Anal. Bioanal. Chem. 382 (2005) 1380-1388

Anexos / Annexes

Anal Bioanal Chem (2005) 382: 1380–1388
DOI 10.1007/s00216-005-3275-4

ORIGINAL PAPER

José Manuel Amigo · Jordi Coello · Santiago Maspoch

Three-way partial least-squares regression for the simultaneous kinetic-enzymatic determination of xanthine and hypoxanthine in human urine

Received: 22 February 2005 / Accepted: 22 April 2005 / Published online: 7 June 2005
© Springer-Verlag 2005

Abstract The performance of three-way principal component analysis and three-way partial least-squares regression when applied to a complex kinetic-enzymatic system is studied, in order to investigate the analytical potential of the combined use of these chemometric technologies for non-selective enzymatic systems. A enzymatic-kinetic procedure for the simultaneous determination of hypoxanthine and xanthine in spiked samples of human urine is proposed. The chemical system involves two consecutive reactions catalyzed by xanthine oxidase (EC 1.17.3.2). This enzyme catalyzes the oxidation of hypoxanthine, first to xanthine and then to uric acid, a competitive inhibitor of the reactions. The influence of uric acid during quantitative determination was considered in the design of the calibration set. The sample and enzyme solution were mixed in a stopped-flow module and the reaction was monitored using a diode array spectrophotometer. The recorded data have an intrinsical three-component structure (samples, time and wavelength). This data array was studied via three-way principal component analysis and was modeled for quantitative purposes using a three-way partial least-squares calibration procedure. Results are compared with those obtained by applying classical bilinear PLS to the previously unfolded data matrix.

Keywords Oxipurines · Urine analysis · Three-way principal component analysis · Three-way partial least-squares regression · Simultaneous kinetic determination

Introduction

While enzymes are usually visualized as highly selective catalysts and are used with the aim of obtaining a specific product, there are a number of them that, in fact, show group selectivity. The development of multivariate calibration procedures and their application to kinetics [1–4] has renewed analytical interest in non-selective enzymes, since they can potentially be applied to the simultaneous and rapid determination of several analytes using relatively simple instrumentation. The use of enzymes showing group selectivity becomes an analytical choice when applied to the kinetic resolution of similar compounds with the same functional groups. The potential for this kind of resolution has already been proved during the simultaneous determination of ethanol and methanol using an alcohol oxidase [5]. Since the same product is formed from both compounds, discrimination is only possible via kinetic information, and this non-linear system was solved by applying an artificial neural network to the data recorded at a single wavelength.

The system studied in this work is more complex in the sense that it consists of two consecutive reactions ($A \rightarrow B \rightarrow C$) catalyzed by the same enzyme, where A and B are the analytes and C, the product, acts as an inhibitor [6]. However, although the kinetic system is complex, there is a large amount of information available since all three compounds absorb in the ultraviolet region, although with highly overlapped absorption bands. Therefore, the spectral and kinetic data from the system could contribute useful information that may be used to discriminate the responses of the analytes, and therefore to resolve the mixtures.

The oxipurine (hypoxanthine and xanthine) and uric acid levels in blood, plasma and urine may provide sensitive indicators of certain pathologic states, including xanthinuria, gout, renal failure, toxaemia during pregnancy, and other diseases [7]. These compounds have similar chemical structures and UV–Vis spectra (Fig. 1a), which has prompted the use of analytical separation techniques such as high performance liquid chromatography (HPLC) [8–10] or high performance capillary electrophoresis (HPCE) [11] for their simultaneous analysis. In addition, several electrochemical methods have been proposed [12, 13]. Until now, the

J. M. Amigo · J. Coello (✉) · S. Maspoch
Universitat Autónoma de Barcelona,
Bellaterra, 08193 Barcelona, Spain
E-mail: jordi.coello@uab.es
Tel.: +34-93-5812122
Fax: +34-93-5812379

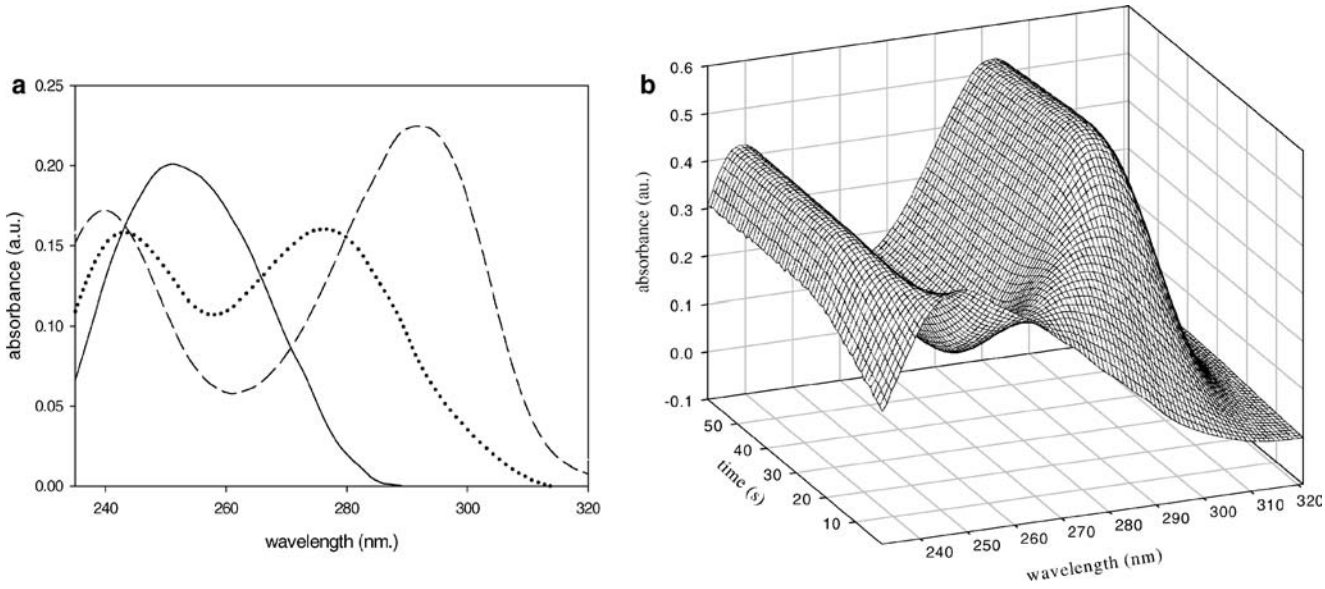


Fig. 1a–b **a** Pure spectra of xanthine (*dotted line* 30 μmol L^{-1}), hypoxanthine (*solid line* 60 μmol L^{-1}) and uric acid (*dashed line* 60 μmol L^{-1}). The important wavelengths are 238 and 293 nm (formation of uric acid), 249 nm (hypoxanthine transformation), 277 nm (xanthine transformation). **b** Kinetic profile of a mixture of hypoxanthine (30 μmol L^{-1}) and xanthine (15 μmol L^{-1}) reacting with 0.07 U mL^{-1} of xanthine oxidase

simultaneous determination of xanthine and hypoxanthine without any previous separation has not been performed due to the lack of spectral selectivity. This lack of spectral selectivity is not a problem for kinetic-chemometric methods, where information about the spectral evolution with time can be collected and several multivariate methods can be used.

The enzyme xanthine oxidase (EC 1.17.3.2) has been used as a group enzyme to develop methodologies for the determination of complex mixtures of analytes [14], as well as for hypoxanthine and xanthine [7, 13, 15]. In all of these cases, the enzyme was immobilized in a specific electrode or in a chromatographic pre-column.

Kinetic-spectrophotometric data have an intrinsic three-way structure (sample, time and wavelength). Application of bilinear partial least-squares (PLS) to multivariate kinetic data, prior to unfolding the data, procures a high predictive capability and versatility, being able to solve a large number of different kinetic models. However, the large number of variables (vectors with over 10,000 data points are easily obtained) containing both kinetic and spectral information, make difficult to extract the qualitative information needed to understand what is really happening during the reaction.

Several chemometric methodologies have been applied in order to decompose three-way data matrices [16]. Some of these methodologies are: tri-linear decomposition (TLD) [17], three-way principal components analysis (3W-PCA) [18, 19], parallel factor analysis (PARAFAC) [20], Tucker models [21], and three-way-multivariate curve resolution–alternating least-squares (3W-MCR-ALS) [22]. All of these methodologies are able to split the tri-linear systems into separate spectral and kinetic profiles, although each algorithm produces results that are affected by different restrictions and, in some cases, the requirement for linearity restricts their application. Three-way PLS (3W-PLS) was developed in order to obtain quantitative information on the systems without the need to unfold the data matrix, using an algorithm similar to the bilinear PLS used for two-dimensional data [23].

Despite the increasing use of these three-way methodologies, there is still very little real practical experience of applying 3W-PCA and 3W-PLS to kinetics. In addition, none of these methodologies has been applied to the resolution of mixtures in enzymatic–kinetic systems. The principal aim of this work is to show the analytical potential of the use of combined chemometric calibration methods for non-selective enzyme systems. This paper is also a practical example of the use of 3W-PCA and 3W-PLS, exploring their ability to deal with a relatively complex kinetic–enzymatic system. Thus, we prepose a simple and accurate methodology for the determination of xanthine and hypoxanthine in human urine using an enzymatic reaction and three-way methodologies. The results from this technique are compared with those obtained using the best-known two-way PCA and PLS with unfolded data.

Enzymatic system

It is known that hypoxanthine (HX) and xanthine (X) are oxidized to uric acid (UA) in a reaction catalysed by the bifunctional enzyme milk xanthine oxidase (XOD, xanthine/oxygen oxidoreductase, EC 1.17.3.2) (see reactions 1, 2 below) (Fig. 1b). This enzyme is one of the most complex of the flavoproteins, and it is known to be an enzyme of low specificity, catalyzing the oxidations of many substances, including purines, pyrimidines, pteridins and aldehydes [6, 24, 25].

$$HX + O_2 \xrightarrow{XOD} X + H_2O_2 \quad (1)$$

$$X + O_2 \xrightarrow{XOD} UA + H_2O_2 \quad (2)$$

The overall process is influenced by factors such as pH, the buffer chosen and its concentration, enzyme concentration, working temperature and the initial presence of uric acid. It is known that the uric acid acts as a linear competitive inhibitor of the reactions involved, even at oxygen saturation [6]. Furthermore, there is evidence that uric acid is oxidized in alkaline solution [26].

Three-way analysis: principal component analysis (3W-PCA) and partial least-squares regression (3W-PLS)

Enzymatic analysis is usually quite far from being linear and, more importantly, some matrix compounds may affect the enzyme activity, making the processes that occur in samples different to those that occur in calibration mixtures prepared from pure analyte solutions.

When dealing with two-way data matrices, PCA is a very useful tool for getting an overall description of the system prior to the development of the calibration model (PLS model). In a three-way system, a 3W-PCA may be performed before applying a 3W-PLS, with the same aim as in the two-way analysis. A sound description of the 3W-PCA and 3W-PLS algorithms used can be found in the references [19, 23, 27, 28].

As has been mentioned before, there are several ways to decompose the array $\underline{\mathbf{X}}$. In this paper, the more straightforward approach has been considered, that of multiway-PCA [18]. This decomposition is equivalent to PCA and has the same goals and benefits.

During 3W-PCA, the three-way matrix $\underline{\mathbf{X}}$ is unfolded in each of the possible ways shown in Fig. 2a, where *i, j* and *k* indicate samples, time and wavelength, respectively. The algorithm proceeds by decomposing the matrix $\underline{\mathbf{X}}$ as the summation of the product of score vectors ($\mathbf{S}$) and loading matrices ($\mathbf{L}$), with the correct number of principal factors (f) chosen by cross-validation. Depending on the particular unfolding performed, it is possible to analyze variability among samples, times or wavelengths. Qualitative information about each mode and the behavior of the system can be obtained by looking at the score matrices ($\mathbf{S}_{if}$, $\mathbf{S}_{jf}$, $\mathbf{S}_{kf}$). The principal advantage of the three-way PCA over two-way PCA is

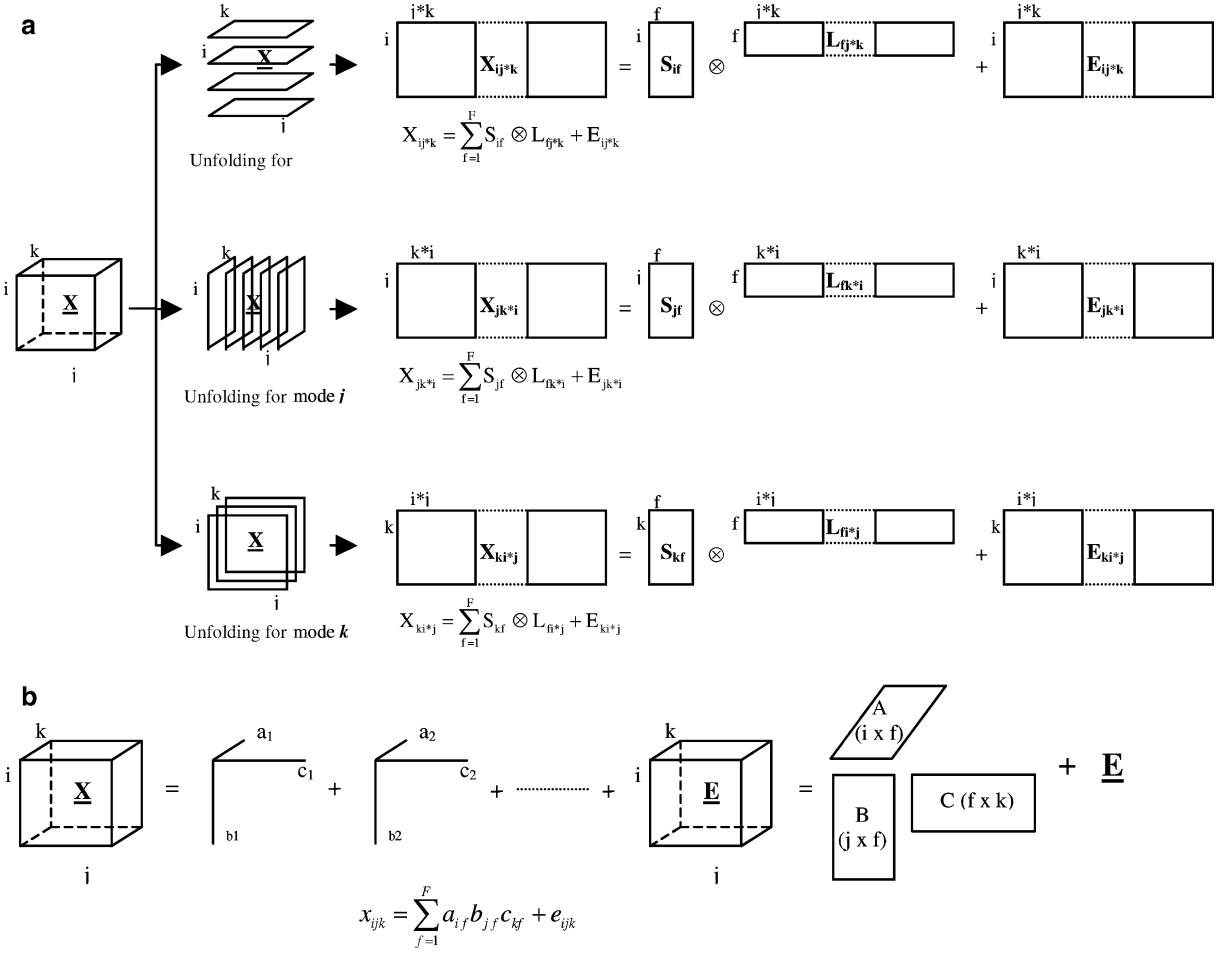


Fig. 2a–b **a** Three-way PCA decomposition. **b** Three-way PLS decomposition

that the score plots obtained from three-way methods are easier to interpret than those obtained using two-way methods.

During *N*-way PLS, and especially during three-way PLS (3W-PLS), the three-way matrix is decomposed into a 3W-model like the PARAFAC model [20] (Fig. 2b). The three-way matrix $\underline{\mathbf{X}}$ is decomposed into a set of rank-one cubes describing $\underline{\mathbf{X}}$ in some optimal sense, so the number of cubes could be higher than the number of chemical components in the system. The main feature of the algorithm is that it produces score vectors that, in a tri-linear sense, have maximum covariance with the unexplained part of the dependent variable. Mathematically speaking, there is no difference between scores and loadings, because both are calculated in the same way, so $\underline{\mathbf{X}}$ is defined with the matrices **A** ($i \times f$), **B** ($j \times f$) and **C** ($k \times f$), whose elements are a_{if}, b_{jf} and c_{kf}, respectively. f is the number of factors. In our case, the dimensions i,j and k are samples, time and wavelength, respectively.

When there are a large number of variables, it is important to note that the solution is easy to interpret compared with to unfolding methods [2]. However, the main drawback of 3W-PLS is the loss of fit when compared with a bilinear model due to the more severe constraints [29].

The 3W-PLS algorithm can be extended into a 3W-PLS2 in the same way as for 2W-PLS. The algorithm for 3W-PLS2 is seen to be a straightforward extension of the 2W-PLS2 algorithm, and is equal to 3W-PLS1 only if just one dependent variable is modeled. In PLS2 algorithms (three- and two-way), several dependent variables are modeled simultaneously. $\underline{\mathbf{X}}$ is decomposed, producing score vectors that have the maximum covariance with the unexplained parts of all of the dependent variables. This fact means that we can take advantage of possible correlations or co-linearity between dependent variables.

Experimental and procedure

Reagents

A 50 mmol L^{-1}, pH 8.5, TRIS buffer solution was prepared from tris(hydroxymethyl)-aminomethane (Merck p.a.) and hydrochloric acid 37% (Panreac p.a.) using Milli-Q water. 5 mmol L^{-1} stock solutions of hypoxanthine, xanthine and uric acid (Merck p.a.) and 0.14 U mL^{-1} of enzyme xanthine oxidase (E.C. 1.17.3.2; from buttermilk; 15% protein; 0.4–1.0 U mg^{-1} protein; Merck) were prepared in the buffer solution. Lyophilized Lyphochek urine (Quantitative Urine Control from BIO-RAD) was regenerated in buffer solution.

Apparatus and software

Kinetic profiles were recorded using an HP-8453 diode array spectrophotometer and a stopped-flow Applied Photophysics RX 2000 system. A spectrophotometric cell of 1.00 cm path length thermostated at 25.0 ± 0.1 °C was used throughout. The temperature was maintained using a Peltier module (Agilent 89090-A). The room temperature was kept at 24 ± 1 °C. Recorded spectra were imported from the spectrophotometer into different software. The 3W-PLS1 and 3W-PLS2 algorithms were implemented in MatLab Version 6.5 (MathWorks, Natick, MA, USA) and PLS-TOOLBOX 3.0 (Eigenvector Research, WA, USA). Unscrambler Version 8.5 (Camo, Norway) was used for 2W-PLS1 and 2W-PLS2 calculations.

Calibration, validation and spiked sample design

Even though the goal is to determine hypoxanthine and xanthine, the inhibitory effect of uric acid obliges us to take into account its concentration when designing the three different groups of samples (calibration, validation and spiked) that were prepared. Therefore, a calibration design incorporating three components and five concentration levels was used.

The 23 calibration mixtures comprised the following ranges of analyte concentrations: hypoxanthine, 0.0–3.0 mmol L^{-1} urine; xanthine, 0.0–1.5 mmol L^{-1} urine; uric acid, 0.0–3.0 mmol L^{-1} urine. All concentrations in the calibration mixtures refer to the original concentrations of the analytes in urine and in the reaction cell.

The 15 mixtures used as the validation set and the 15 spiked samples contained analytes present in the same concentration ranges as defined for the calibration samples. In order to test the repeatability of the procedure, two replicates of each sample were tested.

Working pH was fixed at 8.5; at this pH, the enzyme was found to be most active in both reactions.

Procedure

Lyophilized urine was regenerated with the buffer solution and was filtered through nylon cartridges (0.45 μm). This solution was diluted 1:25 and spiked properly. All solutions were prepared just before they were analyzed.

The stopped–flow mixture system consisted of two parallel syringes. Buffered xanthine oxidase solution was placed in the first syringe, and the buffered calibration, validation or spiked mixture was placed in the second. The reactions started when some of these solutions (15 μl of each) were pushed into the reaction cell. The final concentration of enzyme in the reaction cell was 0.07 units of enzyme per milliliter.

The reference solution used for the calibration and validation sets was the working buffer plus enzyme. For the spiked samples, a reference solution containing the working buffer, enzyme and non-spiked urine was used instead.

0.3 s after mixing, the system began to record the UV-Vis spectra, and these were recorded every 1 s for 160 s, at 1 nm intervals over the wavelength range 225–320 nm. After 160 s, all of the reactions—except for uric acid decomposition— had completed (Fig. 1).

Data recording and processing

The UV–Vis spectrum for each sample I was recorded at J different times(t_1,..., t_j) at K different wavelengths in order to construct 3W-data arrays. To apply the bilinear PLS, the three-way data matrix was unfolded to obtain a classical two-dimensional data matrix in such a way that each row contained the spectrum for a mixture recorded at a particular time, and all of the spectra were sequentially linked: (λ_1 t_1, λ_2 t_1,...,λ_K t_1,...,λ_i t_j, ...,λ_1 t_j, λ_2 t_j,...,λ_K t_j).

Taking into account the intrinsic relationship between hypoxanthine and xanthine, the 3W-PLS2 and 2W-PLS2 algorithms were assayed and compared with 3W-PLS1 and 2W-PLS1 algorithms. Models were internally validated by leave-one-out cross-validation, with each segment being formed by the two replicates of the same mixture.

For each model, the number of significant PLS factors was chosen to be the lower number that yielded a predicted error sum of squares (PRESS) value that was not significantly different from the minimum PRESS (see 3 below).

$$\mathrm{PRESS} = \sum_{i=1}^{n} (\hat{c}_i - c_i)^2 \tag{3}$$

where n is the number of samples, $\hat{c}_i$ is the calculated concentration of sample i and c_i its reference value.

To evaluate the precision and the predictive capabilities of the different models, the standard errors of calibration (SEC) and prediction (SEP) were calculated (see 4, 5 below) taking into account the bias of the model (see 6 below). The standard deviation between replicates (SDBR) (see 7 below) was calculated from the difference between the results for both replicates of each sample. It is, thus, a measure of experimental repeatability and it can be considered the target precision or the best precision attainable under these particular experimental conditions. Values of SEP that are clearly higher than those of SDBR are indicative of a lack of fit for the model.

$$\mathrm{SEC} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{c} - c_i - \mathrm{bias})^2}{n - f - 1}} \tag{4}$$

$$\mathrm{SEP} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{c} - c_i - \mathrm{bias})^2}{n - 1}} \tag{5}$$

$$\mathrm{bias} = \frac{\sum_{i=1}^{n} (\hat{c} - c_i)}{n} \tag{6}$$

$$\mathrm{SDBR} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{c}_{i,1} - \hat{c}_{i,2})^2}{2n}} \tag{7}$$

In these evaluation parameters, c_i is the reference concentration and $\hat{c}_i$ the calculated concentration, n is the number of samples in the prediction or calibration set, and f is the number of PLS factors.

Results and discussion

Comparison between calibration mixtures and spiked samples

The 3W-PCA model with four components (model A) explains more than 99% of the variance in each decomposition mode, even though it must be pointed out that the first PC accounts for around 90% of the variance (Table 1).

A plot of the scores for the different modes is shown in Fig. 3. Evidence that the PC3 score in the spiked samples differs from that in the non-spiked samples is apparent from the sample mode (Fig. 3a). Looking at the score plot in the wavelength mode (Fig. 3c), a difference in the first fifteen wavelengths can be seen for the third PC; however, there is no real difference in the PC3 of the kinetic mode. This fact agrees with the explained variance for PC3 in the sample and wavelength modes (3%), whereas only 0.66% of the variance is explained for the PC3 in the kinetic mode. The score plot of PC1 versus PC3 for the samples (Fig. 4a) clearly shows that PC3

Table 1 Explained variance (%) for 3W-PCA decomposition in each mode

	Sample		Time		Wavelength	
	Model A	Model B	Model A	Model B	Model A	Model B
PC1	90.24	91.20	93.26	93.19	88.71	91.42
PC2	6.12	7.97	5.91	5.98	8.31	8.18
PC3	2.99	0.58	0.66	0.67	2.57	0.39
PC4	0.42	0.10	0.12	0.12	0.38	0.01

Model A refers to the model with all of the wavelengths. *Model B* refers to the model without the first 15 wavelengths

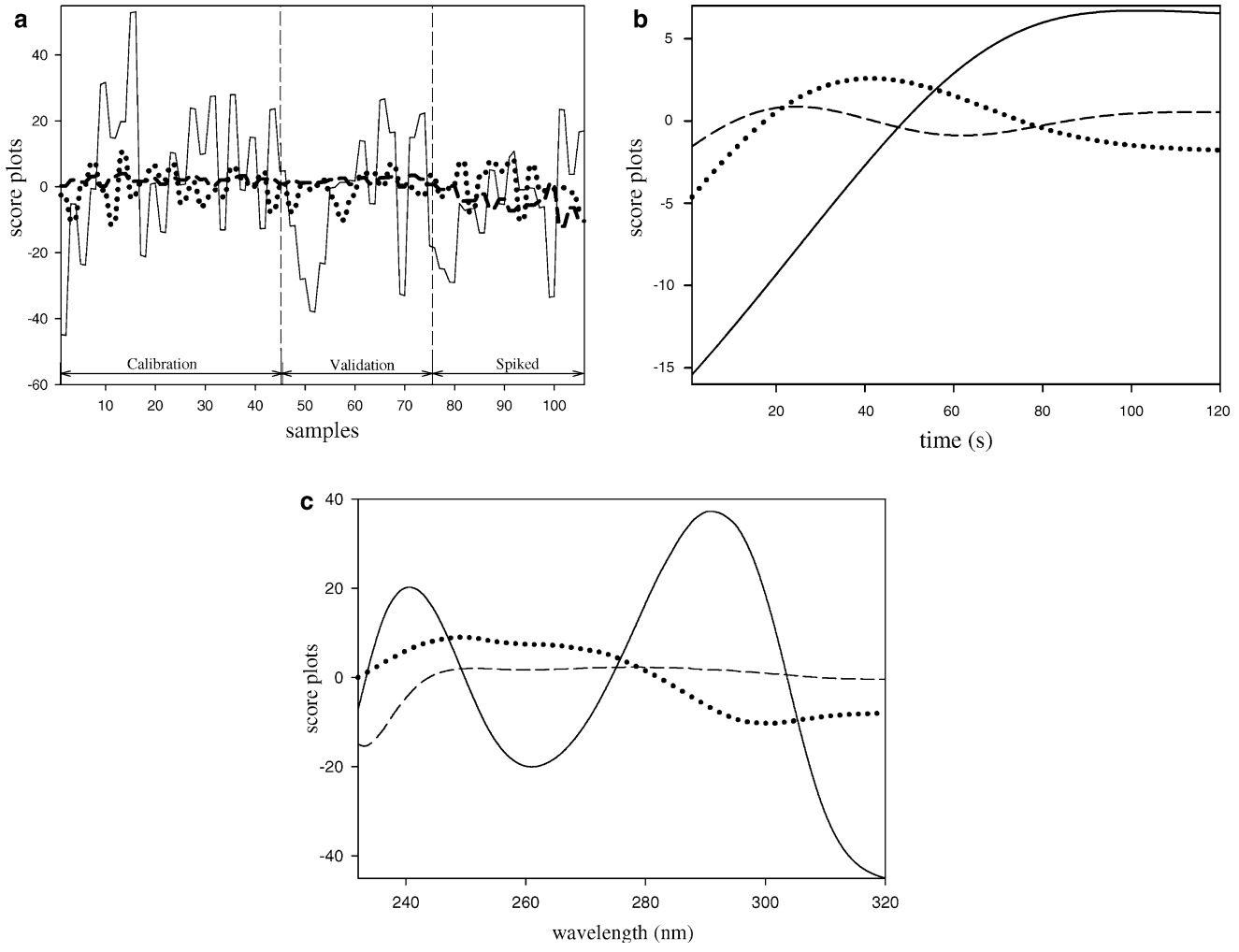


Fig. 3a–c Score plots for the **a** sample, **b** kinetic and **c** wavelength modes (*solid line* PC1, *dotted line* PC2, *dashed line* PC3)

differentiates laboratory mixtures from spiked samples. This difference affects the concentration and spectral mode, but does not affect the kinetics of the system.

After deleting the first 15 wavelengths of the spectra (model B), 3W-PCA shows that all of the samples are merged and become undistinguishable (Fig. 4b). In this case, the third PC for each mode explains no more than 0.67% of the variance (Table 1).

It must be pointed out that PC1 in the three modes of the new model explain the behaviour of uric acid. This is not surprising: according to the kinetic system, the predominant feature is the formation of the broad absorption band of uric acid at 290 nm (Fig. 1a).

Determination of hypoxanthine and xanthine

Overall results for the determination of hypoxanthine and xanthine are shown in Tables 2 and 3, respectively. Validation and spiked samples shows similar figures, with low values of SEP and SBDR, so there were no differences between these groups. This is an indication of a good correction for the matrix effect, and that the models offer good predictive capabilities and repeatabilities. Results for hypoxanthine are a little better than those for xanthine. The xanthine results may indicate a slight non-linearity of the system (SEP is approximately double SDBR), but the predictive capabilities of the models are good enough to quantify both analytes with good results.

As we noted before, the samples in the calibration and validation sets contained different amounts of uric acid and some were without uric acid. There were no differences in the quantification of hypoxanthine and xanthine for both cases, indicating either little influence of the uric acid, or that it has been corrected by the calibration. This means that, despite the fact that the principal source of variance in matrix $\underline{\mathbf{X}}$ is the uric acid, 3W-PLS models decompose $\underline{\mathbf{X}}$ taking into account the covariance with the unexplained part of the dependent variable: the concentration matrix (formed with hypoxanthine and xanthine).

The PLS2 algorithm offers better results than PLS1 due to the intrinsic relationship between hypoxanthine and xanthine. Three-way methods also perform better than two-way methods.

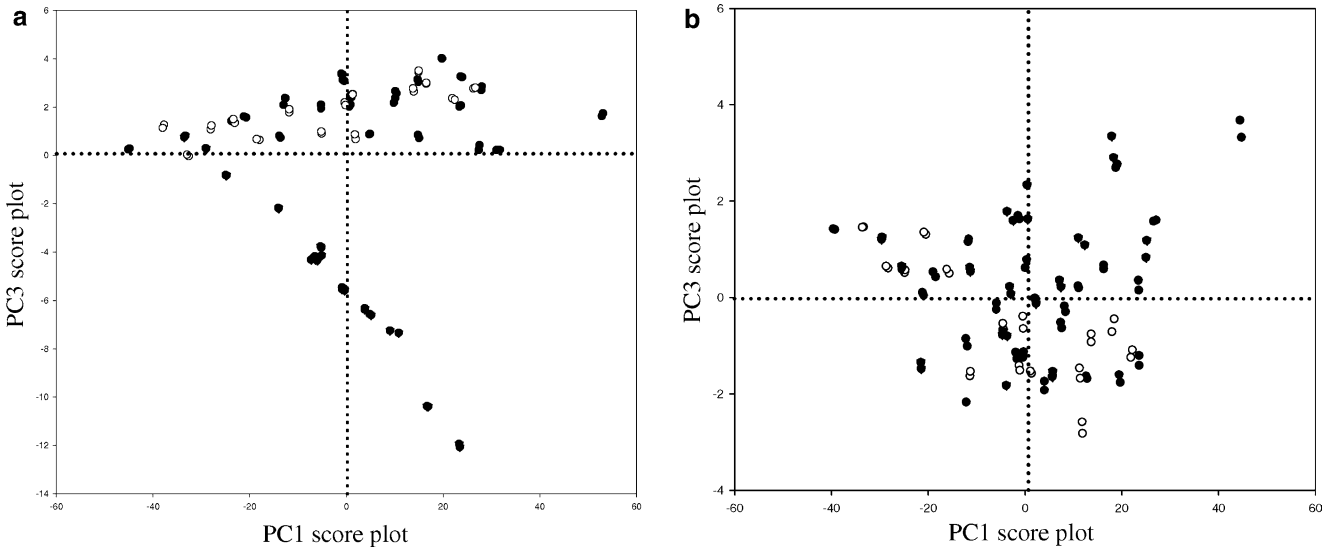


Fig. 4a–b PC1 versus PC3 score plots for sample modes (*open circles* calibration; (*filled circles*) validation; (*filled inverted triangles*) spiked) using **a** model A, with all the wavelengths, and **b** model B, without the first 15 wavelengths

Table 2 Model results for hypoxanthine in calibration, validation and spiked samples

	PLS1	PLS2	3W-PLS1	3W-PLS2
Calibration				
Num. PC	4	7	4	7
SEC[a]	2.07	1.28	1.94	0.96
SDBR[a]	0.39	0.63	0.49	0.62
Validation				
SEP[a]	1.60	0.68	2.04	0.61
SDBR[a]	0.34	0.47	0.42	0.44
Spiked				
SEP[a]	2.26	1.41	2.20	1.03
SDBR[a]	0.98	1.06	0.97	0.96

[a] μmol L^{-1} urine

Table 3 Model results for xanthine in calibration, validation and spiked samples

	PLS1	PLS2	3W-PLS1	3W-PLS2
Calibration				
Num. PC	7	7	7	7
SEC[a]	1.54	1.6	1.08	1.09
SDBR[a]	0.50	0.48	0.61	0.54
Validation				
SEP[a]	0.95	1.30	0.94	0.96
SDBR[a]	0.59	0.60	0.65	0.54
Spiked				
SEP[a]	1.49	1.54	1.31	1.26
SDBR[a]	0.73	0.60	0.71	0.56

[a] μmol L^{-1} urine

The loading plots shown in Fig. 5 give us useful information. The shape of the sample profile of PC1 is related to the hypoxanthine concentration in the calibration set. The shape of the kinetic loading for PC1 suggests that all times are equally important in the determination of hypoxanthine and xanthine; whereas the loading of PC2 indicates that the times between 20 and 90 s are less important (a value of zero is reached); a similar explanation applies to the kinetic loadings for PC3 and PC4, indicating that data at the beginning of the reaction have more importance to the quantification of the analytes. The spectral loadings, for all the PCs, show two regions of importance, one with maxima at 250–270 nm, mainly related to hypoxanthine, and the other with maxima around 290 nm, related to uric acid. One maximum in the xanthine spectrum appears at 280 nm (Fig. 1a), but there is no important spectral loading at this wavelength.

Table 4 presents the figures of merit for the regression lines of the reference values versus the calculated ones obtained using the 3W-PLS2 model. The standard deviation of the slope and intercept are given in brackets. For the three sets of samples (calibration, validation and spiked), the slopes are very close to one, as are the determination coefficients (R^2).

Results obtained by other authors for the quantification of hypoxanthine and xanthine in human urine and spiked buffered solutions can be found in the literature [8, 10, 12]. Results obtained using the the kinetic–spectrophotometric method and 3W-PLS2 are comparable with the results obtained using various analytical methods (including HPLC and gas chromatography).

Conclusions

The potential application of three-way methodologies to enzymatic processes has been tested with good results. 3W-PCA and 3W-PLS have been shown to be an excellent alternative to unfolding methods. Qualitative information on the kinetic behavior of the system and differences between laboratory and spiked samples has been obtained using 3W-PCA.

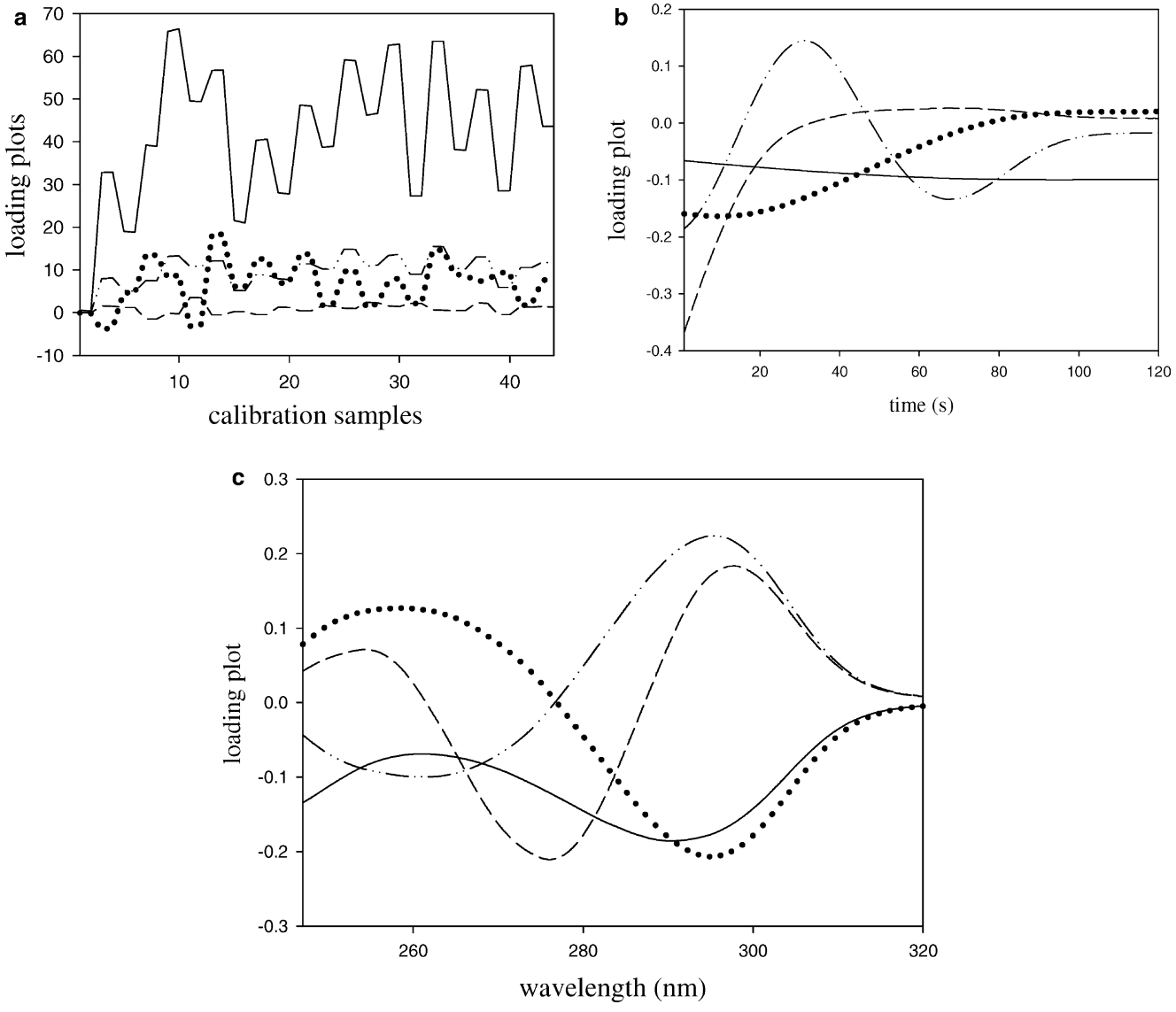


Fig. 5a–c 3W-PLS2 loadings for the quantification of hypoxanthine and xanthine (*solid line* PC1, *dotted line* PC2, *dashed line* PC3, *dash-dotted line* PC4). **a** Sample loading, **b** kinetic loading, **c** spectral loading

Table 4 Figures of merit for the regression lines of the reference values versus those calculated using the model 3W-PLS2

	Calibration	Validation	Spiked
Hypoxanthine			
Slope	0.987 (0.006)	1.000 (0.006)	0.978 (0.011)
Intercept	0.649 (0.217)	−0.438 (0.211)	0.072 (0.340)
R^2	0.999	0.999	0.998
Xanthine			
Slope	1.005 (0.019)	1.001 (0.023)	0.987 (0.030)
Intercept	–0.248 (0.358)	0.045 (0.385)	0.407 (0.509)
R^2	0.990	0.989	0.982

Values of standard deviation are in *parentheses*. R^2 is the determination coefficient

In addition, a new procedure for the simultaneous determination of mixtures of hypoxanthine and xanthine in human urine has been tested and presented, yielding good results. This analysis uses several chemometric methods and does not require any previous chromatographic separation.

The 3W-PLS2 algorithm yielded the best results for xanthine and hypoxanthine quantification. The presence of uric acid did not influence the results. The SEP and SBDR are low, indicating that the models offer good predictive capabilities and repeatabilities. The results for xanthine may indicate that the system is slightly non-linear (SEP greater than SDBR), but the predictive capabilities of the models are good enough to quantify both analytes with good results.

Acknowledgements The authors gratefully acknowledge the financial support of this work by the Ministerio de Ciencia y Tecnología (DGI BQU2001–2019) and the Comissionat per a Universitats i Recerca of the Generalitat de Catalunya (2001-SGR-00176).

References

1. Pettersson Å, Karlberg B (1997) Anal Chim Acta 354:241–248
2. Crouch SR, Coello J, Maspoch S, Porcel M (2000) Anal Chim Acta 424:115–126
3. Coello J, Maspoch S, Villegas N (2000) Talanta 53:627–637
4. Ni Y, Huang C, Kokot S (2004) Chemom Intell Lab Syst 71:177–193
5. Blanco M, Coello J, Iturriaga H, Maspoch S, Porcel M (1999) Anal Chim Acta 398:83–92
6. Escribano J, Garcia-Canovas F, Garcia-carmona F (1988) J Biochem 254:829–833
7. Tawa R, Kito M, Hirose S (1981) Chem Lett 745–748
8. Putterman GJ, Shaikh B, Hallmark MR, Sawyer CG, Hixson CV, Perini F (1979) Anal Biochem 98:18–26
9. Czauderna M, Kowalczyk J (1997) J Chromatogr B 704:89–98
10. Di Pietro MC, Vannoni D, Leoncini R, Liso G, Guerranti R, Marinello E (2001) J Chromatogr B 751:87–92
11. Chen G, Chu Q, Zhang L, Ye J (2002) Anal Chim Acta 457:225–233
12. Zen J-M, Lai Y-Y, Yang H-H, Senthil Kumar A (2002) Sens Actuator B 84:237–244
13. Pei J, Li XY (2000) Anal Chim Acta 414:205–213
14. Foppoli C, Coccia R, Cini C, Rosei MA (1997) Biochim Biophys Acta 1334:200–206
15. Carsol M-A, Mascini M (1998) Talanta 47:335–342
16. Geladi P (1989) Chemom Intell Lab Syst 7:11–30
17. Sanchez E, Kowalski BR (1990) J Chemom 4:29–45
18. Wold S, Geladi P, Esbensen K, Öhman J (1987) J Chemom 1:41–56
19. Henrion R (1994) Chemom Intell Lab Syst 25:1–23
20. Bro R (1997) Appl Spectrosc Rev 32–3:237–261
21. Tucker L (1963) Problems of measuring change. University of Wisconsin Press, Madison, WI, pp 122–137
22. Tauler R (1995) Chemom Intell Lab Syst 30:133–146
23. Bro R (1996) J Chemom 10:47–61
24. Massey V, Brumby PE, Komai H, Palmer G (1969) J Biol Chem 244–247:1682–1691
25. Jeëwska MM (1973) Eur J Biochem 36:385–390
26. Lindsey AS, Sharma RK (1981) Anal Lett 14-B10:799–811
27. Smilde AK (1997) J Chemom 11:367–377
28. Jong S (1998) J Chemom 12:77–81
29. Kiers HAL (1991) Psychometrica 56:449–470

Anal Bioanal Chem (2005) 382: 1380–1388
DOI 10.1007/s00216-005-3275-4

ORIGINAL PAPER

José Manuel Amigo · Jordi Coello · Santiago Maspoch

Three-way partial least-squares regression for the simultaneous kinetic-enzymatic determination of xanthine and hypoxanthine in human urine

Received: 22 February 2005 / Accepted: 22 April 2005 / Published online: 7 June 2005
© Springer-Verlag 2005

Abstract The performance of three-way principal component analysis and three-way partial least-squares regression when applied to a complex kinetic-enzymatic system is studied, in order to investigate the analytical potential of the combined use of these chemometric technologies for non-selective enzymatic systems. A enzymatic-kinetic procedure for the simultaneous determination of hypoxanthine and xanthine in spiked samples of human urine is proposed. The chemical system involves two consecutive reactions catalyzed by xanthine oxidase (EC 1.17.3.2). This enzyme catalyzes the oxidation of hypoxanthine, first to xanthine and then to uric acid, a competitive inhibitor of the reactions. The influence of uric acid during quantitative determination was considered in the design of the calibration set. The sample and enzyme solution were mixed in a stopped-flow module and the reaction was monitored using a diode array spectrophotometer. The recorded data have an intrinsical three-component structure (samples, time and wavelength). This data array was studied via three-way principal component analysis and was modeled for quantitative purposes using a three-way partial least-squares calibration procedure. Results are compared with those obtained by applying classical bilinear PLS to the previously unfolded data matrix.

Keywords Oxipurines · Urine analysis · Three-way principal component analysis · Three-way partial least-squares regression · Simultaneous kinetic determination

Introduction

While enzymes are usually visualized as highly selective catalysts and are used with the aim of obtaining a specific product, there are a number of them that, in fact, show group selectivity. The development of multivariate calibration procedures and their application to kinetics [1–4] has renewed analytical interest in non-selective enzymes, since they can potentially be applied to the simultaneous and rapid determination of several analytes using relatively simple instrumentation. The use of enzymes showing group selectivity becomes an analytical choice when applied to the kinetic resolution of similar compounds with the same functional groups. The potential for this kind of resolution has already been proved during the simultaneous determination of ethanol and methanol using an alcohol oxidase [5]. Since the same product is formed from both compounds, discrimination is only possible via kinetic information, and this non-linear system was solved by applying an artificial neural network to the data recorded at a single wavelength.

The system studied in this work is more complex in the sense that it consists of two consecutive reactions ($A \rightarrow B \rightarrow C$) catalyzed by the same enzyme, where A and B are the analytes and C, the product, acts as an inhibitor [6]. However, although the kinetic system is complex, there is a large amount of information available since all three compounds absorb in the ultraviolet region, although with highly overlapped absorption bands. Therefore, the spectral and kinetic data from the system could contribute useful information that may be used to discriminate the responses of the analytes, and therefore to resolve the mixtures.

The oxipurine (hypoxanthine and xanthine) and uric acid levels in blood, plasma and urine may provide sensitive indicators of certain pathologic states, including xanthinuria, gout, renal failure, toxaemia during pregnancy, and other diseases [7]. These compounds have similar chemical structures and UV–Vis spectra (Fig. 1a), which has prompted the use of analytical separation techniques such as high performance liquid chromatography (HPLC) [8–10] or high performance capillary electrophoresis (HPCE) [11] for their simultaneous analysis. In addition, several electrochemical methods have been proposed [12, 13]. Until now, the

J. M. Amigo · J. Coello (✉) · S. Maspoch
Universitat Autónoma de Barcelona,
Bellaterra, 08193 Barcelona, Spain
E-mail: jordi.coello@uab.es
Tel.: +34-93-5812122
Fax: +34-93-5812379

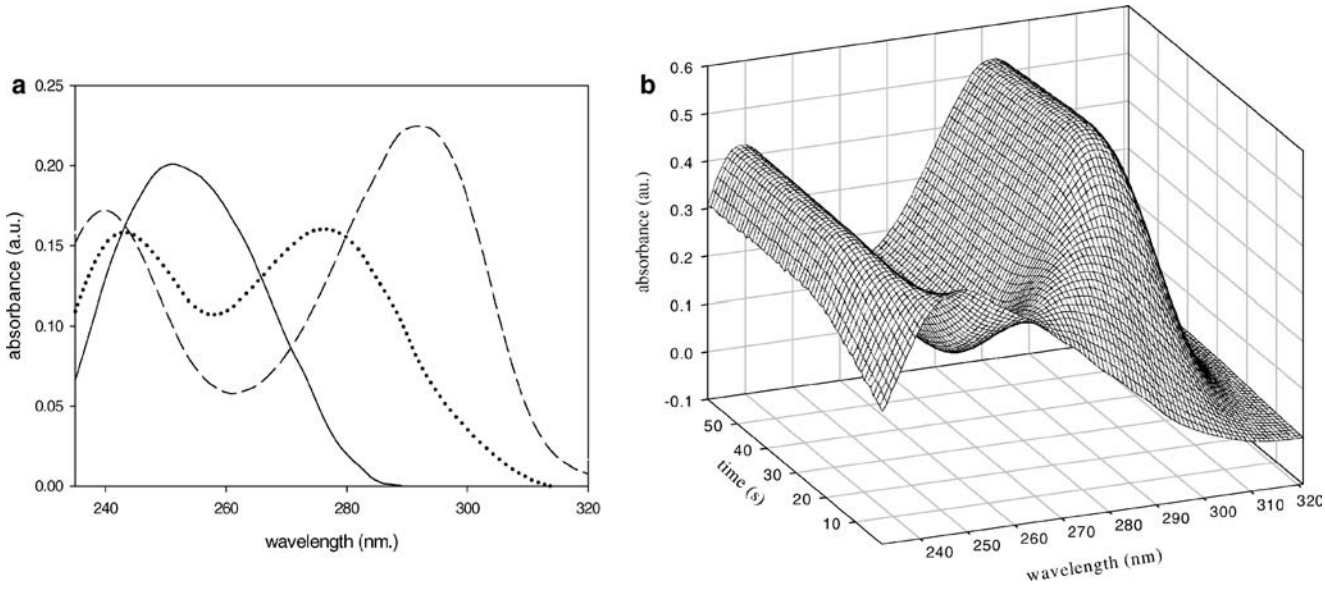


Fig. 1a–b **a** Pure spectra of xanthine (*dotted line* 30 μmol L^{-1}), hypoxanthine (*solid line* 60 μmol L^{-1}) and uric acid (*dashed line* 60 μmol L^{-1}). The important wavelengths are 238 and 293 nm (formation of uric acid), 249 nm (hypoxanthine transformation), 277 nm (xanthine transformation). **b** Kinetic profile of a mixture of hypoxanthine (30 μmol L^{-1}) and xanthine (15 μmol L^{-1}) reacting with 0.07 U mL^{-1} of xanthine oxidase

simultaneous determination of xanthine and hypoxanthine without any previous separation has not been performed due to the lack of spectral selectivity. This lack of spectral selectivity is not a problem for kinetic-chemometric methods, where information about the spectral evolution with time can be collected and several multivariate methods can be used.

The enzyme xanthine oxidase (EC 1.17.3.2) has been used as a group enzyme to develop methodologies for the determination of complex mixtures of analytes [14], as well as for hypoxanthine and xanthine [7, 13, 15]. In all of these cases, the enzyme was immobilized in a specific electrode or in a chromatographic pre-column.

Kinetic-spectrophotometric data have an intrinsic three-way structure (sample, time and wavelength). Application of bilinear partial least-squares (PLS) to multivariate kinetic data, prior to unfolding the data, procures a high predictive capability and versatility, being able to solve a large number of different kinetic models. However, the large number of variables (vectors with over 10,000 data points are easily obtained) containing both kinetic and spectral information, make difficult to extract the qualitative information needed to understand what is really happening during the reaction.

Several chemometric methodologies have been applied in order to decompose three-way data matrices [16]. Some of these methodologies are: tri-linear decomposition (TLD) [17], three-way principal components analysis (3W-PCA) [18, 19], parallel factor analysis (PARAFAC) [20], Tucker models [21], and three-way-multivariate curve resolution–alternating least-squares (3W-MCR-ALS) [22]. All of these methodologies are able to split the tri-linear systems into separate spectral and kinetic profiles, although each algorithm produces results that are affected by different restrictions and, in some cases, the requirement for linearity restricts their application. Three-way PLS (3W-PLS) was developed in order to obtain quantitative information on the systems without the need to unfold the data matrix, using an algorithm similar to the bilinear PLS used for two-dimensional data [23].

Despite the increasing use of these three-way methodologies, there is still very little real practical experience of applying 3W-PCA and 3W-PLS to kinetics. In addition, none of these methodologies has been applied to the resolution of mixtures in enzymatic–kinetic systems. The principal aim of this work is to show the analytical potential of the use of combined chemometric calibration methods for non-selective enzyme systems. This paper is also a practical example of the use of 3W-PCA and 3W-PLS, exploring their ability to deal with a relatively complex kinetic–enzymatic system. Thus, we prepose a simple and accurate methodology for the determination of xanthine and hypoxanthine in human urine using an enzymatic reaction and three-way methodologies. The results from this technique are compared with those obtained using the best-known two-way PCA and PLS with unfolded data.

Enzymatic system

It is known that hypoxanthine (HX) and xanthine (X) are oxidized to uric acid (UA) in a reaction catalysed by the bifunctional enzyme milk xanthine oxidase (XOD, xanthine/oxygen oxidoreductase, EC 1.17.3.2) (see reactions 1, 2 below) (Fig. 1b). This enzyme is one of the most complex of the flavoproteins, and it is known to be an enzyme of low specificity, catalyzing the oxidations of many substances, including purines, pyrimidines, pteridins and aldehydes [6, 24, 25].

$$HX + O_2 \xrightarrow{XOD} X + H_2O_2 \quad (1)$$

$$X + O_2 \xrightarrow{XOD} UA + H_2O_2 \quad (2)$$

The overall process is influenced by factors such as pH, the buffer chosen and its concentration, enzyme concentration, working temperature and the initial presence of uric acid. It is known that the uric acid acts as a linear competitive inhibitor of the reactions involved, even at oxygen saturation [6]. Furthermore, there is evidence that uric acid is oxidized in alkaline solution [26].

Three-way analysis: principal component analysis (3W-PCA) and partial least-squares regression (3W-PLS)

Enzymatic analysis is usually quite far from being linear and, more importantly, some matrix compounds may affect the enzyme activity, making the processes that occur in samples different to those that occur in calibration mixtures prepared from pure analyte solutions.

When dealing with two-way data matrices, PCA is a very useful tool for getting an overall description of the system prior to the development of the calibration model (PLS model). In a three-way system, a 3W-PCA may be performed before applying a 3W-PLS, with the same aim as in the two-way analysis. A sound description of the 3W-PCA and 3W-PLS algorithms used can be found in the references [19, 23, 27, 28].

As has been mentioned before, there are several ways to decompose the array $\underline{\mathbf{X}}$. In this paper, the more straightforward approach has been considered, that of multiway-PCA [18]. This decomposition is equivalent to PCA and has the same goals and benefits.

During 3W-PCA, the three-way matrix $\underline{\mathbf{X}}$ is unfolded in each of the possible ways shown in Fig. 2a, where *i, j* and *k* indicate samples, time and wavelength, respectively. The algorithm proceeds by decomposing the matrix $\underline{\mathbf{X}}$ as the summation of the product of score vectors ($\mathbf{S}$) and loading matrices ($\mathbf{L}$), with the correct number of principal factors (f) chosen by cross-validation. Depending on the particular unfolding performed, it is possible to analyze variability among samples, times or wavelengths. Qualitative information about each mode and the behavior of the system can be obtained by looking at the score matrices ($\mathbf{S}_{if}$, $\mathbf{S}_{jf}$, $\mathbf{S}_{kf}$). The principal advantage of the three-way PCA over two-way PCA is

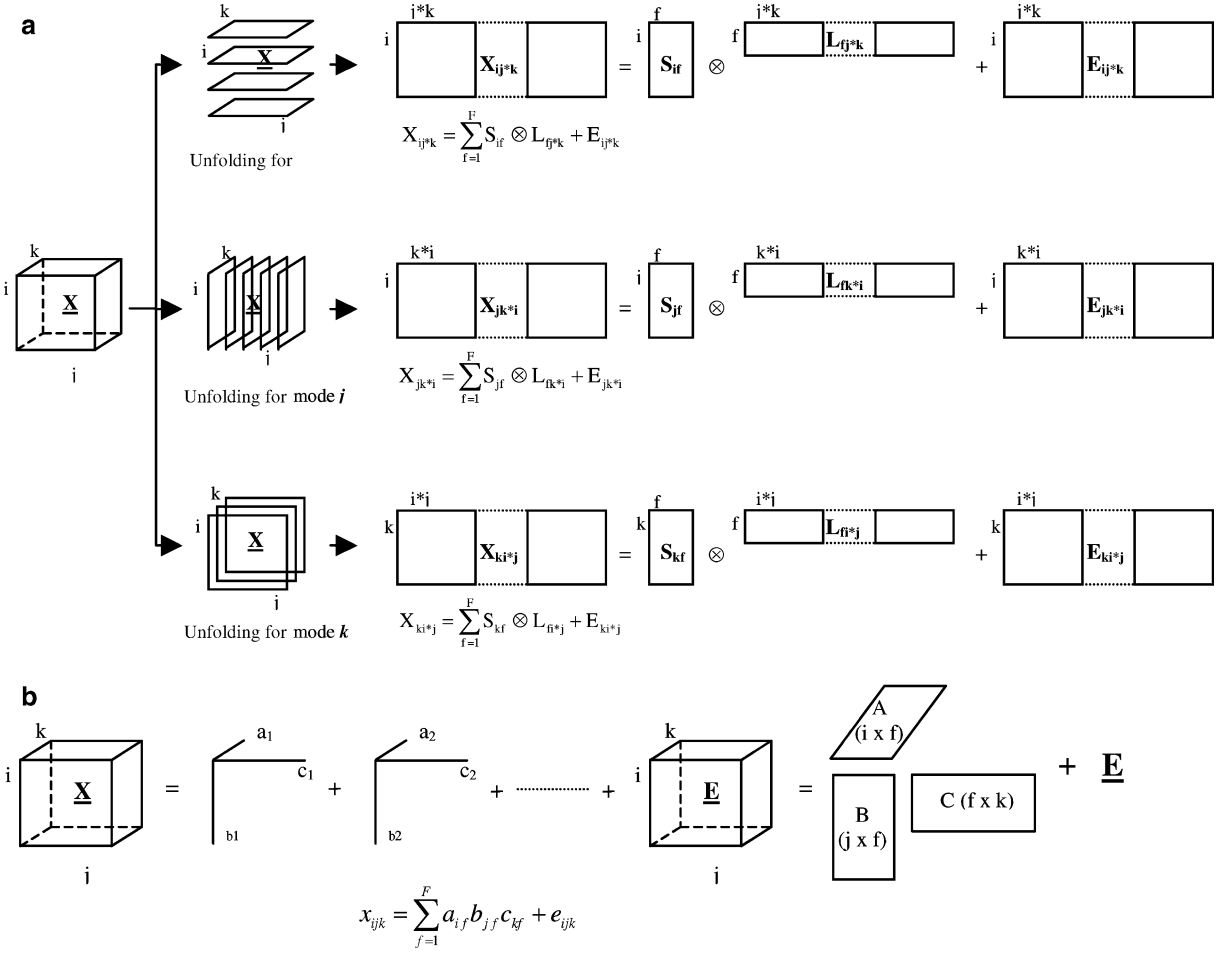


Fig. 2a–b **a** Three-way PCA decomposition. **b** Three-way PLS decomposition

that the score plots obtained from three-way methods are easier to interpret than those obtained using two-way methods.

During *N*-way PLS, and especially during three-way PLS (3W-PLS), the three-way matrix is decomposed into a 3W-model like the PARAFAC model [20] (Fig. 2b). The three-way matrix $\underline{\mathbf{X}}$ is decomposed into a set of rank-one cubes describing $\underline{\mathbf{X}}$ in some optimal sense, so the number of cubes could be higher than the number of chemical components in the system. The main feature of the algorithm is that it produces score vectors that, in a tri-linear sense, have maximum covariance with the unexplained part of the dependent variable. Mathematically speaking, there is no difference between scores and loadings, because both are calculated in the same way, so $\underline{\mathbf{X}}$ is defined with the matrices **A** ($i \times f$), **B** ($j \times f$) and **C** ($k \times f$), whose elements are a_{if}, b_{jf} and c_{kf}, respectively. f is the number of factors. In our case, the dimensions i,j and k are samples, time and wavelength, respectively.

When there are a large number of variables, it is important to note that the solution is easy to interpret compared with to unfolding methods [2]. However, the main drawback of 3W-PLS is the loss of fit when compared with a bilinear model due to the more severe constraints [29].

The 3W-PLS algorithm can be extended into a 3W-PLS2 in the same way as for 2W-PLS. The algorithm for 3W-PLS2 is seen to be a straightforward extension of the 2W-PLS2 algorithm, and is equal to 3W-PLS1 only if just one dependent variable is modeled. In PLS2 algorithms (three- and two-way), several dependent variables are modeled simultaneously. $\underline{\mathbf{X}}$ is decomposed, producing score vectors that have the maximum covariance with the unexplained parts of all of the dependent variables. This fact means that we can take advantage of possible correlations or co-linearity between dependent variables.

Experimental and procedure

Reagents

A 50 mmol L^{-1}, pH 8.5, TRIS buffer solution was prepared from tris(hydroxymethyl)-aminomethane (Merck p.a.) and hydrochloric acid 37% (Panreac p.a.) using Milli-Q water. 5 mmol L^{-1} stock solutions of hypoxanthine, xanthine and uric acid (Merck p.a.) and 0.14 U mL^{-1} of enzyme xanthine oxidase (E.C. 1.17.3.2; from buttermilk; 15% protein; 0.4–1.0 U mg^{-1} protein; Merck) were prepared in the buffer solution. Lyophilized Lyphochek urine (Quantitative Urine Control from BIO-RAD) was regenerated in buffer solution.

Apparatus and software

Kinetic profiles were recorded using an HP-8453 diode array spectrophotometer and a stopped-flow Applied Photophysics RX 2000 system. A spectrophotometric cell of 1.00 cm path length thermostated at 25.0 ± 0.1 °C was used throughout. The temperature was maintained using a Peltier module (Agilent 89090-A). The room temperature was kept at 24 ± 1 °C. Recorded spectra were imported from the spectrophotometer into different software. The 3W-PLS1 and 3W-PLS2 algorithms were implemented in MatLab Version 6.5 (MathWorks, Natick, MA, USA) and PLS-TOOLBOX 3.0 (Eigenvector Research, WA, USA). Unscrambler Version 8.5 (Camo, Norway) was used for 2W-PLS1 and 2W-PLS2 calculations.

Calibration, validation and spiked sample design

Even though the goal is to determine hypoxanthine and xanthine, the inhibitory effect of uric acid obliges us to take into account its concentration when designing the three different groups of samples (calibration, validation and spiked) that were prepared. Therefore, a calibration design incorporating three components and five concentration levels was used.

The 23 calibration mixtures comprised the following ranges of analyte concentrations: hypoxanthine, 0.0–3.0 mmol L^{-1} urine; xanthine, 0.0–1.5 mmol L^{-1} urine; uric acid, 0.0–3.0 mmol L^{-1} urine. All concentrations in the calibration mixtures refer to the original concentrations of the analytes in urine and in the reaction cell.

The 15 mixtures used as the validation set and the 15 spiked samples contained analytes present in the same concentration ranges as defined for the calibration samples. In order to test the repeatability of the procedure, two replicates of each sample were tested.

Working pH was fixed at 8.5; at this pH, the enzyme was found to be most active in both reactions.

Procedure

Lyophilized urine was regenerated with the buffer solution and was filtered through nylon cartridges (0.45 μm). This solution was diluted 1:25 and spiked properly. All solutions were prepared just before they were analyzed.

The stopped–flow mixture system consisted of two parallel syringes. Buffered xanthine oxidase solution was placed in the first syringe, and the buffered calibration, validation or spiked mixture was placed in the second. The reactions started when some of these solutions (15 μl of each) were pushed into the reaction cell. The final concentration of enzyme in the reaction cell was 0.07 units of enzyme per milliliter.

The reference solution used for the calibration and validation sets was the working buffer plus enzyme. For the spiked samples, a reference solution containing the working buffer, enzyme and non-spiked urine was used instead.

0.3 s after mixing, the system began to record the UV-Vis spectra, and these were recorded every 1 s for 160 s, at 1 nm intervals over the wavelength range 225–320 nm. After 160 s, all of the reactions—except for uric acid decomposition— had completed (Fig. 1).

Data recording and processing

The UV–Vis spectrum for each sample I was recorded at J different times(t_1,..., t_j) at K different wavelengths in order to construct 3W-data arrays. To apply the bilinear PLS, the three-way data matrix was unfolded to obtain a classical two-dimensional data matrix in such a way that each row contained the spectrum for a mixture recorded at a particular time, and all of the spectra were sequentially linked: (λ_1 t_1, λ_2 t_1,...,λ_K t_1,...,λ_i t_j, ...,λ_1 t_j, λ_2 t_j,...,λ_K t_j).

Taking into account the intrinsic relationship between hypoxanthine and xanthine, the 3W-PLS2 and 2W-PLS2 algorithms were assayed and compared with 3W-PLS1 and 2W-PLS1 algorithms. Models were internally validated by leave-one-out cross-validation, with each segment being formed by the two replicates of the same mixture.

For each model, the number of significant PLS factors was chosen to be the lower number that yielded a predicted error sum of squares (PRESS) value that was not significantly different from the minimum PRESS (see 3 below).

$$\mathrm{PRESS} = \sum_{i=1}^{n} (\hat{c}_i - c_i)^2 \tag{3}$$

where n is the number of samples, $\hat{c}_i$ is the calculated concentration of sample i and c_i its reference value.

To evaluate the precision and the predictive capabilities of the different models, the standard errors of calibration (SEC) and prediction (SEP) were calculated (see 4, 5 below) taking into account the bias of the model (see 6 below). The standard deviation between replicates (SDBR) (see 7 below) was calculated from the difference between the results for both replicates of each sample. It is, thus, a measure of experimental repeatability and it can be considered the target precision or the best precision attainable under these particular experimental conditions. Values of SEP that are clearly higher than those of SDBR are indicative of a lack of fit for the model.

$$\mathrm{SEC} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{c} - c_i - \mathrm{bias})^2}{n - f - 1}} \tag{4}$$

$$\mathrm{SEP} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{c} - c_i - \mathrm{bias})^2}{n - 1}} \tag{5}$$

$$\mathrm{bias} = \frac{\sum_{i=1}^{n} (\hat{c} - c_i)}{n} \tag{6}$$

$$\mathrm{SDBR} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{c}_{i,1} - \hat{c}_{i,2})^2}{2n}} \tag{7}$$

In these evaluation parameters, c_i is the reference concentration and $\hat{c}_i$ the calculated concentration, n is the number of samples in the prediction or calibration set, and f is the number of PLS factors.

Results and discussion

Comparison between calibration mixtures and spiked samples

The 3W-PCA model with four components (model A) explains more than 99% of the variance in each decomposition mode, even though it must be pointed out that the first PC accounts for around 90% of the variance (Table 1).

A plot of the scores for the different modes is shown in Fig. 3. Evidence that the PC3 score in the spiked samples differs from that in the non-spiked samples is apparent from the sample mode (Fig. 3a). Looking at the score plot in the wavelength mode (Fig. 3c), a difference in the first fifteen wavelengths can be seen for the third PC; however, there is no real difference in the PC3 of the kinetic mode. This fact agrees with the explained variance for PC3 in the sample and wavelength modes (3%), whereas only 0.66% of the variance is explained for the PC3 in the kinetic mode. The score plot of PC1 versus PC3 for the samples (Fig. 4a) clearly shows that PC3

Table 1 Explained variance (%) for 3W-PCA decomposition in each mode

	Sample		Time		Wavelength	
	Model A	Model B	Model A	Model B	Model A	Model B
PC1	90.24	91.20	93.26	93.19	88.71	91.42
PC2	6.12	7.97	5.91	5.98	8.31	8.18
PC3	2.99	0.58	0.66	0.67	2.57	0.39
PC4	0.42	0.10	0.12	0.12	0.38	0.01

Model A refers to the model with all of the wavelengths. *Model B* refers to the model without the first 15 wavelengths

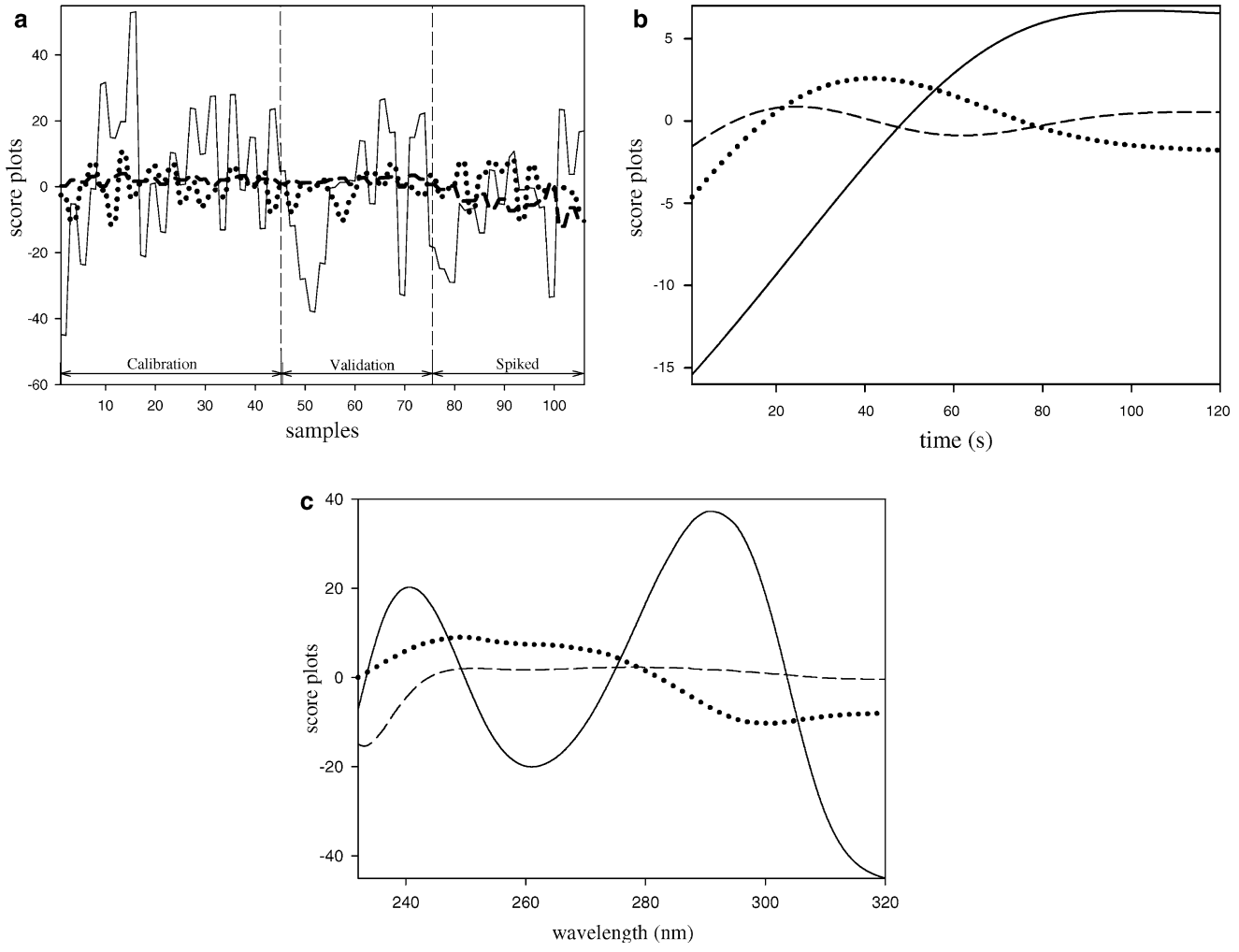


Fig. 3a–c Score plots for the **a** sample, **b** kinetic and **c** wavelength modes (*solid line* PC1, *dotted line* PC2, *dashed line* PC3)

differentiates laboratory mixtures from spiked samples. This difference affects the concentration and spectral mode, but does not affect the kinetics of the system.

After deleting the first 15 wavelengths of the spectra (model B), 3W-PCA shows that all of the samples are merged and become undistinguishable (Fig. 4b). In this case, the third PC for each mode explains no more than 0.67% of the variance (Table 1).

It must be pointed out that PC1 in the three modes of the new model explain the behaviour of uric acid. This is not surprising: according to the kinetic system, the predominant feature is the formation of the broad absorption band of uric acid at 290 nm (Fig. 1a).

Determination of hypoxanthine and xanthine

Overall results for the determination of hypoxanthine and xanthine are shown in Tables 2 and 3, respectively. Validation and spiked samples shows similar figures, with low values of SEP and SBDR, so there were no differences between these groups. This is an indication of a good correction for the matrix effect, and that the models offer good predictive capabilities and repeatabilities. Results for hypoxanthine are a little better than those for xanthine. The xanthine results may indicate a slight non-linearity of the system (SEP is approximately double SDBR), but the predictive capabilities of the models are good enough to quantify both analytes with good results.

As we noted before, the samples in the calibration and validation sets contained different amounts of uric acid and some were without uric acid. There were no differences in the quantification of hypoxanthine and xanthine for both cases, indicating either little influence of the uric acid, or that it has been corrected by the calibration. This means that, despite the fact that the principal source of variance in matrix $\underline{\mathbf{X}}$ is the uric acid, 3W-PLS models decompose $\underline{\mathbf{X}}$ taking into account the covariance with the unexplained part of the dependent variable: the concentration matrix (formed with hypoxanthine and xanthine).

The PLS2 algorithm offers better results than PLS1 due to the intrinsic relationship between hypoxanthine and xanthine. Three-way methods also perform better than two-way methods.

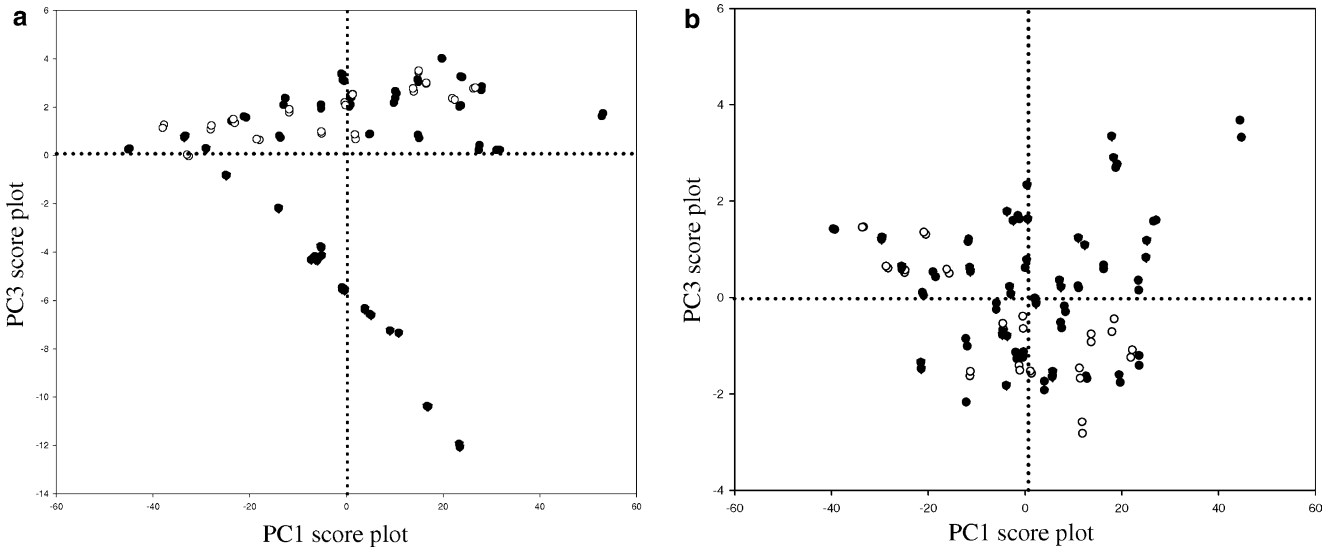


Fig. 4a–b PC1 versus PC3 score plots for sample modes (*open circles* calibration; (*filled circles*) validation; (*filled inverted triangles*) spiked) using **a** model A, with all the wavelengths, and **b** model B, without the first 15 wavelengths

Table 2 Model results for hypoxanthine in calibration, validation and spiked samples

	PLS1	PLS2	3W-PLS1	3W-PLS2
Calibration				
Num. PC	4	7	4	7
SEC[a]	2.07	1.28	1.94	0.96
SDBR[a]	0.39	0.63	0.49	0.62
Validation				
SEP[a]	1.60	0.68	2.04	0.61
SDBR[a]	0.34	0.47	0.42	0.44
Spiked				
SEP[a]	2.26	1.41	2.20	1.03
SDBR[a]	0.98	1.06	0.97	0.96

[a] μmol L^{-1} urine

Table 3 Model results for xanthine in calibration, validation and spiked samples

	PLS1	PLS2	3W-PLS1	3W-PLS2
Calibration				
Num. PC	7	7	7	7
SEC[a]	1.54	1.6	1.08	1.09
SDBR[a]	0.50	0.48	0.61	0.54
Validation				
SEP[a]	0.95	1.30	0.94	0.96
SDBR[a]	0.59	0.60	0.65	0.54
Spiked				
SEP[a]	1.49	1.54	1.31	1.26
SDBR[a]	0.73	0.60	0.71	0.56

[a] μmol L^{-1} urine

The loading plots shown in Fig. 5 give us useful information. The shape of the sample profile of PC1 is related to the hypoxanthine concentration in the calibration set. The shape of the kinetic loading for PC1 suggests that all times are equally important in the determination of hypoxanthine and xanthine; whereas the loading of PC2 indicates that the times between 20 and 90 s are less important (a value of zero is reached); a similar explanation applies to the kinetic loadings for PC3 and PC4, indicating that data at the beginning of the reaction have more importance to the quantification of the analytes. The spectral loadings, for all the PCs, show two regions of importance, one with maxima at 250–270 nm, mainly related to hypoxanthine, and the other with maxima around 290 nm, related to uric acid. One maximum in the xanthine spectrum appears at 280 nm (Fig. 1a), but there is no important spectral loading at this wavelength.

Table 4 presents the figures of merit for the regression lines of the reference values versus the calculated ones obtained using the 3W-PLS2 model. The standard deviation of the slope and intercept are given in brackets. For the three sets of samples (calibration, validation and spiked), the slopes are very close to one, as are the determination coefficients (R^2).

Results obtained by other authors for the quantification of hypoxanthine and xanthine in human urine and spiked buffered solutions can be found in the literature [8, 10, 12]. Results obtained using the the kinetic–spectrophotometric method and 3W-PLS2 are comparable with the results obtained using various analytical methods (including HPLC and gas chromatography).

Conclusions

The potential application of three-way methodologies to enzymatic processes has been tested with good results. 3W-PCA and 3W-PLS have been shown to be an excellent alternative to unfolding methods. Qualitative information on the kinetic behavior of the system and differences between laboratory and spiked samples has been obtained using 3W-PCA.

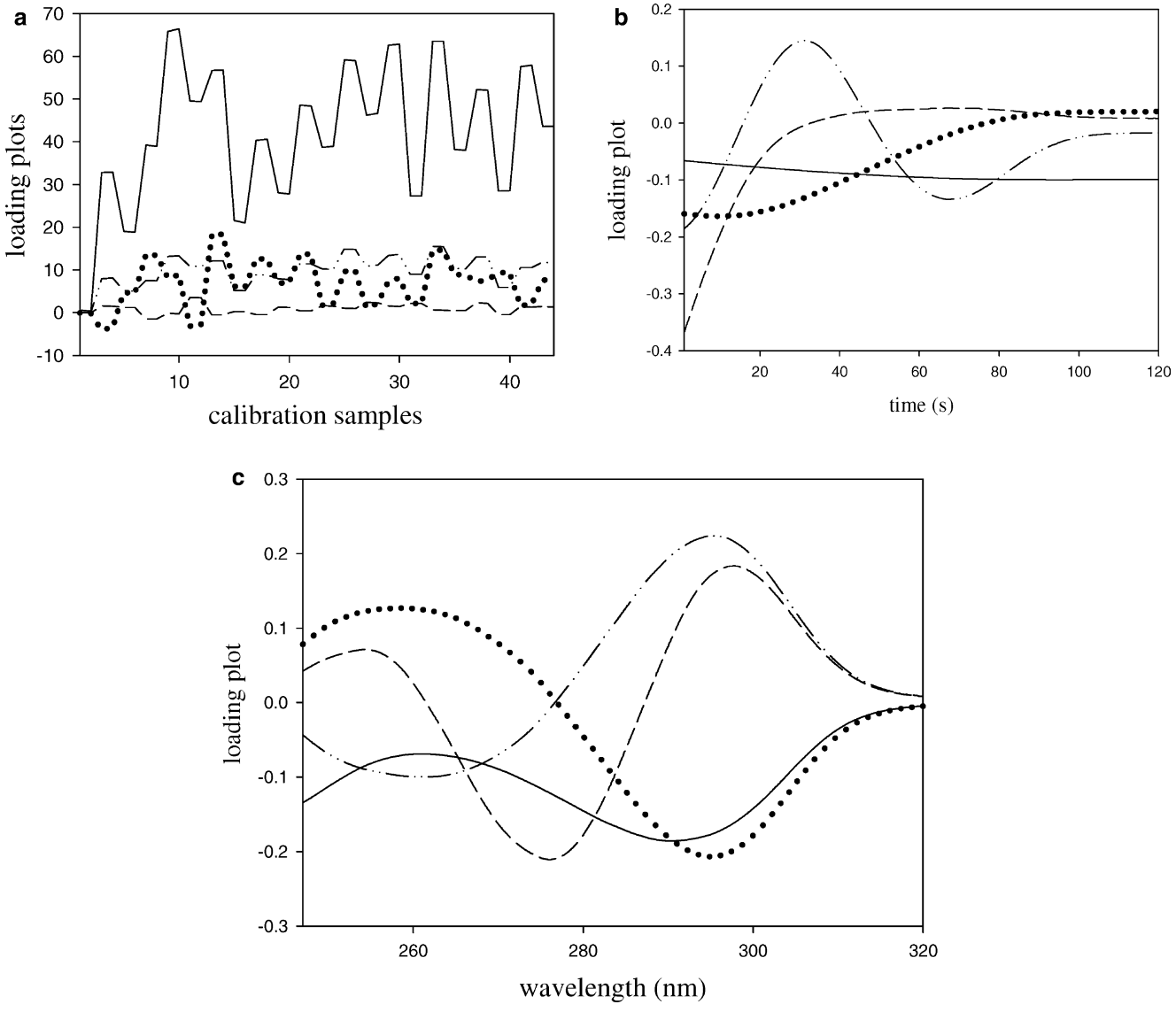


Fig. 5a–c 3W-PLS2 loadings for the quantification of hypoxanthine and xanthine (*solid line* PC1, *dotted line* PC2, *dashed line* PC3, *dash-dotted line* PC4). **a** Sample loading, **b** kinetic loading, **c** spectral loading

Table 4 Figures of merit for the regression lines of the reference values versus those calculated using the model 3W-PLS2

	Calibration	Validation	Spiked
Hypoxanthine			
Slope	0.987 (0.006)	1.000 (0.006)	0.978 (0.011)
Intercept	0.649 (0.217)	−0.438 (0.211)	0.072 (0.340)
R^2	0.999	0.999	0.998
Xanthine			
Slope	1.005 (0.019)	1.001 (0.023)	0.987 (0.030)
Intercept	–0.248 (0.358)	0.045 (0.385)	0.407 (0.509)
R^2	0.990	0.989	0.982

Values of standard deviation are in *parentheses*. R^2 is the determination coefficient

In addition, a new procedure for the simultaneous determination of mixtures of hypoxanthine and xanthine in human urine has been tested and presented, yielding good results. This analysis uses several chemometric methods and does not require any previous chromatographic separation.

The 3W-PLS2 algorithm yielded the best results for xanthine and hypoxanthine quantification. The presence of uric acid did not influence the results. The SEP and SBDR are low, indicating that the models offer good predictive capabilities and repeatabilities. The results for xanthine may indicate that the system is slightly non-linear (SEP greater than SDBR), but the predictive capabilities of the models are good enough to quantify both analytes with good results.

Acknowledgements The authors gratefully acknowledge the financial support of this work by the Ministerio de Ciencia y Tecnología (DGI BQU2001–2019) and the Comissionat per a Universitats i Recerca of the Generalitat de Catalunya (2001-SGR-00176).

References

1. Pettersson Å, Karlberg B (1997) Anal Chim Acta 354:241–248
2. Crouch SR, Coello J, Maspoch S, Porcel M (2000) Anal Chim Acta 424:115–126
3. Coello J, Maspoch S, Villegas N (2000) Talanta 53:627–637
4. Ni Y, Huang C, Kokot S (2004) Chemom Intell Lab Syst 71:177–193
5. Blanco M, Coello J, Iturriaga H, Maspoch S, Porcel M (1999) Anal Chim Acta 398:83–92
6. Escribano J, Garcia-Canovas F, Garcia-carmona F (1988) J Biochem 254:829–833
7. Tawa R, Kito M, Hirose S (1981) Chem Lett 745–748
8. Putterman GJ, Shaikh B, Hallmark MR, Sawyer CG, Hixson CV, Perini F (1979) Anal Biochem 98:18–26
9. Czauderna M, Kowalczyk J (1997) J Chromatogr B 704:89–98
10. Di Pietro MC, Vannoni D, Leoncini R, Liso G, Guerranti R, Marinello E (2001) J Chromatogr B 751:87–92
11. Chen G, Chu Q, Zhang L, Ye J (2002) Anal Chim Acta 457:225–233
12. Zen J-M, Lai Y-Y, Yang H-H, Senthil Kumar A (2002) Sens Actuator B 84:237–244
13. Pei J, Li XY (2000) Anal Chim Acta 414:205–213
14. Foppoli C, Coccia R, Cini C, Rosei MA (1997) Biochim Biophys Acta 1334:200–206
15. Carsol M-A, Mascini M (1998) Talanta 47:335–342
16. Geladi P (1989) Chemom Intell Lab Syst 7:11–30
17. Sanchez E, Kowalski BR (1990) J Chemom 4:29–45
18. Wold S, Geladi P, Esbensen K, Öhman J (1987) J Chemom 1:41–56
19. Henrion R (1994) Chemom Intell Lab Syst 25:1–23
20. Bro R (1997) Appl Spectrosc Rev 32–3:237–261
21. Tucker L (1963) Problems of measuring change. University of Wisconsin Press, Madison, WI, pp 122–137
22. Tauler R (1995) Chemom Intell Lab Syst 30:133–146
23. Bro R (1996) J Chemom 10:47–61
24. Massey V, Brumby PE, Komai H, Palmer G (1969) J Biol Chem 244–247:1682–1691
25. Jeëwska MM (1973) Eur J Biochem 36:385–390
26. Lindsey AS, Sharma RK (1981) Anal Lett 14-B10:799–811
27. Smilde AK (1997) J Chemom 11:367–377
28. Jong S (1998) J Chemom 12:77–81
29. Kiers HAL (1991) Psychometrica 56:449–470

A mixed har- and soft-modelling approach to study and monitor enzymatic systems in biological fluids.

José Manuel Amigo, Anna de Juan, Jordi Coello, Santiago Maspoch.

Anal. Chim. Acta 567 (2006) 244-254

Anexos / Annexes

Available online at www.sciencedirect.com
SCIENCE DIRECT
Analytica Chimica Acta 567 (2006) 245–254
ANALYTICA CHIMICA ACTA
www.elsevier.com/locate/aca

A mixed hard- and soft-modelling approach to study and monitor enzymatic systems in biological fluids

José Manuel Amigo [a,*], Anna de Juan [b], Jordi Coello [a], Santiago Maspoch [a]

[a] *Departament de Química, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Barcelona, Spain*
[b] *Departament de Química Analítica, Universitat de Barcelona, E-08028 Barcelona, Spain*

Received 14 December 2005; received in revised form 9 March 2006; accepted 10 March 2006
Available online 16 March 2006

Abstract

Mixed hard- and soft-modelling multivariate curve resolution (HS-MCR) is applied to study and to monitor complex enzymatic systems. Working under the basis of the soft-modelling technique multivariate curve resolution-alternating least squares (MCR-ALS), a hard constraint is introduced to force some or all concentration profiles to fulfil an enzymatic model. In this way, improvements to the application of pure hard- or pure soft-modelling are achieved.

The enzymatic reactions of different mixtures of hypoxanthine, xanthine and uric acid with xanthine oxidase are studied. This is a complex enzymatic process, where uric acid acts as a linear competitive inhibitor. The reactions were monitored with UV–vis spectrophotometry coupled to a stopped-flow module.

This work has two aims, both of them focusing on different aspects linked to modelling enzymatic systems using HS-MCR. The first goal is related to the elucidation of the real enzymatic mechanism when one of the chemical substances involved in the process apparently deviates from the mechanism found in the literature. The second one focuses on modelling the enzymatic reaction in the presence of a biological interference, such as human urine. The elucidation of the real mechanism of this enzymatic process and of the behaviour of the involved chemical species in a natural absorbing medium are good examples of situations that can benefit from mixed modelling approaches involving the best of hard- and soft-modelling methodologies.
© 2006 Elsevier B.V. All rights reserved.

Keywords: Enzymatic analysis; Xanthine oxidase; Hard-modelling; Soft-modelling; Curve resolution

1. Introduction

Nowadays, enzymes are widely used in analytical chemistry for the determination of analytes in complex matrices [1,2], in industrial process model and control [3], in food industry and in pharmacy [4]. Due to the role of enzymes in the chemical and engineering sciences, modelling of the enzymatically driven processes is needed to control certain parameters of interest, such as the evolution of the concentration of substrates and products, the rate constants and the spectra of the involved species.

Sometimes, the analysis of enzymatic reactions becomes difficult because these processes take place in biological fluids (blood, urine and plasma), which are complex absorbent interferences. A strategy commonly followed consists of studying the behaviour of the system under reliable and known conditions (i.e. aqueous buffered solutions) and, afterwards, extrapolating this knowledge to biological media. Such an approximation does not take into account either the influence of the medium in the process evolution or the contribution of the medium to the spectra recorded.

Enzymatic processes monitored spectroscopically have been traditionally analysed using hard-modelling (HM) approaches, where the parameters for a given or expected enzymatic model are mathematically fitted to equations describing the system. The parameters to calculate are, mainly, the Michaelis–Menten constant and the maximum velocity of the reaction. These classical approaches show excellent performance and results if the proposed enzymatic model is appropriate and if all the variation related to the spectrometric response is linked to the components involved in the process. Results start worsening when some uncontrolled variation in the reaction occurs during the

* Corresponding author. Tel.: +34 935811712; fax: +34 935812379.
E-mail addresses: josemanuel.amigo@uab.es, jmamigo@einstein.uab.es (J.M. Amigo).

0003-2670/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.aca.2006.03.040

process or when inert absorbing interferences are present. In the latter case, HM analysis could still give good estimated parameters, but the pure spectral profiles for the species involved in the process would be wrong [5].

Multivariate curve resolution (MCR-ALS) is a soft modelling approach (SM) to describe and model processes without explicitly using the underlying chemical model linked to them [6]. MCR-ALS decomposes the raw measurements (in our case, a series of spectra as a function of the process variable) into a bilinear model that contains the concentration profiles and pure spectra of each absorbing species contributing to the collected measurements. The main drawback of these methods is the ambiguity associated with the raw data decomposition, which is partially or completely suppressed by means of the addition of several soft-modelling constraints or by the simultaneous analysis of sets of experiments [7]. In contrast to HM approaches, MCR-ALS, as any other pure soft-modelling method, does not provide estimates of process parameters. However, it is able to model all the absorbing contributions in the original measurement, i.e., those contributing to the process of interest and those which do not.

At this point, advantages and drawbacks of pure hard- and pure soft-modelling methodologies prompt to the design of mixed approaches that can take the best of both worlds and overcome the limitations linked to the use of pure approaches. Thus, the versatility and freedom of SM approaches can be linked to the rigorous description of process contributions of HM by introducing the fulfilment of an enzymatic model as a hard constraint in the MCR-ALS algorithm. As any other constraint, the application of this hard-model constraint will be flexible and will only affect the absorbing chemical species involved in the enzymatic process. Hereafter, we use the acronym HS-MCR to refer to this mixed approach.

Compared with pure SM approaches, HS-MCR helps to minimize the ambiguity associated to the soft-modelling data decomposition and provides estimates of the process parameters as additional information. As an improvement to pure HM methods, HS-MCR can fit enzymatic processes in the presence of absorbing interferences applying the hard-modelling constraint exclusively to the species involved in the process and leaving the interferences to be soft-modelled [5,8,9].

The power of the HS-MCR methodology to model enzymatic systems will be shown through two real applications linked to the process induced by xanthine oxidase in mixtures of hypoxanthine, xanthine and uric acid. The first application is related to the elucidation of the real enzymatic mechanism when one of the chemical substances involved in the process potentially deviates from the proposed model behaviour found in the literature. The second one refers to model the enzymatic reaction when there is an absorbing interference from a biological fluid.

2. The enzymatic system: description and experimental monitoring

2.1. Process description

The enzymatic system studied is the consecutive transformation of hypoxanthine (H) and xanthine (X) to uric acid (U) by the action of xanthine oxidase (E) (xanthine:oxygen oxidoreductase, EC 1.17.3.2) (Fig. 1). This enzyme is one of the most complex enzymes of the flavoproteins. It has been recognized as an enzyme of low specificity that can catalyse the oxidation of many substances, such as purines, pyrimidines, pteridins and aldehydes [10].

Some mechanistic hypotheses have been developed based on different methodologies for one step reactions [11–14]. Escribano et al. [15] proposed a ping-pong mechanism for the oxidation of hypoxanthine and xanthine by milk xanthine oxidase including a competitive inhibition of the product, the uric acid, according to this reaction scheme:

$$\mathrm{H + E \Leftrightarrow [HE] \rightarrow X + E}, \qquad \mathrm{X + E \Leftrightarrow [XE] \rightarrow U + E},$$

$$\mathrm{U + E \Leftrightarrow [UE]}$$

Assuming the steady-state conditions, the following differential equations were obtained to describe the H, X and U variation as a function of time [16]:

$$\frac{-\mathrm{d[H]}}{\mathrm{d}t} = \frac{V^{\mathrm{X}}_{\mathrm{mx}}[\mathrm{H}]}{K^{\mathrm{X}}_{\mathrm{m}} + [\mathrm{H}] + \frac{K^{\mathrm{H}}_{\mathrm{m}}[\mathrm{U}]}{K_{\mathrm{i}}} + \frac{K^{\mathrm{H}}_{\mathrm{m}}[\mathrm{X}]}{K^{\mathrm{X}}_{\mathrm{m}}}} \tag{1}$$

$$\frac{\mathrm{d[X]}}{\mathrm{d}t} = \frac{V^{\mathrm{H}}_{\mathrm{mx}}[\mathrm{H}]K^{\mathrm{X}}_{\mathrm{m}} - V^{\mathrm{X}}_{\mathrm{mx}}[\mathrm{X}]K^{\mathrm{H}}_{\mathrm{m}}}{K^{\mathrm{X}}_{\mathrm{m}}K^{\mathrm{H}}_{\mathrm{m}} + K^{\mathrm{H}}_{\mathrm{m}}[\mathrm{X}] + \frac{K^{\mathrm{X}}_{\mathrm{m}}K^{\mathrm{H}}_{\mathrm{m}}[\mathrm{U}]}{K_{\mathrm{i}}}} \tag{2}$$

$$\frac{\mathrm{d[U]}}{\mathrm{d}t} = \frac{V^{\mathrm{X}}_{\mathrm{mx}}[\mathrm{X}]}{K^{\mathrm{X}}_{\mathrm{m}} + [\mathrm{X}] + \frac{K^{\mathrm{X}}_{\mathrm{m}}[\mathrm{U}]}{K_{\mathrm{i}}} + \frac{K^{\mathrm{X}}_{\mathrm{m}}[\mathrm{H}]}{K^{\mathrm{H}}_{\mathrm{m}}}} \tag{3}$$

In this mechanism, variation of substrate (H), intermediate (X) and product (U) concentration with time, t, is shown as a function of the concentration of these species and of the kinetic parameters of the system. $K^{\mathrm{H}}_{\mathrm{m}}$, $K^{\mathrm{X}}_{\mathrm{m}}$ are the Michaelis–Menten constants of H and X, respectively, $V^{\mathrm{H}}_{\mathrm{mx}}$, $V^{\mathrm{X}}_{\mathrm{mx}}$ are their maximum velocities and K_{i} is the competitive inhibition constant.

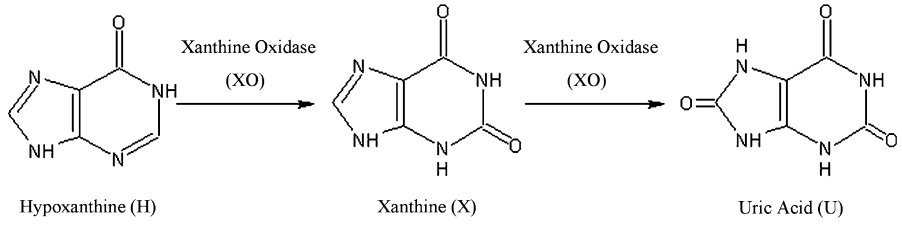


Fig. 1. Two steps enzymatic process of hypoxanthine catalysed by xanthine oxidase.

Although this is a generally accepted starting point model, some authors have postulated that uric acid is not stable in alkaline solutions and it is oxidized to other secondary products [16,17]. This degradation of the product can affect the behaviour of the uric acid and will be further studied.

The model postulated refers to aqueous solutions and a more complex situation may occur when these catalytic reactions are carried out in biological fluids, such as blood, plasma or urine [18]. The possible influence of the medium in the enzymatic process will also be assessed in this work.

2.2. Experimental monitoring

2.2.1. Reagents

All chemicals were of analytical reagent grade and purchased from Sigma–Aldrich. Stock solutions of 2.5 mmol L^{-1} of hypoxanthine, xanthine and uric acid were prepared. 0.15 U mL^{-1} enzyme xanthine-oxidase solution was used (EC 1.17.3.2; from buttermilk; 15% protein; 0.4–1.0 U mg $protein^{-1}$). All sample solutions were prepared just before use in 50 mmol L^{-1} Tris(hydroxymethyl)–aminomethane–HCl buffer at pH 8.5. Lyophilized urine Lyphochek® (Quantitative Urine Control from BIO-RAD) was regenerated with 10 mL of the buffer solution, filtered through nylon cartridges (0.45 μm) and spiked properly.

2.2.2. Instruments

An HP-8453 diode array spectrophotometer and a stopped flow system Applied Photophysics RX 2000 were used to monitor the reactions as a function of time. A spectrophotometric flow-cell of 1.00 cm path length was used throughout. The system was thermostatted by using a Peltier module Agilent 89090-A. Recorded spectra were imported from the spectrophotometer into the suitable software.

2.2.3. Experimental procedure and spectra acquisition

Enzymatic processes are highly influenced by factors such as pH, the selected buffer and its concentration, enzyme concentration and working temperature. The working conditions have to be established to guarantee reproducibility of the measurements. pH of the medium was established to be 8.5. At this pH, the enzyme has the maximum catalytic power. Temperature of the system was kept at 25.0 ± 0.1.

The performance of the experiments was as follows. The stopped-flow mixture system consisted of two parallel syringes (Fig. 2). The first syringe contained buffered xanthine oxidase solution and the second syringe the different sample mixtures (these sample mixtures were formed by combinations of some or all the species involved in the enzymatic process (H, X and U) and urine, when needed). The reactions started when 15 μL of each of these solutions were pushed into the reaction cell. The mixture of working buffer solution and enzyme was used as the reference in all experiments.

Recording of the UV–vis spectra was started 0.1 s after mixture. Spectra were recorded at 1 nm intervals over the wavelength range 225–320 nm and every 0.5 s during 180 s.

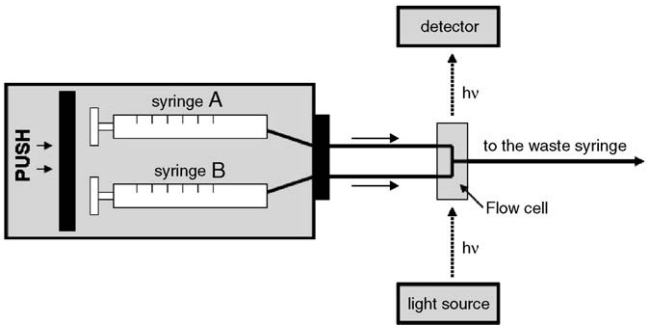


Fig. 2. Stopped-flow system coupled to an UV–vis spectrophotometer used for monitoring the enzymatic reactions.

The obtained spectral data were converted into ASCII files for further data analysis with the HS-MCR algorithm.

Table 1 shows the initial concentrations of the analytes and urine in the reaction cell for all the experiments performed. The concentration ranges of hypoxanthine, xanthine and uric acid chosen for all experiments cover the expected concentration levels in urine for healthy people and people with renal failures (gout, toxaemia, etc.) [18].

3. Data sets and data treatment

3.1. Data sets

Each experiment in Table 1 results in a data matrix, sized (number of spectra × number of wavelengths) that contains the series of spectra recorded during the enzymatic process.

In the first data set, *A*, experiments with mixtures of hypoxanthine, xanthine and uric acid at different concentration levels comprised in the range 10–60 μmol L^{-1} were carried out. The data matrices of this data set were used to elucidate the correct model for the enzymatic reaction.

Table 1
Initial concentrations (μmol L^{-1}) of hypoxanthine, xanthine, uric acid and urine in the mixtures used for the different data sets *A* and *D*

Data set	Experiments	Initial concentrations (μmol L^{-1})			
		H	X	U	% Urine
A	A_1	60	0	0	
	A_2	0	30	0	
	A_3	30.2	15.1	0	
	A_4	45.2	0	20.2	
	A_5	10.1	22.2	15.1	
	A_6	0	15.1	30.3	
D	D_1	50.0	0	0	0
	D_2	0	50.0	0	0
	D_3	50.0	0	0.9	3
	D_4	50.0	0	1.8	6
	D_5	0	50.0	0.9	3
	D_6	0	50.0	1.8	6
	D_7	30.0	30.0	0.9	3
	D_8	30.0	30.0	1.8	6
	D_9	0	30.0	30.9	3
	D_{10}	0	30.0	31.8	6

The data set D was formed by experiments involving mixtures of hypoxanthine, xanthine and uric acid in the range of concentration of 30–55 μmol L^{-1} plus different amount of urine added. The urine range was comprised from 3% to 6% of interference level. The little quantity of uric acid in the human urine, certified by the provider, has also been taken into account. The data set is composed by 10 experiments, some with urine interference and some without. This data set will be used to model the enzymatic process in the presence of biological absorbing interferences.

3.2. Data treatment

3.2.1. The underlying model of spectroscopic measurements

The data collected during the spectroscopic monitoring of a process follow the Beer–Lambert's law bilinear model. Thus, any data matrix **D** containing the series of registered spectra can be expressed as the outer product of two matrices, **C** and $\mathbf{S}^{\mathrm{T}}$, which contains the concentration profiles of each chemical species as a function of the process variable and the related pure spectra of those species, respectively. This model is expressed in Eq. (4):

$$\mathbf{D} = \mathbf{C}\mathbf{S}^{\mathrm{T}} + \mathbf{E} \tag{4}$$

D contains the raw data and is sized (number of spectra × number of wavelengths), **C** is sized (number of spectra × number of absorbing components) and $\mathbf{S}^{\mathrm{T}}$ is sized (number of absorbing components × number of wavelengths). **E** is the matrix of experimental error, not described by the $\mathbf{C}\mathbf{S}^{\mathrm{T}}$ model, and is sized as **D**. The contribution of each particular species to the raw data is described by a dyad of profiles $\mathbf{c}_i\mathbf{s}_i^{\mathrm{T}}$, where $\mathbf{c}_i$ shows the variation in concentration of species i along the process and $\mathbf{s}_i^{\mathrm{T}}$ the related pure spectrum. Any row in **D** (spectrum collected during the process) is described as the concentration-weighted sum of the pure spectra of the absorbing species present while monitoring the process.

3.2.2. General description of MCR-ALS algorithm

Methods that allow the recovery of matrices **C** and $\mathbf{S}^{\mathrm{T}}$ from the sole raw measurements, **D**, receive the name of curve resolution methods [5]. Among those, multivariate curve resolution-alternating least squares (MCR-ALS) has been widely proved to be a powerful tool to model systems of different nature [7,19,20].

MCR-ALS is an iterative resolution method widely described in previous papers [7,19,21] and the general steps of the algorithm can be summarized as follows:

1. Determination of the number of components in **D**.
2. Generation of non-random initial estimates of either the **C** or the $\mathbf{S}^{\mathrm{T}}$ matrix.
3. Given **D** and **C**, least-squares calculation of $\mathbf{S}^{\mathrm{T}}$ under constraints.
4. Given **D** and $\mathbf{S}^{\mathrm{T}}$, least-squares calculation of **C** under constraints.
5. Reproduction of **D** as $\mathbf{C}\mathbf{S}^{\mathrm{T}}$. Go to 3 till the quality in the data reproduction is satisfactory and convergence in the iterative optimisation is achieved.

MCR-ALS can be applied to an individual experiment or to a series of them simultaneously. When more than one experiment comes into play, the model in Eq. (4) extends to:

$$\begin{bmatrix} D_1 \\ D_2 \\ \cdots \\ D_n \end{bmatrix} = \begin{bmatrix} C_1 \\ C_2 \\ \cdots \\ C_n \end{bmatrix} S^{\mathrm{T}} \tag{5}$$

where now **D** and **C** are column-wise augmented data matrices, with the submatrices of the individual experiments, $\mathbf{D}_i$ and $\mathbf{C}_i$, one on top of each other. $\mathbf{S}^{\mathrm{T}}$ is a single data matrix with the shape of the pure spectra of the chemical species, common and valid for all experiments. The data arrangement in Eq. (5) implies that the shape of the concentration profile of a particular species can change in a completely free manner from experiment to experiment.

Constraints are the corner stone of the MCR-ALS algorithm. Defined as any chemical or mathematical feature that should be obeyed by the modelled profiles, they modify the **C** and $\mathbf{S}^{\mathrm{T}}$ profiles during the optimisation process so that they become chemically meaningful. As a consequence of the limitation in the shape of the profiles recovered, the ambiguity linked to the bilinear decomposition of the data set is greatly decreased. Typical constraints are non-negativity, unimodality (presence of only one maximum per profile), closure (mass balance in the concentration direction) or mathematical features linked to local rank information, such as selectivity. In augmented data sets, other conditions, like trilinearity, can be imposed under the form of constraints [19,20].

Because of the mathematical simplicity of the algorithm, constraints can be used in a very flexible way. Thus, different constraints can be applied to the profiles in **C** and in $\mathbf{S}^{\mathrm{T}}$, to the different chemical species within each of these two matrices or to the different $\mathbf{C}_i$ matrices when a series of experiments is analysed simultaneously.

3.2.3. The introduction of the enzymatic constraint: the HS-MCR variant

Based on the classical MCR-ALS algorithm and taking as starting point the consideration of the enzymatic hard model as an additional constraint, the general steps of the HS-MCR algorithm are the same as those described in Section 3.2.2, except for step 4, where the introduction of the hard model constraint is carried out.

As in previous mixed approaches of hard- and soft-modelling proposed by the authors [5,9], the hard modelling constraint acts directly on the matrix of concentration profiles, **C** (note that pure hard-modelling approaches work on the complete matrix of measurements, **D**). Thus, the soft-modelled constrained **C** profiles obtained in step 4 (see previous section) and related to absorbing species involved in the enzymatic process are used as input for a non-linear multivariate fit according to the enzymatic model in Eqs. (1)–(3). The selected columns of the matrix **C** (or of a submatrix $\mathbf{C}_i$ in the case of a series of experiments) are directly fitted to the model and there are no spectral contributions (linear parameters) to be taken into account. The resulting fitted profiles

update the soft-modelled ones and the parameters linked to the enzymatic process (Michaelis–Menten constants, maximum velocities of reaction) are obtained as additional information.

Because of the intrinsic flexibility of the basic MCR-ALS algorithm, several advantages in the modelling of enzymatic processes are derived. The first point to stress is the possibility to model enzymatic processes in the presence of inert absorbing interferences. In this case, only the concentration profiles of species participating in the process are fitted, whereas the inert contributions are simply soft-modelled. In the analysis of a series of experiments, different models could be applied to the different experiments and different species could be selected. Although a global fit of all experiments sharing the same model (as in pure HM methods) is feasible [22], individual fit of each experiment is also helpful to detect outlying experiments which, with the participation of the same chemical species (same pure spectra), can show an unexpected behaviour. Finally, the algorithm foresees the inclusion of the behaviour of non-absorbing species (sometimes called *silent* species) into the enzymatic model used in the fitting step. When this is the case, the final results of HS-MCR are the usual $\mathbf{C}$ and $\mathbf{S}^{T}$ matrices, i.e., concentration profiles and spectra of the absorbing species (no matter if in or out of the enzymatic process), the parameters estimated from the hard-model fitting and the concentration profiles of the silent species.

To our knowledge, it only exists so far an attempt to apply a mixed hard- and soft-modelling approach to enzymatic processes [23]. The present algorithm has different aspects from the one in this work and shows new real complementary applications, such as modelling these processes in the presence of interferences and using this methodology to tune a previously proposed enzymatic mechanism.

MCR-ALS with the enzymatic model constraint was implemented into the MCR-ALS algorithm working under MatLab v. 6.5 (MathWorks, Massachusetts). This modification of the algorithm is available from the authors.

3.2.4. Hard-modelling fitting

Pure hard-modelling fitting (coded HM in the text), based on non-linear fitting of the original data set D according to the procedure in Ref. [22], has been performed with the Berkeley Madonna v. 8.0.1 software. This methodology has only been applied to data sets where the enzymatic model is completely known and in the absence of interferences.

4. Results and discussion

4.1. Elucidation of the enzymatic model

As a way of example, Fig. 3a shows the raw spectra obtained for the spectroscopic monitoring of the experiment containing 30 μmol L^{-1} of hypoxanthine and xanthine. Fig. 3b corresponds to the pure spectra of H, X and U. As can be seen in Fig. 3c, the absorbance at 290 nm (maximum of uric acid) as a function of time reaches a maximum and decreases some seconds after the reaction has started. This decrease of the absorbance may support the hypothesis related to the instability of uric acid in alkaline solutions (our working pH is 8.5) where it would be oxidized, mainly, to allantoin (a species that does not absorb in the UV–vis range measured) [16,17].

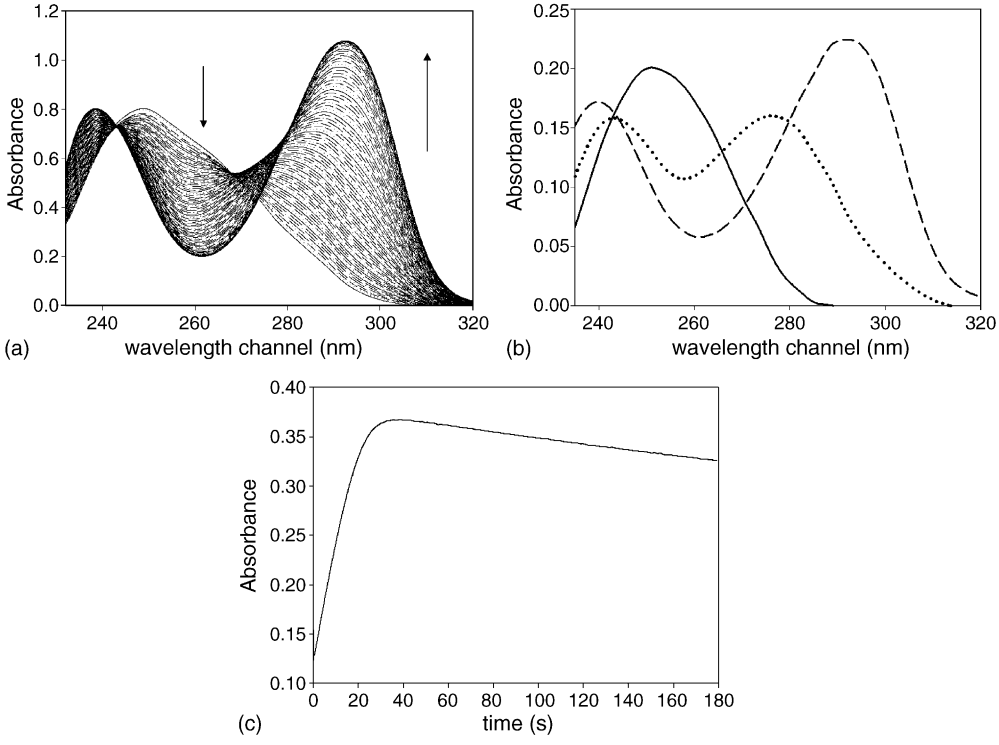


Fig. 3. (a) UV–vis monitoring of a sample containing 30 μmol L^{-1} of hypoxanthine and xanthine, respectively (A_3 experiment for the A data set). (b) Spectra of hypoxanthine (solid); xanthine (dotted) and uric acid (medium dash). (c) Time trace monitoring at 290 nm (maximum in absorbance of uric acid) of the A_3 experiment. The shape shows the decaying in the absorbance due to the oxidation of uric acid.

Table 2
Characteristics of the methods applied to elucidate the mathematical model (data set *A*) and to fit the enzymatic system in the presence of urine (data set *D*)

Analysis run no.	Data set	Data analysis[a]	Initial estimates	Constraints[b]	% Lack of fit[c]
1	A_1	SM	C (EFA)[d]	[1, 2]	1.10
2	A_2	HS-MCR	S^T (run 1)	[1, 2]	3.19
3	A_2	HM old[e]	Constants[f]		3.16
4	A_2	HM new[g]	Constants		2.99
5	$[A_1–A_6]$	HM new	Constants		1.33
6	D_1	SM	C (EFA)	[1, 2]	2.40
7	$[D_1; D_3]$	HSMCR	S^T (run 6)[h]	[1, 2, 3]	2.23
8	$[D_1–D_{10}]$	HSMCR	S^T (run 7)	[1, 2, 3]	1.78

[a] HM: hard-modelling; SM: soft-modelling; HS-MCR: mixed approach (hard-modelling + soft-modelling).
[b] Constraints applied to the concentration profiles: 1, non-negativity; 2, unimodality; 3, kinetic constraint applied to H, X and U according to the new model for uric acid. The spectra are always constrained to be non-negative.
[c] Lack off fit (%) $= 100 \times \sqrt{\sum_{i,j} e_{ij}^2 / \sum_{i,j} d_{ij}^2}$, where e_{ij} are residuals and d_{ij} are the elements of the raw data matrix.
[d] C: initial estimates are the concentration profiles; S^T: initial estimates are the spectra found in the indicated run number.
[e] Hard-modelling according to the model by Ref. [15].
[f] Initial estimates for the hard-modelling are initial values of the enzymatic constants.
[g] Hard-modelling according to the model in Eqs. (1), (2) and (7).
[h] Initial estimates are spectral profiles from run 6 plus first spectrum of matrix D_3.

To illustrate the example of application of HS-MCR to the elucidation of the enzymatic model, the experiment A_2 has been analysed (run 2 in Table 2). This experiment has been chosen because it only contains xanthine, so the kinetic reaction involved is the simplest one where uric acid participates. Initial estimates of spectra were those obtained from a previous pure SM analysis of the A_1 experiment (see Tables 1 and 2). Nonnegativity and unimodality constraints were imposed to the concentration profiles, whereas only non-negativity constraint was imposed to the spectral profiles. The application of the hard-modelling constraint in this analysis run was as follows. The enzymatic model used to fit the data was the one in Eqs. (1)–(3). The species in the model, xanthine and uric acid, have been described as absorbing (X) and non-absorbing (U). This definition of the species makes that, effectively, only the xanthine concentration profile is taken as an input profile in the enzymatic fitting; the concentration profile of the uric acid, coded as non-absorbing species, is soft-modelled.

The output of the application of the HS-MCR algorithm will be formed by the pure spectra of xanthine and uric acid (in matrix $\mathbf{S}^T$), the concentration profile of xanthine fitted according to the enzymatic model and the soft-modelled concentration profile of uric acid (in matrix $\mathbf{C}$) and an additional concentration profile, from the enzymatic fitting, which would correspond to the behaviour of the so defined silent species, uric acid.

This strategy is used to compare the concentration profile of the uric acid obtained from the absorbing contribution of this species to the raw data set with the concentration profile that would present if it followed the behaviour in the postulated enzymatic model. If the model assumed is correct, both concentration profiles would match; if systematic deviations from the theoretical behaviour are seen, the model should be reformulated.

Fig. 4 shows all the concentration profiles and spectra obtained after the application of the HS-MCR algorithm. Solid lines are profiles in $\mathbf{C}$ and $\mathbf{S}^T$ matrices and the dashed profile is the evolution that uric acid would present according to the enzymatic model. There is no match between the theoretical and the soft-modelled concentration profile contribution of uric acid. The U concentration decreases visibly with time 30 s after the reaction has begun. The shape of the degradation suggests that this decrease in absorbance could be associated with a first-order

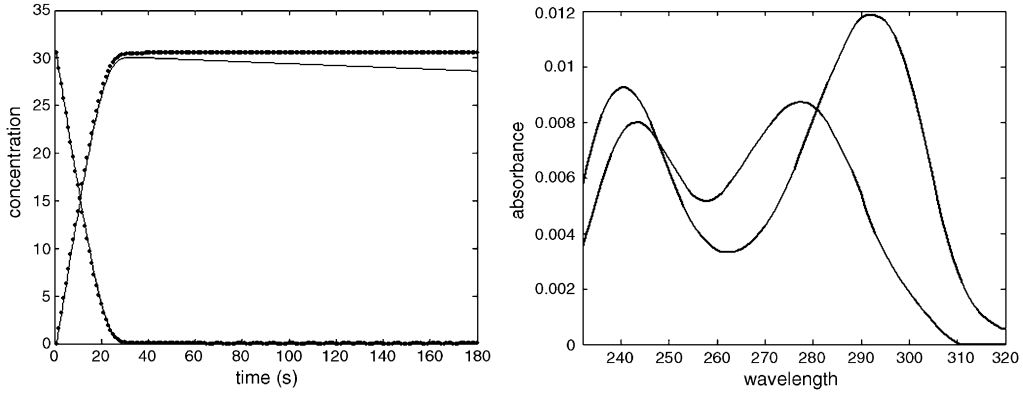


Fig. 4. Kinetic profiles and spectra recovered for A_2 experiment. Solid lines are the HS-MCR profiles. Dashed lines are the HM profiles applying the model by Ref. [15].

kinetic degradation (Eq. (6))

$$\frac{\mathrm{d[U]}}{\mathrm{d}t} = -K_\mathrm{d}[\mathrm{U}] \tag{6}$$

In this equation, the decay of uric acid is linearly dependent on its concentration and K_d is the degradation constant. The global expression for the behaviour of uric acid may be proposed by linking expressions (3) and (6) as follows:

$$\frac{\mathrm{d[U]}}{\mathrm{d}t} = \frac{V^\mathrm{X}_\mathrm{mx}[\mathrm{X}]}{K^\mathrm{X}_\mathrm{m} + [\mathrm{X}] + \frac{K^\mathrm{X}_\mathrm{m}[\mathrm{U}]}{K_\mathrm{i}} + \frac{K^\mathrm{X}_\mathrm{m}[\mathrm{H}]}{K^\mathrm{H}_\mathrm{m}}} - K_\mathrm{d}[\mathrm{U}] \tag{7}$$

Taking into consideration the behaviour for uric acid in Eq. (7), classical HM calculations have been performed for the A_2 experiment with this new enzymatic model (run 4).

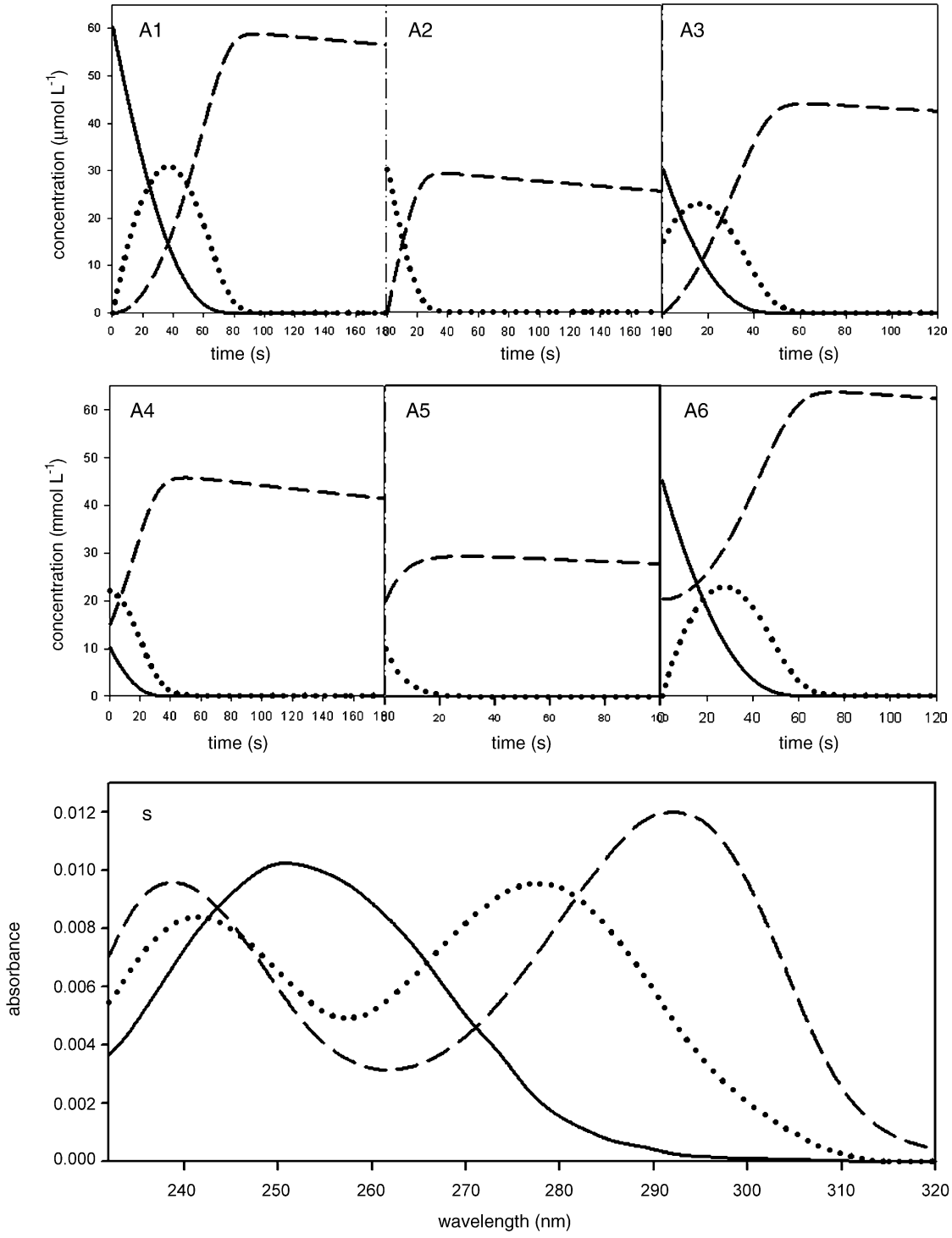


Fig. 5. Kinetic profiles and pure spectra related to the analysis of the data set *A* by means of HM application of the new model (hypoxanthine, solid; xanthine, dotted; uric acid, medium dash).

Table 3
Enzymatic constants obtained by applying the different models and methodologies proposed

Data set	Methodology	Enzymatic model	Enzymatic constants					
			V_H	K_H	K_X	K_i	V_X	K_d
A	Hard-modelling	Escribano et al. [15]	1.9 (0.1)[a]	2.2 (0.3)	4 (1)	178.1 (0.4)	1.8 (0.2)	–
	Hard-modelling	New model	1.77 (0.08)	2.9 (0.4)	6.8 (0.7)	178.0 (0.3)	2.1 (0.1)	0.0005 (0.0001)
D[b]	HS-MCR	New model	1.2 (0.2)	2.6 (0.4)	7.0 (1.0)	176 (2)	1.1 (0.1)	0.0006 (0.0002)

[a] Standard deviation between different experiments in brackets.
[b] Rate constants obtained from D_3 to D_{10} experiments.

Table 2 shows the comparison between HM fitting performed according to the old model [15] and the new model proposed. The lack of fit value is better when the new model is applied.

The general validity of the new proposed model has been checked through the global HM analysis of the whole data set *A* (run 5). In Fig. 5, the HM fitting for each experiment of data set *A* and the derived least-squares spectra are shown (results from run 4). The LOF value is satisfactory and the kinetic parameters in Table 3 agree with values in the literature [13,15]. K_d is very small in comparison with the other rate constants; this fact indicates that the degradation of uric acid is a slow process, only visible when most of the uric acid has been already formed.

HS-MCR has been applied in this section to help in the elucidation of the suitable enzymatic model of the process. Once the mechanism is set and since all the absorbing species in the *A* data set are involved in the process to be modelled, classical HM can be applied.

4.2. Process modelling in the presence of biological fluids

The second application of HS-MCR focuses on the modelling of the enzymatic process in an absorbing biological fluid. In this situation, the process takes place in the presence of different interference levels of urine, which produce an absorption that remains constant throughout the experiment.

In this kind of experiment, the presence of an inert and constant absorbing interference along the process results also in a problem of rank-deficiency. To show this fact, a rank-analysis of D_1 and D_3 experiments (the same hypoxanthine solution without and with urine, respectively) has been performed. The number of chemical compounds related to each experiment has been estimated using evolving factor analysis (EFA) [23]. Fig. 6 shows the EFA plots related to D_1 and D_3 experiment. Lines above the noise level refer to chemical compounds. Solid lines refer to the emergence of the compounds and dotted lines to the decay. As can be seen, both EFA analyses indicate that three components are present. This is a wrong assumption in D_3 since there

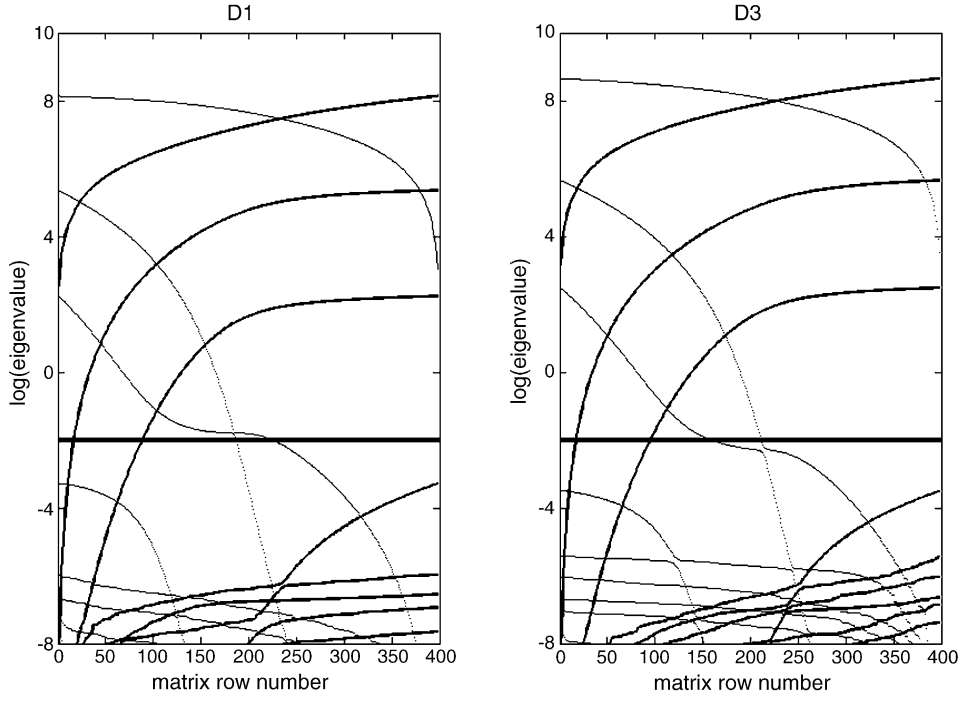


Fig. 6. Evolving factor analysis for the experiments D_1 and D_3. Bold-solid lines above noise level (thick line) point out the number and location of components appearing during the kinetic process; meanwhile dotted lines indicate the number and location of components disappearing.

is a fourth inert absorbing component. The urine has not been detected as an additional component because two closed systems evolve together, the one formed by the species involved in the enzymatic process and the one formed by the constant contribution of the urine along time. A way to resolve this rank-deficiency problem is by performing the data analysis with concatenated experiments; some without urine, D_1 and D_2, and some others with urine interference, from D_3 to D_{10}.

A way to start the data analysis is with the soft-modelling resolution of the D_1 experiment, as a way to obtain good initial estimates of the spectra linked to the H, X and U species (see Table 2, run 6). For run 6, initial estimates were built from EFA analysis. Non-negativity and unimodality constraints were imposed to the concentration profiles and the UV spectra were forced to be non-negative.

To understand the application of HS-MCR to the present example, a simple analysis (run 7) including analogous D_1 and D_3 experiments (without and with urine, respectively) is described. The initial estimates for the augmented matrix [D_1; D_3] were the spectral profiles obtained from run 6 (for H, X and U species) plus the initial spectrum of D_3 experiment (to account for the urine absorbing contribution). HS-MCR allows us to constrain the two experiments independently. HS-MCR works considering four absorbing contributions, three related to the species in the kinetic process and an additional one due to the absorbing interference. Both kinds of contributions have to be constrained according to their different nature. Thus, the concentration profiles of H, X and U have been forced to be non-negative, unimodal and to fit the enzymatic model postulated in Section 4.1 in the two experiments, whereas the concentration profile of the urine has been forced to be non-negative in the experiment D_3 because no other features about its behaviour can be assumed beforehand. UV–vis spectral profiles of the four species have been forced to be non-negative in the single matrix $\mathbf{S}^{\mathrm{T}}$. Using the appropriate set of experiments in the same analysis allows the elimination of the rank-deficiency of the experiment D_3 and the correct modelling of the whole system, leaving the urine interference out from the kinetic process, as can be seen through the good lack of fit in Table 2 and in the profiles recovered in Fig. 7.

The power of the HS-MCR approach is shown in the analysis of multiple experiments together (the 10 experiments in the complete data set D). The resolution procedure of the D data set is summarised in Table 2 (see run 8). The HS-MCR constraints have been applied to the 10 experiments according to their different nature in the same way as in run 7. Initial estimates are the pure spectral profiles resolved from run 7.

Tables 2 and 3 show the results provided by HS-MCR on the D data set. It has to be pointed out that rate constants shown in Table 3 for D data set only refer to D_3–D_{10} experiments,

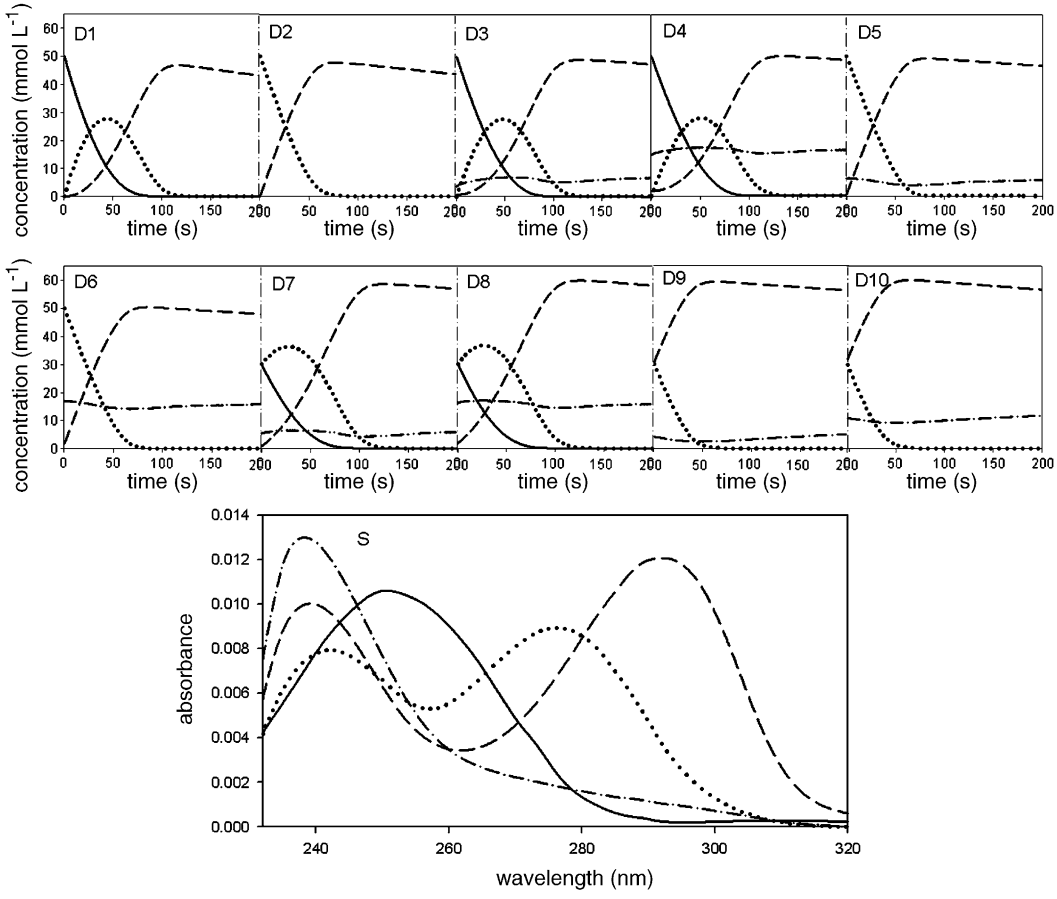


Fig. 7. Kinetic profiles and pure spectra related to the analysis of the data set D by means of HS-MCR application (hypoxanthine, solid; xanthine, dotted; uric acid, medium dash; urine, dotted-dash).

those with urine interference. The results in Table 3 are averaged values from the individual hard-modelling fit in experiments D_3–D_{10}; the small dispersion on the obtained values confirms the general model validity and the quality of the experimental data. Fig. 7 shows the resolved concentration profiles and spectra for the D data set. Spectral profiles very similar to those from the A data set (without urine) have been recovered for H, X and U compounds since the contribution due to the urine interference could be modelled separately. As expected, the urine concentration profile appears as a constant contribution along time and the two levels of urine concentration have been modelled correctly in the different experiments. Concentration profiles of H, X and U behave according to the proposed enzymatic model.

In general, rate constants obtained in the presence of urine do not differ significantly from those obtained without it (see results for data set A with the same model) and only slight differences are seen in some other parameters, like V_X. It is important to note that the individual fit of the data in each experiment to the enzymatic model would have allowed the detection of differences in these parameters among experiments with and without urine if they had existed.

5. Conclusions

The modelling of the reaction of hypoxanthine, xanthine and uric acid mixtures induced by xanthine oxidase has been successfully carried out. The combination of hard-modelling and soft-modelling in the HS-MCR algorithm has overcome the potential of both pure methodologies for the study of enzymatic processes. Different application of the enzymatic model constraint in the HS-MCR algorithm has helped to elucidate the real enzymatic mechanism of the process and to model the process in the presence of the absorbing interference originating from a biological fluid.

The proposal of the new enzymatic mechanism, that includes the oxidation of uric acid, was suggested by the mismatch in behaviour between the concentration profile of the soft-modelled uric acid and the concentration profile related to the model proposed in the literature (obtained from the inclusion of uric acid as silent species in the non-linear fitting step of HS-MCR). This same strategy can be used in other process examples to check whether the real evolution of a species, described by the variation in the recorded experimental absorbance, matches the behaviour postulated by a particular model.

HS-MCR has demonstrated to be a useful tool when the enzymatic reaction takes place into an absorbing biological fluid. By applying the enzymatic model constraints only to hypoxanthine, xanthine and uric acid and allowing urine to be soft-modelled, the estimation of the rate constants and the recovery of the pure spectra and the concentration profiles of the four components in the system were possible. This data analysis helps to separate and model correctly the model driven contributions from the inert species. Moreover, the effect of the presence of biological fluids or, in a general case, any other absorbing media, on the process parameters, could also be detected through variations in the parameters estimated from the individually fitted models in the different experiments.

Acknowledgements

José Manuel Amigo, Jordi Coello and Santiago Maspoch want to thank to the Spanish Programme of Chemical Process Technologies (Projects CTQ 2004-2013) for the finantial support.

References

[1] J.M. Amigo, J. Coello, S. Maspoch, Anal. Bioanal. Chem. 382 (2005) 1380.
[2] M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, M. Porcel, Anal. Chim. Acta 398 (1999) 83.
[3] N. López, R. Pérez, F. Vázquez, F. Valero, A.J. Sánchez, Chem. Technol. Biotechnol. 77 (2002) 175–182.
[4] http://www.enzymes.co.uk/.
[5] A. de Juan, M. Maeder, M. Martínez, R. Tauler, Chemom. Intell. Lab. Syst. 54 (2000) 123.
[6] A. de Juan, E. Casassas, R. Tauler, Encyclopedia of analytical chemistry: instrumentation and applications, in: Soft-Modelling of Analytical Data, Wiley, New York, 2000.
[7] J. Jaumot, R. Gargallo, A. de Juan, R. Tauler, Chemom. Intell. Lab. Syst. 76 (2005) 101.
[8] A. de Juan, M. Maeder, M. Martínez, R. Tauler, Anal. Chim. Acta 442 (2001) 337.
[9] J. Diewok, A. de Juan, M. Maeder, R. Tauler, B. Lendl, Anal. Chem. 75 (2003) 641.
[10] M.M. Jeżewska, Eur. J. Biochem. 36 (1973) 385.
[11] R.B. McWhirter, R.J. Hille, Biol. Chem. 266 (35) (1991) 23724.
[12] J.S. Olson, D.P. Ballou, G. Palmer, V. Massey, J. Biol. Chem. 249 (14) (1974) 4363.
[13] V. Massey, P.E. Brumby, H.J. Komai, Biol. Chem. 244–247 (1969) 1682.
[14] J.S. Olson, D.P. Ballou, G. Palmer, V. Massey, J. Biol. Chem. 249 (14) (1974) 4350.
[15] J. Escribano, F. Garcia-Cánovas, F. Garcia-Carmona, Biochem. J. 254 (1988) 829.
[16] P.M.S. Clark, L.J. Kricka, A. Patel, Anal. Lett. 12 (B15) (1979) 1537.
[17] A.S. Lindsey, R.K. Sharma, Anal. Lett. 14 (B10) (1981) 799.
[18] R. Tawa, M. Kito, S. Hirose, Chem. Lett. (1981) 745.
[19] R. Tauler, A.K. Smilde, B.J. Kowalski, J. Chemom. 9 (1995) 31.
[20] A. de Juan, R. Tauler, Anal. Chim. Acta 500 (2003) 195.
[21] R. Tauler, Chemom. Intell. Lab. Syst. 30 (1995) 133.
[22] M. Maeder, Anal. Chem. 59 (1987) 527.
[23] R. Sánchez-Ponce, S.C. Rutan, Chemom. Intell. Lab. Syst. 77 (2005) 50.

A mixed hard- and soft-modelling approach for the quantitative determination of oxipurines and uric acid in human urine.

José Manuel Amigo, Anna de Juan, Jordi Coello, Santiago Maspoch.

Anal. Chim. Acta 567 (2006) 236-244

Anexos / Annexes

Available online at www.sciencedirect.com
SCIENCE DIRECT
Analytica Chimica Acta 567 (2006) 236–244
ANALYTICA CHIMICA ACTA
www.elsevier.com/locate/aca

A mixed hard- and soft-modelling approach for the quantitative determination of oxipurines and uric acid in human urine

José Manuel Amigo [a,*], Anna de Juan [b], Jordi Coello [a], Santiago Maspoch [a]

[a] *Departament de Química, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Barcelona, Spain*
[b] *Departament de Química Analítica, Universitat de Barcelona, E-08028 Barcelona, Spain*

Received 14 December 2005; received in revised form 9 March 2006; accepted 10 March 2006
Available online 16 March 2006

Abstract

A new treatment of kinetic-enzymatic data, joining the benefits of hard- and soft-modelling methods, is proposed for the kinetic resolution and simultaneous quantification of xanthine oxidase induced oxipurines (hypoxanthine and xanthine) and uric acid in human urine without any further separation.

These three analytes show highly overlapped spectral bands in the range of 232–320 nm and urine contains absorbing species in this wavelength range that behave as a constant interference throughout the monitoring of the kinetic experiment. The proposed method allows the determination of the analytes in human urine by simultaneous analysis of the urine sample and a few (even only one) standards of pure analyte mixtures of hypoxanthine and xanthine in aqueous solution.

An iterative mixed hard- and soft-modelling multivariate curve resolution (HS-MCR) algorithm, which includes a hard-modelling constraint based on the enzymatic model in the parent multivariate curve resolution-alternating least squares method (MCR-ALS), is applied to a data set formed by the standards and the urine samples. Hard-modelling is applied to the concentration profiles of the three analytes in the standards, whereas soft-modelling on the analytes and the urine in the samples gives the quantitative information. As a result, quantitative information similar in quality to that obtained by separation techniques (HPLC and HPCE) and other chemometric approaches (PLS and N-PLS) is obtained with less experimental effort and a much smaller number of standards in aqueous solution that do not need to contain the interferences present in the samples.
© 2006 Elsevier B.V. All rights reserved.

Keywords: Enzymatic data analysis; Oxipurines; Human urine; Hard-modelling; Soft-modelling; Multivariate curve resolution

1. Introduction

The spectroscopic monitoring of kinetic processes has experienced a great development in the last years and, nowadays, there are instruments that are able to follow fast catalytic reactions in real time. Kinetic monitoring has two main and different purposes: the first one focused on obtaining qualitative information, through the knowledge of the evolution of a particular process as a function of time, identifying the species involved, the underlying reaction model and the derived kinetic parameters, and the second one focused on getting quantitative information, taking advantage of the different behaviour that diverse species show within a kinetic process. No matter what the final goal is, there is an increase in the demand of powerful mathematical algorithms, able to extract and select the relevant analytical information from evolving kinetic systems. The joint of instruments recording high quality kinetic information, i.e., full spectrum responses in very short time scales, and chemometrics has allowed the direct qualitative and quantitative analysis of unresolved complex mixtures in natural samples [1].

When kinetic multicomponent analyses are used for quantitative purposes, multivariate calibration methods, such as partial least squares (PLS), are often the choice to correlate all the spectra registered in an experiment with the concentration of some or all the involved species and to predict the analyte(s) concentrations in new samples [2]. The main drawback of these multivariate regression techniques is the large number of calibration samples needed to build the calibration model, since all possible analytes and interferences have to be included in the calibration set in suitable concentration levels to

* Corresponding author. Tel.: +34 935811712; fax: 34 935812379.
E-mail address: josemanuel.amigo@uab.es (J.M. Amigo).

0003-2670/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.aca.2006.03.041

obtain a robust regression model with a good predictive ability [3,4].

Within the group of kinetic processes, enzymatic processes, particularly those using non-selective enzymes, have become a powerful tool for the resolution and quantitation of multicomponent mixtures [5–11]. An enzymatic process is an evolving system, where the catalytic reaction of some analytes with the same enzyme is monitored by means of the acquisition of multiple spectra along time. Enzymatic reactions work under mathematical relationships established between the kinetic evolution of the involved species and their concentration. The spectroscopic variation of the monitored system can be properly described by these enzymatic laws through the use of hard-modelling methodologies, provided that there are no species other than those involved in the process contributing to the spectroscopic signal measured. This is the reason why the application of pure hard-modelling methodology can be difficult in natural samples, where the reactions usually take place in the presence of absorbing interferences from the natural matrix which do not participate in the monitored process [3,12].

Multivariate curve resolution-alternating least squares (MCR-ALS) is a well-known resolution methodology that has been often used to model processes [13–16]. Being a soft-modelling methodology, MCR-ALS is able to extract the information of the evolving systems without using the underlying kinetic/enzymatic model. MCR-ALS models the data as the sum of all the pure concentration-weighted contributions to the measured signal, those linked to species in the enzymatic process and those to species out of it. As a consequence, errors caused by the assumption of a wrong model are avoided and the presence of absorbing interferences is no longer a problem. The main drawback of MCR-ALS with respect to hard-modelling methodologies is the ambiguity associated with the resolved profiles, which can affect to a major or minor extent the quality of the derived qualitative and quantitative information [13,17].

A step forward in solving the drawbacks of pure hard- and soft-modelling methods involves joining both methodologies in mixed hard- and soft- multivariate curve resolution" methods (HS-MCR) [3,12,18–21]. Thus, introducing the fulfilment of an enzymatic model by the species in the process as an additional constraint in MCR-ALS helps to minimize the ambiguities in the resolution results and to obtain additional information about the system, i.e., apart from the spectra and time-dependent profiles of all the absorbing species in the data set, the kinetic parameters linked to the process are obtained.

In a previous work, the HS-MCR methodology was developed and used to obtain the correct enzymatic model and the related enzymatic parameters of a process in samples containing urine [22]. The aim of the present work is to demonstrate that HS-MCR approaches can be used to obtain quantitative information. This possibility is shown through the kinetic resolution and simultaneous quantification of oxipurines (hypoxanthine and xanthine) and uric acid in human urine, without any further separation. MCR-ALS is applied to an augmented matrix formed by only two standards of hypoxanthine and xanthine, respectively, and samples of urine. The hard-modelling constraint is only applied to the three analytes involved in the kinetic reaction in the standard experiments, whereas pure soft-modelling is used to model the three analytes and the urine interference in the rest of the samples. As a result, correct quantitative information similar in quality to that obtained by classical methods (for example, HPLC and GC) and other chemometrics approaches (unfolded PLS and N-PLS) is obtained.

2. Experimental

2.1. Reagents and instrumentation

Hypoxanthine (H), xanthine (X) and uric acid were purchased from Merck and were of analytical grade. Xanthine oxidase (EC 1.17.3.2; from buttermilk; 15% protein; 0.4–1.0 U mg^{-1} protein; Merck) was used as catalytic agent. TRIS buffer solution was prepared from tris(hydroxymethyl)-aminomethane (Merck, p.a.) and hydrochloric acid (Panreac, 37%, p.a.). Lyophilized Lyphocheck urine was purchased from BIO-RAD.

A stopped-flow system (Applied Photophysics RX 2000) coupled to a diode array spectrophotometer (HP-8453) was used to record the spectral kinetic profiles. A spectrophotometric cell of 1.00 cm path length was used throughout the experiments. All the system was thermostatted at 25.0 ± 0.1 °C with a Peltier module (Agilent 89090-A). The room temperature was kept at 24 ± 1 °C.

2.2. Chemical system and experimental procedure

Xanthine oxidase (EC 1.17.3.2) catalyzes the oxidation of hypoxanthine to xanthine and to uric acid, the last acting as a competitive inhibitor of the previous reactions. The real mechanism of this enzymatic process has been elucidated in a previous work [22]. When the enzymatic reaction of hypoxanthine or xanthine in urine is monitored using a diode array UV–vis spectrophotometer coupled to a stopped-flow module, it can be observed that the three involved analytes show highly overlapped bands in the range of 232–320 nm (Fig. 1). Furthermore, urine contains absorbing species in this wavelength range that behave as a constant interference throughout the monitoring of the kinetic experiment. This fact prevents the correct prediction of the concentration of the analytes in the presence of urine by pure hard-modelling approaches. This enzymatic process is followed in all the experiments described below.

To carry out the experiments, a 50 mmol L^{-1}, pH 8.5, Tris buffer solution was prepared using Milli-Q water. Lyophilized urine was regenerated with Milli-Q water and filtered through nylon cartridges (0.45 μm). This solution was diluted 1:50 with the buffer solution. A solution of 0.14 U mL^{-1} of xanthine oxidase was prepared, also by diluting it in the buffer solution.

Two buffered aqueous solutions of H and X of 100 μmol L^{-1} were prepared and used as calibration samples. It has to be pointed out that it is not necessary the inclusion of a standard solution of uric acid because it is the product of the reactions involved in the enzymatic system and the whole kinetic information is involved into the two standards prepared.

Nineteen mixture samples of H, X and uric acid were also prepared in the range of concentration of 0–100 μmol L^{-1}. These

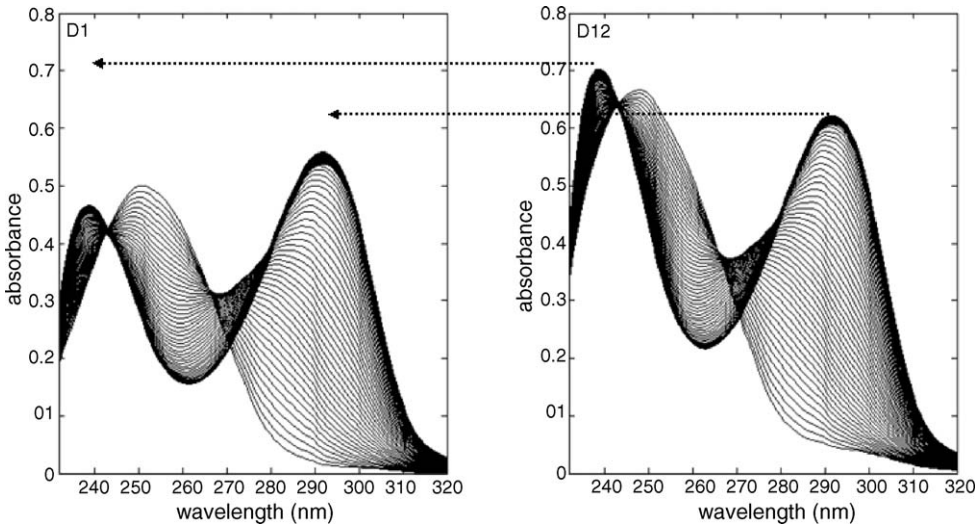


Fig. 1. Comparison between two kinetic evolutions recorded containing 50 μmol L^{-1} of hypoxanthine (D_1) and the same quantity of hypoxanthine plus urine (D_{12}).

solutions were spiked with different amounts of regenerated urine to obtain different levels of interference. In Table 1, the composition of the mixtures is described. The urine used in this experiment was free of hypoxanthine and xanthine, but contained a small quantity of uric acid. This uric acid concentration is reported by the manufacturer and has been taken into account in the quantification, as it can be seen in Table 1. The concentration values in Table 1 are the concentrations of the analytes in the reaction cell at time zero, before the enzymatic process starts. All buffered sample solutions of analytes and urine were freshly prepared before their use.

Table 1
Composition of two standard solutions (D_1 and D_2) and 19 mixtures (from D_3 to D_{21}) containing urine (expressed in % of volume)

Sample	Concentration[a] (μmol L^{-1})			Urine (%)
	Hypoxanthine	Xanthine	Uric acid	
D_1	50	0	0	0
D_2	0	50	0	0
D_3	20	0	0.23	0.8
D_4	20	0	0.69	2.4
D_5	20	0	1.85	6.4
D_6	0	50	1.85	6.4
D_7	0	20	0.69	2.4
D_8	0	20	0.23	0.8
D_9	0	20	0.69	2.4
D_{10}	0	20	1.85	6.4
D_{11}	0	50	0.69	2.4
D_{12}	50	0	0.69	2.4
D_{13}	0	0	51.85	6.4
D_{14}	30	30	0.69	2.4
D_{15}	30	30	1.85	6.4
D_{16}	30	0	30.69	2.4
D_{17}	30	0	31.85	6.4
D_{18}	0	30	30.69	2.4
D_{19}	0	30	31.85	6.4
D_{20}	30	30	30.69	2.4
D_{21}	30	30	31.85	6.4

[a] Related to the initial concentration in the flow-cell.

The stopped flow system consisted of two parallel syringes, A and B [22]. In syringe A, the buffered xanthine oxidase was placed; while syringe B contained the mixture samples or the standard solutions. The reaction started when aliquots of the two solutions in syringes A and B were pushed into the reaction cell in a 1:1 ratio. The volume of the reaction cell was 30 μL and the final concentration of the enzyme was 0.07 U mL^{-1}. A solution containing buffered xanthine oxidase was used as a blank. 0.3 s after mixing, the spectrophotometer started to record the UV–vis spectra every 1 s for 180 s, at 1 nm interval over the wavelength range of 232–320 nm.

Recorded spectra were converted into ASCII files to be imported into the data analysis software. For MCR-ALS, an enzymatic hard-model adaptation of the algorithm implemented by the authors as Matlab code was used under MatLab 7.0. The implementation of the enzymatic model as a hard-model constraint to MCR-ALS was presented in the previous work [22] and is available from authors.

3. Data treatment

3.1. Hard- and soft-modelling constrained MCR (HS-MCR) for quantitative purposes

Since this work is closely connected to a previous article, where the implementation of the hard-modelling enzymatic constraint is described in detail [22], a summarized explanation of the algorithm, stressing the aspects related to the quantitative purpose of the work, will be presented.

HS-MCR is an iterative resolution method that decomposes a data matrix with mixed information about a multicomponent system, **D**, into the product of two small matrices, **C** and $\mathbf{S}^T$, that contain the information about the pure component contributions to the global measured signal in their column and row profiles, respectively [12]. For a spectroscopically monitored enzymatic process, the matrices in the resolution model:

$$\mathbf{D} = \mathbf{C}\mathbf{S}^T + \mathbf{E}$$

are the spectra collected as a function of time (**D**), the kinetic concentration profiles of the different species in the process (**C** columns) and their related pure spectra ($\mathbf{S}^T$ rows). **E** represents the experimental error.

HS-MCR shares the main steps of the parent algorithm, multivariate curve resolution-alternating least squares (MCR-ALS) [14] that are summarised below:

1. Determination of the number of components in the raw data set.
2. Generation of initial estimates of either concentration profiles (**C** matrix) or spectra ($\mathbf{S}^T$ matrix).
3. Taking the original data matrix **D** and e.g., the initial **C** matrix, least-squares calculation of $\mathbf{S}^T$ under constraints.
4. Taking the original **D** and the calculated $\mathbf{S}^T$, least-squares calculation of **C** under constraints.
5. Reproduce the original data with the $\mathbf{CS}^T$ model and go back to step 3 if convergence is not yet achieved.

Convergence is achieved when the relative difference in lack of fit between two consecutive iterations goes below a threshold value (often 0.1%) or when a preselected number of iteration is exceeded.

The percentage of lack of fit (% LOF) gives a measure of the fit quality and has been calculated as follows:

$$\%\,\mathrm{LOF} = 100 \times \sqrt{\frac{\sum_i^I \sum_j^J e_{mn}^2}{\sum_i^I \sum_j^J d_{mn}^2}} \quad (2)$$

where e_{mn} is the mnth element of the residual matrix **E**, and d_{mn} is the mnth element of the **D** matrix.

Constraints are chemical or mathematical conditions of our system that are systematically fulfilled by some (or all) of the profiles in the **C** and/or $\mathbf{S}^T$ matrices. The most usual ones are described in the MCR literature [15] and those specifically applicable to the concentration profiles and spectra of our enzymatic data are discussed in detail in the previous article of this series [22]. The main difference between the HS-MCR algorithm and the pure soft-modelling MCR-ALS is the introduction of the physicochemical law that controls the enzymatic process as a hard-modelling constraint. Thus, the concentration profiles of the absorbing species involved in the process are fitted according to the suitable enzymatic law. To do so, a definition of the differential equations controlling the process is needed [22] and so are the concentrations of the species involved at time zero. Other absorbing species in the system, e.g., inert absorbing interferences, are simply soft-modelled.

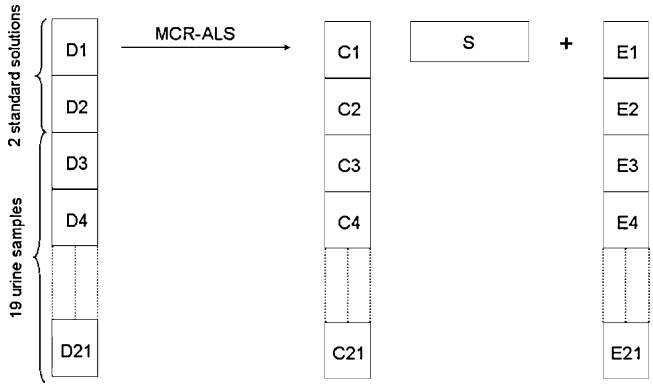


Fig. 2. Column-wise arrangement of the matrices when MCR-ALS is applied.

Getting quantitative information from a HS-MCR algorithm needs the simultaneous analysis of several matrices (standards and samples) in which the enzymatic process has taken place. These matrices are arranged in a column-wise fashion, as can be seen in Fig. 2. As a resolution result, a single spectra matrix $\mathbf{S}^T$ is obtained, valid for all standards and samples, and as many C_i submatrices as total number of samples and standards present

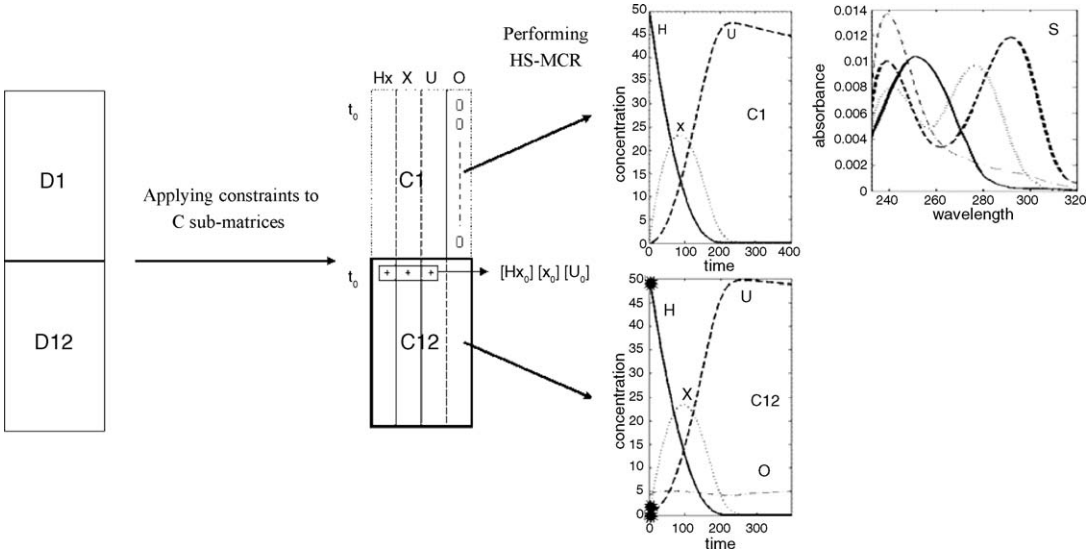


Fig. 3. Graphical scheme related to the HS-MCR resolution of the augmented matrix $[D_1; D_{12}]$ containing an experiment without urine interference (D_1) and with urine (D_{12}). D_1 is fitted to an enzymatic model and D_{12} is soft-modelled. Concentrations at time zero for D_{12} (quantitative information) are obtained (asterisks (*) in the plots).

in the analysis. The quantitative information sought, i.e., the amount of each of the analytes (H, X and U) in the original mixture urine sample, is contained in the first row of the related C_i submatrix. This row has the concentrations of the analytes at time zero, before the enzymatic process has started (see Fig. 3).

There are some constraints that do help to improve the quality of the quantitative information. One of them is the soft-modelling constraint called correspondence between species [23]. This constraint uses the information related to the identification of the species in the different samples, if known, and forces the absent species to have null concentration profiles. Such information is coded in the so-called isp matrix. This matrix has a number of rows equal to the number of concentration submatrices (samples) in the analysis and a number of columns equal to the total number of components present in the system. The presence or absence of a particular species in a submatrix is coded by 1 or 0, respectively. In our enzymatic runs, this information is only strictly available for the standards because the samples, as happens in reality, are assumed to have an unknown composition. It is important to note that, since we work with enzymatic experiments, we should consider the species existing in our original sample and those formed by the evolution of the process with time as present species in the isp matrix. An isp matrix related to our two standards of H (D_1) and X (D_2) and a urine sample (D_3) would look as follows:

$$\begin{array}{c} \\ D_1 \\ D_2 \\ D_3 \end{array} \begin{array}{c} \begin{array}{cccc} Hx & X & U & O \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \end{array}$$

O stands for urine, the fourth absorbing species in the system. The known information of our standards is that they are prepared in aqueous solution and, hence, no contribution of urine should present (0 in O column) and that the H standard contains this species and X and U, by kinetic evolution of the process, whereas the X standard has this analyte and can only yield U through the process. In the case of any of the urine samples, there should always be a 1's vector because urine is always present (1 in the O column) and the other three species (analytes) can be potentially present as well since the sample composition is unknown.

The application of the hard-modelling constraint to the concentration profiles in the standard matrices plays also an important role in the quality of the quantitative information recovered. The positive effect of this constraint responds to two main reasons:

(a) The decrease in ambiguity of the resolved profiles due to the fulfilment of the enzymatic law. As a hard-modelling constraint, the concentration profiles and the linked spectra are much better defined than if only soft-modelling constraints were applied during the resolution process. This results in a better quality in the qualitative and quantitative information.
(b) The fact that the quantitative information carried by the standards is embedded in the enzymatic model. Thus, the enzymatic model fitting needs both the enzymatic laws and the initial concentrations of the analytes to be carried out. These concentrations at time zero, which are fixed during the fitting, are the quantitative information of the standards. Not only this is fixed during the resolution process but, at the same time, this breaks the typical intensity ambiguity linked to resolution methods, which usually leads to obtaining relative quantitative information [23]. In this case, the concentrations at time zero in the samples are absolute concentrations since the reference in the standards are also real concentration values.

The hard-modelling constraint cannot be applied to the urine samples because the initial concentrations of the three analytes, needed to fit the enzymatic model, are actually the parameters sought in the quantitative analysis.

3.2. Validation of quantitative results

As in any quantitative analysis performed by other techniques, the following figures of merit have been calculated to validate the results obtained: root mean square error of prediction (RMSEP) and standard error of prediction (SEP) (Eqs. (2) and (3)) where the bias of the model was taken into account (average value of residuals) (Eq. (4)) to assess the precision and accuracy of concentration prediction for the three analytes [1]:

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{c}_i - c_i)^2}{n}} \tag{3}$$

$$\text{SEP} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{c}_i - c_i - \text{bias})^2}{n-1}} \tag{4}$$

$$\text{bias} = \frac{\sum_{i=1}^{n}\hat{c}_i - c_i}{n} \tag{5}$$

In these expressions, c_i is the reference concentration, $\hat{c}_i$ the calculated concentration, and n is the number of samples.

4. Results and discussion

The steps followed in the analysis of the samples prepared are described below. Pure soft-modelling MCR-ALS was applied to the experiment D_1, which contains the H standard, to obtain initial estimates of the spectra of H, X and uric acid (U) (run 0). The constraints of non-negativity and unimodality were applied to the kinetic profiles, whereas the spectra were constrained to be non-negative. Other options to have non-random initial estimates could have been considered, such as SIMPLISMA for spectra selection [15] or evolving factor analysis [24] for concentration profiles.

To show the use of HS-MCR for quantitative analysis with a simple example, the analysis of the augmented matrix $[D_1; D_{12}]$ (run 1, Table 2), formed by the H standard and a urine sample, is carried out. As it was shown in the previous work [22], one of the problems of the enzymatic system in the urine samples is the constant interference of urine throughout the reaction (see Fig. 1 and note the significant spectral differences between both experiments, although D_{12} has exactly the same composition as D_1,

Table 2
Method application for the analysis of enzymatic system

Analysis run no.	Dataset	Data analysis[a]	Initial estimates[b]	Constraints[c]		% Lack of fit (% LOF)
				Standards	Samples	
0	D_1	SM	**C** (EFA)	[1, 2]	–	1.24
1	[D_1; D_{12}]	HS-MCR	$\mathbf{S}^T$ (run 0)[d]	[1, 2, 3, 4]	[1, 2, 4]	1.40
2	[D_1; D_2; D_3–D_{21}]	HS-MCR	$\mathbf{S}^T$ (run 1)	[1, 2, 3, 4]	[1, 2, 4]	1.35
3	[D_1; D_2; D_3–D_{21}]	HS-MCR	$\mathbf{S}^T$ (run 1)	[1,2, 3, 4]	[1, 2, 4]	1.80

For a further description of the data matrices, see related text.

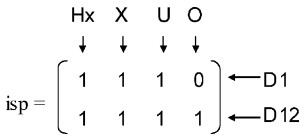
[a] SM: pure soft-model (MCR-ALS). HS-MCR: hard–soft-multivariate curve resolution.

[b] **C**: concentration profiles; $\mathbf{S}^T$: spectral profiles obtained in the indicated run.

[c] For the concentration profiles: 1, non-negativity; 2, unimodality; 3, kinetic fit; correspondence of species. The spectra were always constrained to be non-negative. Standards refer to D_1 and D_2 experiments; whereas samples refer to D_3 to D_{21} experiments (those that contain urine).

[d] The initial estimates correspond to the spectral profiles obtained in the run 0 plus the first spectrum obtained in D_{12} matrix as an estimate of the urine spectrum.

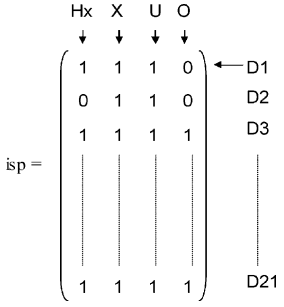
except for the presence of urine). Initial estimates are the spectra obtained in run 0 for the three analytes H, X and U, and the first spectrum of the D_{12} matrix as a representation of the urine contribution to the signal. The versatility of applying HS-MCR methodology lays in the possibility of constraint in a different way the C_i profiles of each sub-matrix D_i of the augmented matrix. As mentioned in the data treatment section, the kinetic profiles of the experiment D_1, which contained a pure standard of H without urine, were constrained to obey the enzymatic model describing the system $H \rightarrow X \rightarrow U$. In this way, the rotational and intensity ambiguity on the profiles of H, X and U was suppressed. On the contrary, all the concentration profiles linked to the D_{12} urine sample experiment were only soft-modelled. In both experiments, non-negativity and unimodality constraints were imposed to the concentration profiles, whereas only the non-negativity constraint was imposed to spectral profiles. For this analysis, the isp matrix used was:



It is known beforehand that D_1 did not contain urine. In addition, H will be catalyzed to X and U, two absorbing species. On the contrary, the only fact known from experiment D_{12} is that it contains urine. Therefore, the input for D_{12} was a 1's row, confirming the presence of urine and the potential occurrence of the three analytes.

As the outcome of the HS-MCR methodology, quite good kinetic profiles (C_1 and C_{12}) and spectra for H, X and U were obtained, the related kinetic parameters for experiment D_1 and, as the main purpose was, the concentration of the three analytes in the urine sample used for the D_{12} experiment. This information is contained in the first row of the C_{12} submatrix, which represents the amount of these species at time zero. The urine interference was soft-modelled and an approximately constant concentration profile (expected for an inert interference) and a spectral profile closer to the real contribution (Fig. 3) were obtained. The resolved urine spectrum will be used as initial estimate for this contribution in the next analysis.

The analysis of all 21 data sets presented in this article, which includes few standards and many samples containing urine, is now presented. The augmented matrix used in this run (called run 2) includes two standard samples without urine, D_1 and D_2, and the rest of experiments containing urine (D_3–D_{21}). The spectra obtained from run 1 were used as initial estimates for this analysis. The new augmented matrix [D_1–D_{21}] was modelled analogously to run 1. In this analysis, the hard-modelling constraint was imposed to the kinetic profiles of D_1 and D_2 standard experiments and the rest of urine sample experiments were soft-modelled, without taking into account the information known a priori about the species present in the experiments with urine because these experiments are used as unknown samples. So, the isp matrix used for this run was



The two first rows correspond to the standards; whereas the rest of rows are 1' vectors and correspond to the experiments containing urine.

The resolved concentration profiles and spectra are consistent with the chemistry of the samples and the known reference spectra (Fig. 4). The concentrations of the different analytes were obtained for each of the samples. These values were compared with the reference values and the figures of merit of the correlation line were calculated (Table 3, Fig. 5a). SEP and RMSEP obtained values are low showing that the predictive capability of this method is good. This idea is corroborated by bias values close to cero. It has to be pointed out that only two standards in aqueous solutions have been used to obtain this quantitative information. It is also remarkable that the reconstruction of the original data set is also good, with very low lack of fit values (Table 2).

From the results obtained, it can be observed (Fig. 5a) that the contribution of xanthine and hypoxanthine in some samples is extremely small. This may suggest that these analytes are not present in the sample. A last series of calculations was performed

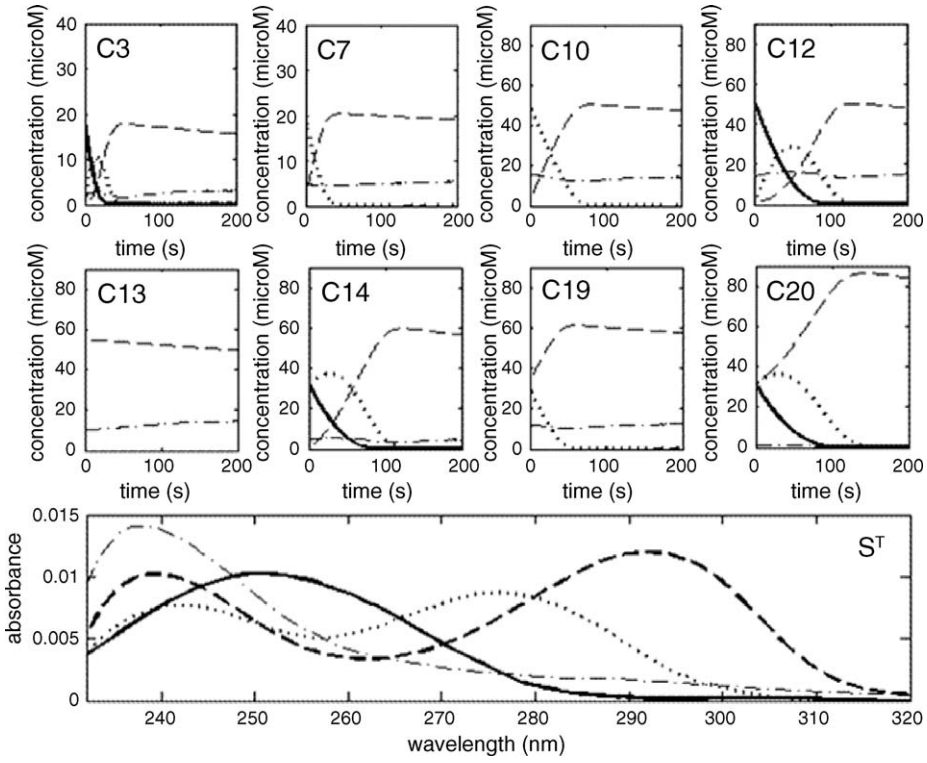


Fig. 4. Concentration and spectral profiles obtained from run 2 (hypoxanthine, solid; xanthine, dotted; uric acid, medium dash; urine, dotted-dash). Only the concentration profiles of some matrices are shown.

(run 3) using the same set of experimental data in run 2 and the same constraints, but modifying the isp matrix to take into account the information about the absence of some analytes in certain samples. Now, the rows corresponding to D_6–D_{10} experiments had a "0" in the H column and the row of the experiment D_{13} had a "0" in the H and X columns. The similar concentrations obtained for the absorbing species for each unknown sample, the similar shape in the resolved spectra and concentration profiles and the acceptable lack of fit in the analysis confirm the validity of the new information about the absence of species included in the run 3 analysis. As a consequence of the quantitative information incorporated in the analysis, RMSEP and SEP values, as well as the figures of merit for the correlation lines (Fig. 5b), for the three analytes are now slightly better than those obtained in run 2, where samples were treated as unknowns. The differences between the runs 2 and 3 results, though, are small, and the results would have been acceptable if only the absence-presence information about the species in the standards had been used (Table 3).

Table 4 shows that the results obtained by HS-MCR (Tables 2 and 3) are comparable in quality to others obtained by separation techniques, such as GC [25] and other analytical methodologies [26] and by other chemometric methods like unfold-PLS and N-PLS [1] (Table 4). Advantages over separation techniques are that no clean-up and removal of interferences are needed for the analysis. Compared with multivariate calibration methodologies, the number of standards required is much smaller and there is no need to incorporate the information on the

Table 3
Comparison between figures of merit obtained for different ways of applying HS-MCR in the quantification of H, X and U in urine: statistical parameters

Run no.	Slope	Offset	Correlation	RMSEP[b]	SEP	Bias
Hypoxanthine						
Run 2	1.02 (0.02)[a]	−1.7 (0.7)	0.986	2.393	2.057	1.311
Run 3	1.01 (0.02)	−0.4 (0.5)	0.993	1.488	1.527	−0.059
Xanthine						
Run 2	1.03 (0.02)	−2.1 (0.5)	0.994	1.964	1.465	1.350
Run 3	0.96 (0.01)	0.9 (0.3)	0.997	1.143	1.120	0.341
Uric acid						
Run 2	0.97 (0.02)	−0.8 (0.3)	0.995	1.666	1.226	1.163
Run 3	1.02 (0.01)	0.5 (0.3)	0.997	1.412	1.085	0.937

[a] Standard deviation in parentheses.
[b] RMSEP, SEP and bias expressed as μmol L^{-1}.

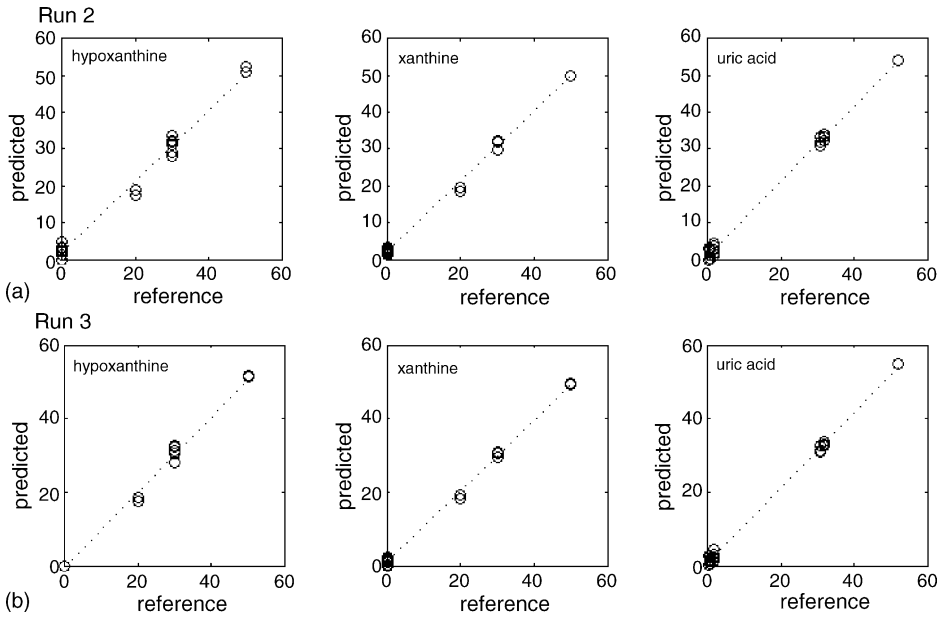


Fig. 5. Correlation plots of predicted concentrations by HS-MCR in (a) run 2 and (b) run 3 versus reference concentrations of hypoxanthine, xanthine and uric acid. Concentration units are expressed in $\mu mol\,L^{-1}$.

Table 4
Comparison of the results with other analysis of hypoxanthine and xanthine performed by different authors

Reference	Sample	Method[a]	Analyte	Concentration range[b] (μmol/L)	r^2
Putterman et al. [25]	Human urine	HPLC and GC	Hypoxanthine	0–1300	0.95
			Xanthine	0–2000	0.98
Zen et al. [26]	Spiked buffered solutions	NSPEs	Xanthine	2–40	–
Amigo et al. [1]	Spiked human urine	3W-PLS2	Hypoxanthine	0–3000	0.998
			Xanthine	0–1500	0.982
New method	Spiked human urine	HS-MCR	Hypoxanthine	0–5000	0.993
			Xanthine	0–5000	0.997
			Uric acid	0–3000	0.997

[a] HPLC, high performance liquid cromatography; GC, gas chromatography; NSPEs, preanodized nontroite-coated screen-printed carbon electrodes; 3W-PLS2, three-way partial least squares 2.
[b] All concentrations are related to volume of urine.

matrix interferences (neither the identity nor the concentration range) in the standard solutions.

5. Conclusions

The application of mixed hard- and soft-modelling resolution methodologies (HS-MCR) has been proven to be a useful tool for a complete quantitative and kinetic description of samples of human urine containing hypoxanthine, xanthine and uric acid. This example shows that HS-MCR can work for quantitative purposes in any multicomponent system, where the analytes evolve according to a known enzymatic or kinetic process in the presence of absorbing interferences.

In these systems, HS-MCR is applied to a column-wise data set, formed by the enzymatic processes carried out in the aqueous standards and in the unknown samples. The hard-modelling constraint is applied to the concentration profiles of the aqueous standards and forces the species involved to obey the enzymatic model. This constraint decreases dramatically the rotational ambiguity (good recovery of profiles) of the resolution results and incorporates the absolute quantitative information of the standards as parameters needed in the enzymatic fitting, i.e., as the concentrations of the species at time zero. The quantitative information of each of the unknown samples, the concentration profiles of which are resolved using only soft-modelling constraints, can be found in the first row of the related C_i submatrix. This row contains the concentration of the analytes at time zero, before the enzymatic process has taken place.

As a method for quantitative analysis, the experimental approach is simple and fast and there is no sample pre-treatment needed to suppress interferences. The HS-MCR algorithm shows also clear advantages over other multivariate calibration methods, such as the small amount of standards needed (one or two standard samples are sufficient) and the fact that they can be prepared in aqueous solution, with no need to know or include the interferences present in the samples. Furthermore, the kinetic profiles and the spectra of the compounds involved are obtained. This amount of information obtained with HS-

MCR methodology clearly overcomes the information provided by pure soft-modelling and pure hard-modelling procedures.

Acknowledgements

José Manuel Amigo, Jordi Coello and Santiago Maspoch want to thank to the Spanish Programme of Chemical Process Technologies (Projects CTQ 2004–2013) for the finantial support.

References

[1] J.M. Amigo, J. Coello, S. Maspoch, Anal. Bioanal. Chem. 382 (2005) 1380.
[2] J. Coello, S. Maspoch, N. Villegas, Talanta 53 (3) (2003) 627.
[3] J. Diewok, A. de Juan, M. Maeder, R. Tauler, B. Lendl, Anal. Chem. 75 (2003) 641.
[4] E.V. Thomas, Anal. Chem. 66 (1994) 795.
[5] S.R. Crouch, J. Coello, S. Maspoch, M. Porcel, Anal. Chim. Acta 424 (2000) 115.
[6] C. Foppoli, R. Coccia, C. Cini, M.A. Rosei, Biochim. Biophys. Acta 1334 (1997) 200.
[7] I.S. Park, N. Kim, Anal. Chim. Acta 394 (1999) 201.
[8] H. Okuma, E. Watanabe, Biosens. Bioelectron. 17 (2002) 367.
[9] M.A. Carsol, M. Mascini, Talanta 47 (1998) 335.
[10] M. Stredansky, A. Pizzariello, S. Miertus, J. Svorc, Anal. Biochem. 285 (2000) 225.
[11] M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, M. Porcel, Anal. Chim. Acta 398 (1999) 83.
[12] A. de Juan, M. Maeder, M. Martínez, R. Tauler, Chemom. Intell. Lab. Syst. 54 (2000) 123.
[13] R. Tauler, Chemom. Intell. Lab. Syst. 30 (1995) 133.
[14] R. Tauler, A.K. Smilde, B.R. Kowalski, J. Chemom. 9 (1995) 31.
[15] A. de Juan, R. Tauler, Anal. Chim. Acta 500 (2003) 195.
[16] J. Jaumot, R. Gargallo, A. de Juan, R. Tauler, Chemom. Intell. Lab. Syst. 76 (2005) 101.
[17] R. Manne, Chemom. Intell. Lab. Syst. 27 (1995) 89.
[18] M.S. Díaz-Cruz, J.M. Díaz-Cruz, J. Mendieta, R. Tauler, M. Esteban, Anal. Biochem. 277 (2000) 189.
[19] J.M. Díaz-Cruz, J. Agulló, M.S. Díaz-Cruz, C. Ariño, M. Esteban, R. Tauler, Analyst 126 (2001) 371.
[20] M.C. Antunes, J.E. Simão, A.C. Duarte, R. Tauler, Analyst 127 (2002) 809.
[21] A. de Juan, M. Maeder, M. Martínez, R. Tauler, Anal. Chim. Acta 442 (2001) 337.
[22] J.M. Amigo, A. de Juan, J. Coello, S. Maspoch, A mixed hard- and soft-modelling approach to study and monitor enzymatic systems in biological fluids, Anal. Chim. Acta 567 (2006) 236.
[23] R. Tauler, D. Barceló, Trends Anal. Chem. 12 (1993) 319.
[24] M. Maeder, Anal. Chem. 59 (1987) 527.
[25] G.J. Putterman, B. Shaikh, M.R. Hallmark, C.G. Sawyer, C.V. Hixson, F. Perini, Anal. Biochem. 98 (1979) 18.
[26] J.-M. Zen, Y.-Y. Lai, H.-H. Yang, A. Senthil Kumar, Sens. Actuators B 84 (2002) 237.

An introduction to Multivariate Curve Resolution-Alternating Least Squares: Spectrophotometric study of the acid-base equilibria of 8-Hydroxyquinoline-5-sulfonic acid.

Cristina Rodríguez-Rodríguez, José Manuel Amigo, Jordi Coello, Santiago Maspoch.

J. Chem. Edu 2007

Anexos / Annexes

MS # 2005-0543

An Introduction to Multivariate Curve Resolution-Alternating Least Squares: Spectrophotometric Study of the Acid–Base Equilibria of 8-Hydroxyquinoline-5-sulfonic Acid

Abstract

We describe an experiment in which the UV–vis spectrum of 8-hydroxyquinoline-5-sulfonic acid (8HQS) is measured as a function of pH. This experiment introduces students to the chemometric treatments of evolving signals by means of multivariate curve resolution-alternating least squares (MCR-ALS). This is a well established decomposition method to obtain the concentration and spectral profiles of the absorbing components of an evolving system without using any previous assumptions about the chemical model. The student may estimate, by an additional simple graphical treatment, the acid–base constants of 8HQS. Since the spectra of the pure components can be obtained experimentally and the pK_a values are well established in the literature, the student will be able to compare his or her results with the reference values and draw conclusions about the potential value of applying MCR-ALS to unknown systems.

Keywords List

Acids / Bases; Analytical Chemistry; Aqueous Solution Chemistry; Calculator-Based Learning; Chemoinformatics; Chemometrics; Computer-Based Learning; Equilibrium; Graduate Education / Research; Hands-On Learning / Manipulatives; Laboratory Instruction; pH; Upper-Division Undergraduate; UV-Vis Spectroscopy

An Introduction to Multivariate Curve Resolution-Alternating Least Squares: Spectrophotometric Study of the Acid–Base Equilibria of 8-Hydroxyquinoline-5-sulfonic Acid W

Cristina Rodríguez-Rodríguez, José Manuel Amigo, Jordi Coello, and Santiago Maspoch*
Departament de Química, Universitat Autònoma de Barcelona, 08193-Bellaterra, Spain; **santiago.maspoch@uab.es*

Data treatment in analytical chemistry has undergone rapid evolution in the last 15 years. These new advances are being used in academic research laboratories and in industrial laboratories, so the teaching of chemometrics has become common in most university curricula. Chemistry students, particularly those majoring in analytical chemistry, should acquire basic knowledge of experimental design, principal components analysis, multivariate calibration, and pattern recognition *(1–3).* Recently, an introduction to curve resolution has also been proposed as a part of a chemometrics module for an undergraduate course *(4).*

Monitoring a process in an industrial reactor is not so different from monitoring any process where, from fixed initial conditions, the system evolves as a function of a continuous variable (time, pH, temperature, etc.). The development of chemometric methodologies for the treatment of these kinds of evolving signals is one of the key points in current chemometrics research *(5, 6).* Most of the proposed methodologies derive from the multivariate curve resolution-alternating least squares algorithm (MCR-ALS) *(7).* This is a well-established decomposition method to obtain the concentration and spectral profile of all substances present in the mixture, assuming that all of them absorb in the recorded spectral range. The possibility of obtaining concentrations and spectra account for the interest in this methodology in the field of industrial processes and in basic research.

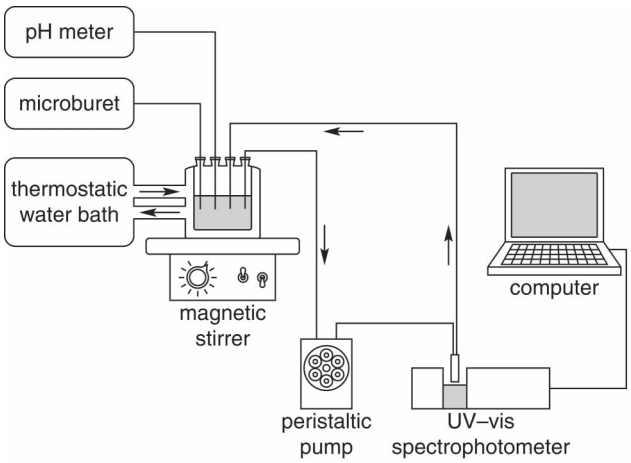


Figure 1. Schematic of the spectrophotometric titration.

In this context, the monitoring of the UV–vis spectral evolution of 8-hydroxyquinoline-5-sulfonic acid (8HQS) with pH is proposed with the aim of illustrating and introducing the MCR-ALS methodology to undergraduate students. 8HQS has been chosen because it has well-defined bands in UV–vis spectra, the acid–base constants (protonation of quinolic nitrogen and deprotonation of the phenolic group) are well-established in the literature *(8, 9),* and they are spaced out enough to obtain experimentally the pure spectrum of the neutral component (the intermediate of the process) and compare it with the spectral pattern obtained by means of MCR-ALS. In addition, this experiment introduces the students to the MatLab code for chemometric purposes *(10, 11).*

Experiment Procedure

The experiment may be performed in two sessions. In the first session the students titrate and record the spectra. Estimated time is 3–4 hours. In the second session the work is focused on the chemometric treatment. With some previous knowledge of Matlab and chemometrics, this session may take 3 hours.

Acquisition of the Spectra

A constant 0.50 M ionic strength ($NaClO_4$) was kept throughout the experiment. To avoid the precipitation of $KClO_4$ in the liquid junction porous plug of the pH electrode, the inner KCl solution of the electrode was replaced by 3 M $NaClO_4$ solution saturated with AgCl. Prior to the spectrophotometric titration, the pH electrode was calibrated following the methodology proposed by Gran *(12, 13).*

A block diagram of the experimental setup is shown in Figure 1. The acid–base titration was carried out in a 50 mL double-walled vessel kept at 25.0 ± 0.2 °C with continuous stirring. The following solutions were placed in the vessel: 10 mL of 0.023 M perchloric acid, 10 mL of 1.5 M sodium perchlorate, 10 mL of 0.02 M tris(hydroxymethyl)-aminomethane buffer at pH 7.4, and 10 mL of 4.0×10^{-4} M of 8HQS solution. The buffer solution was placed in the solution to avoid a sharp shape of pH and to allow a more gradual spectra change. Micro-additions of 0.5 M NaOH solution were made by using a microburet. The pH range was from 2 to 11. A combined pH electrode from Crison and a Knick pH meter were used. Absorbance was measured every 2 nm in the range of 250 to 500 nm by using an HP-8452A UV–vis diode array spectrophotometer.

The solution was continuously propelled through a flow quartz cuvette of 10 mm path length by a peristaltic pump at a flow rate of 2 mL min^{-1}. Tygon pumping tubes were used; all other tubing was 0.5 mm i.d. Teflon.

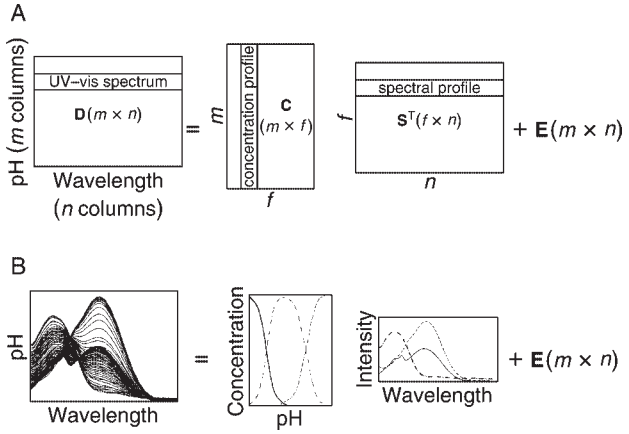


Figure 2. Schematic decomposition of MCR-ALS: (A) mathematically and (B) visualizing the obtained spectral matrix decomposition.

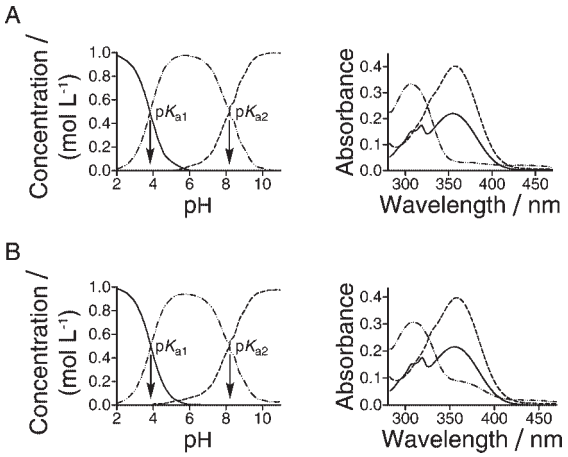


Figure 3. Concentration and Spectral profiles obtained by MCR-ALS: (A) S1 initial estimates and (B) S3 initial estimates. The arrows show the qualitative equilibrium constants obtained.

Table 1. Results Obtained by MCR-ALS and Literature

Initial Estimations	pH Values[a]			pK_{a1}	pK_{a2}	Correlation[b]
S1	1.95	5.98	10.96	3.80	8.22	0.9998
S2	2.54	6.42	9.04	3.79	8.22	0.9983
S3	1.95	3.10	10.96	3.79	8.22	0.9917
S4	1.95	8.61	10.96	3.78	8.22	0.9996
Beltrán et al. *(9)*				3.80	8.35	
Martell et al. *(8)*				3.80	8.20	

[a] pH values of the spectra used as initial estimates. [b] Correlation coeficient corresponding to the intermediate of the reaction.

Chemometric Treatment of the Recorded Spectra

Recorded spectra were converted into ASCII files and arranged in a **D** matrix ($m \times n$) for further analysis by means of MCR-ALS software supported with MatLab v. 6.0 (MathWorks, MA). In the **D** matrix spectra are sorted in rows as a function of pH in a unique MatLab file (Figure 2A). The m rows correspond to the spectra obtained at m different pHs and the n columns refer to the n different wavelengths. The absorbance at each wavelength is assumed to be the sum of the absorbances of each component of the mixture. MCR-ALS decomposes iteratively the matrix **D** yielding concentration profiles **C** ($m \times f$) and the spectral profiles $\mathbf{S}^T$ ($f \times n$) for the previously chosen number of components f (Figure 2B).

As the absorbance is the product of two unknown variables, some constraints (non-negativity, unimodality, and concentration closure) have to be applied to reach to an unambiguous solution. In addition, initial estimates of the concentration or spectral profiles are needed to start the algorithm to iterate. These aspects are discussed in the Supplemental Material.[W]

Software for the MCR-ALS algorithm is available from different sources *(14, 15),* all of them working under MatLab code. Easy-to-follow instructions for new users can be found in the literature *(14–17).* The software provided by Tauler et al. was used in this experiment *(14, 17).*

Hazards

8-Hydroxyquinoline-5-sulfonic acid and tris(hydroxymethyl)aminomethane are irritants. Sodium hydroxide and hydrochloric acid are corrosives. Sodium perchlorate and perchloric acid are oxidizing and corrosives.

Results and Discussion

MCR-ALS produces an estimation of the spectral and concentration profiles of all absorbing species involved in the monitorized process (Figure 3). The reliability of these results has to be proven to convince the student of the potentiality of this technique when applied to new (unknown) processes. Apart from the simple visual comparison, the fit between computed and true experimental spectra can be measured by the correlation coefficient between both spectra (Table 1). Also the quality of the estimation of the variation of concentration with pH can be measured by calculating the pK_a values by any graphical procedure, for example, the maximum or minimum of the corresponding first derivative of the concentration profiles. Comparison of these calculated pK_a values with those obtained in the literature is a valid way determining whether the procedure is working as desired (Table 1).

Experimental spectra have been used as starting points of the decomposition. Because MCR-ALS results may be affected by this selection, different approaches of the initial spectra (S1–S4) have been tested (Table 1). S1 corresponds to the "logical" selection from a visual study of the whole set of recorded spectra. pH values denote which spectra were selected. S1 is also a logical hypothesis for a system with three evolving species: the first and last spectrum (best estimations

of the first and third species) and one spectrum from the middle for the intermediate compound. S2–S4 shows the effect of shifting some of the "best" conditions. These are only four of the multiple choices that can be tested. As can be seen, good correlation coefficients with similar pK_a values were obtained in all cases. The MCR-ALS results for the best and worst (S1 and S3) correlations are shown in Figure 3.

Conclusion

MCR-ALS is a chemometric tool that may extract a large quantity of information about a chemical system without many previous assumptions. In this case, the student only knows initially that the system is a pH-modulated equilibrium. He or she has to discover that the system is composed by three absorbing components, has to choose the initial estimates, and has to impose adequate constraints. The student will also discover the concentration and spectral profiles of the components and will estimate by a simple graphical treatment the acid–base constants of the system. Computed results will be compared with true experimental values.

It is clear that the more chemical knowledge we have about a system, the easier and more reliable our interpretation will be. However, the student will learn that it is possible to obtain good information by applying the MCR-ALS methodology even with little knowledge about an evolving system.

Acknowledgments

José Manuel Amigo, Jordi Coello, and Santiago Maspoch gratefully acknowledge the financial support of this work by the Ministerio de Ciencia y Tecnología (DGI BQU2001-2019). Cristina Rodriguez-Rodriguez acknowledges the financial support from the Ministerio de Ciencia y Tecnología (EET2002-05157-C05-03).

[w]Supplemental Material

Detailed students procedures, instructors' notes, a complete demo, equipment and reagent list, detailed procedures for preparing reagents, and hazards alerts are available in this issue of *JCE Online*.

Literature Cited

1. Kellner, R.; Mermet, J.-M.; Otto, M.; Varcárcel, M.; Widmer, H. M. *Analytical Chemistry;* Wiley-VCH: Weinheim, 2004.
2. Howery, Darryl G.; Hirsch, Roland F. *J. Chem. Educ.* **1983,** *60,* 656.
3. Delaney, Michael F.; Warren, F. Vicent, Jr. *J. Chem. Educ.* **1981,** *58,* 646.
4. Msimanga, Huggins Z.; Elkins, Phet; Tata, Segmia K.; Smith, Dustin Ryan. *J. Chem. Educ.* **2005,** *82,* 415.
5. Tauler, R. *Chemom. Intell. Lab. Syst.* **1995,** *30,* 133–146.
6. Amigo, J. M.; Coello, J.; Maspoch, S. *Anal. Bioanal. Chem.* **2005,** *382,* 1380–1388.
7. Tauler, R.; Smilde, A. K.; Kowalski, B. J. *J. Chemom.* **1995,** *9,* 31–58.
8. Smith, R. M.; Martell, A. E. *Critical Stability Constants;* Plenum: New York, 1975; Vol. 2.
9. Beltrán, J. L.; Codony, R.; Prat, M. D. *Anal. Chim. Acta* **1993,** *276,* 441–454.
10. O'Haver, T. C. *Chemom. Intell. Lab. Sys.* **1989,** *6,* 95–103.
11. Chau, F. T.; Chung, W. H. *J. Chem. Educ.* **1995,** *72,* 78–85.
12. Gran, G. *Analyst* **1952**, *77,* 661–671.
13. Boiani, J. A. *J. Chem. Educ.* **1986**, *63,* 724–726.
14. Multivariate Curve Resolution Home Page. *http://www.ub.es/gesq/mcr.htm* (accessed Mar 2007).
15. Paul Gemperline Home Page. *http://personal.ecu.edu/gemperlinep/* (accessed Mar 2007).
16. Gemperline, P. J.; Cash, E. *Anal. Chem.* **2003,** *75,* 4236–4243.
17. Jaumot, J.; Gargallo, R.; de Juan, A.; Tauler, R. *Chemom. Intell. Lab. Syst.* **2005,** *76,* 101–110.

Lab Documentation

PERFORM THE EXPERIMENT

Session 1: pH electrode calibration and 8HQS spectrophotometric titration

Apparatus and software

Double-walled vessel

Thermostatized bath: *Thermo Haake DC10*

Microburette: *Hamilton*

Peristaltic pump: *Ismatec Reglo*

Spectrophotometer: HP-8452A UV-Vis Diode Array

Flow quartz cell

pH-electrode: combined-pH electrode from *Crison*

pH-meter: *Knick*

Magnetic stirrer

Pumping Tygon and Teflon tubes

MatLab v6.0

Procedure:

1) Replace the inner KCl solution of the pH electrode by a 3 M $NaClO_4$ solution saturated with AgCl.

2) Prepare the following solutions in bi-distilled water:

0.023 M perchloric acid

1.5 M sodium perchlorate

0.02 M Tris(hydroxymethyl)-aminomethane buffer at pH 7.4 (pH adjustment made with the perchloric acid solution)

0.5 M sodium hydroxide

3) Prepare the stock solution of 4.0×10^{-4} M of 8-hydroxyquinoline-5-sulfonic acid (8HQS) in $NaClO_4$ 0.5M.

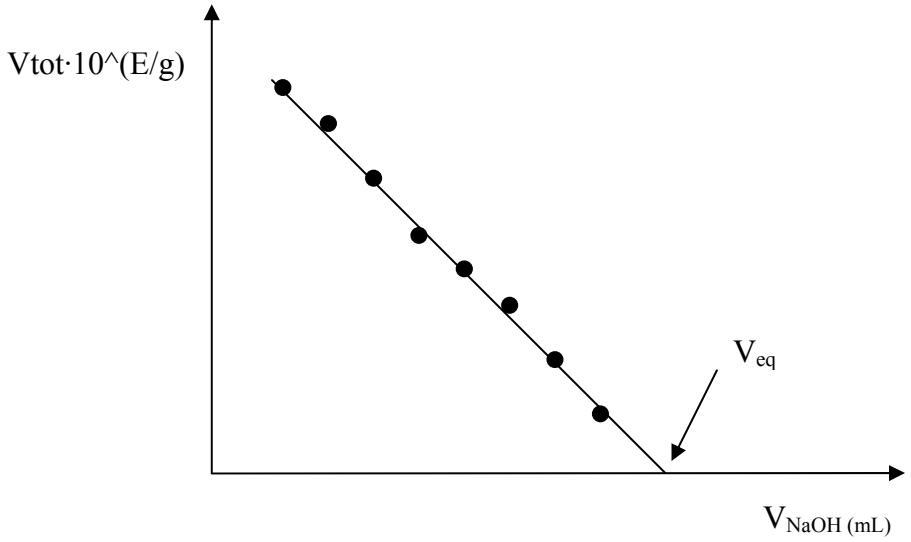
4) Place 2 ml of $HClO_4$ 0.023M (V_{HClO4}) and 10ml of 0.6M NaClO4 (V_{NaClO4}) in the titration vessel.

5) Switch on the water bath and keep the temperature at 25 ºC

6) Titrate with NaOH 0.05M (C_{NaOH}) and record the values of electrode potential E (mV) after each addition (the cumulative volume added is V_{NaOH}). Keep a constant stirring.

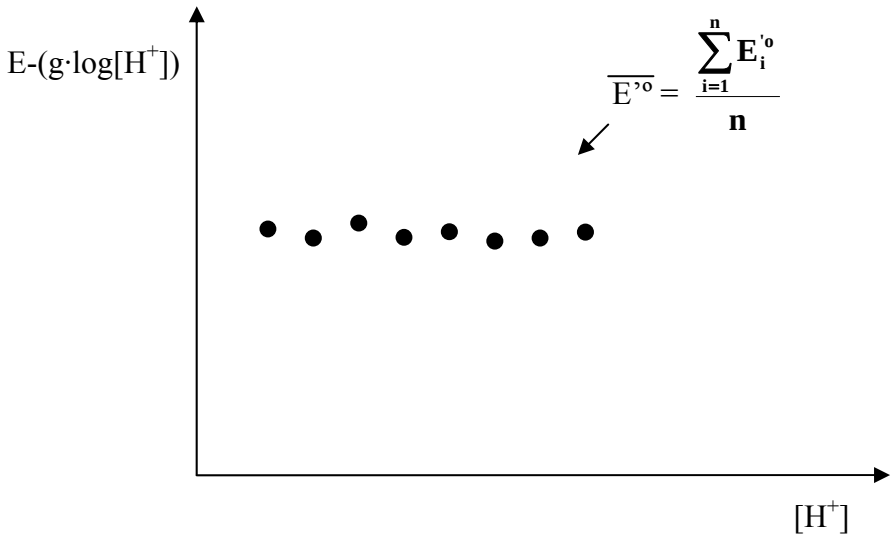
7) Plot Vtot·10^(E/g) versus V_{NaOH}.

Consider g (25 ºC) = 59.16 and Vtot = $V_{HClO4} + V_{NaClO4} + V_{NaOH}$



Calculate the equivalence volume of NaOH (Veq) .

8) The conditional standard potential (E'º) can be obtained by plotting E-g·log[H] versus $[H^+]$, being $[H^+] = ((Veq-V_{NaOH})\cdot C_{NaOH})/Vtot$



9) Add 10 mL of 0.023 M perchloric acid, 10 mL of 1.5 M sodium perchlorate and 10 mL of 0.02 M Tris(hydroxymethyl)-aminomethane.

10) Switch on the peristaltic pump.

11) Record the blank spectrum.

12) Place into the vessel 10 ml of 8HQS stock solution and leave to equilibrate the solution for 5 minutes.

13) Titrate with 0.5M NaOH. After each micro-addition leave to equilibrate for 1 minute. Record the spectrum and write down the E value. Transform the E value into pH value according to this equation:

$$\mathbf{pH} = \frac{\mathbf{(\overline{E'^{o}} - E)}}{\mathbf{g}} \qquad (1)$$

Hazard Notes

8-hydroxyquinoline-5-sulfonic acid monohydrated (Chemical Abstract Service No. 84-88-8). Harmful. Risk notes: 22-36; Safety notes: 26-36

Tris(hydroxymethyl)-aminomethane (CAS No. 77-86-1) Irritant. Risk notes: 36/37/38; Safety notes: 26-36

Sodium hydroxide (CAS No. 1310-73-2) Corrosive. Risk notes: 34; Safety notes: 26-37/39-45

Sodium perchlorate (CAS No. 7601-89-0) Oxidizing and harmful. Risk notes: 9-22; Safety notes: 13-22-27

Perchloric acid (CAS No. 7601-90-3) Oxidizing and corrosive. Risk notes: 5-8-35; Safety notes: 23-26-36-45

Chlorhydric Acid (CAS No. 7647-01-0) Toxic and corrosive. Risk notes: 23-35; Safety notes: 9-26-36/37/39-45

Session 2: MCR-ALS computation

1) Import the data matrices

Different spectrophotometers may store the spectra in different formats. Most of them as ASCII files that can be opened by Excel and stored in a sheet as *.xls. or *.csv. Both formats can be imported from MatLab software:

Open MatLab

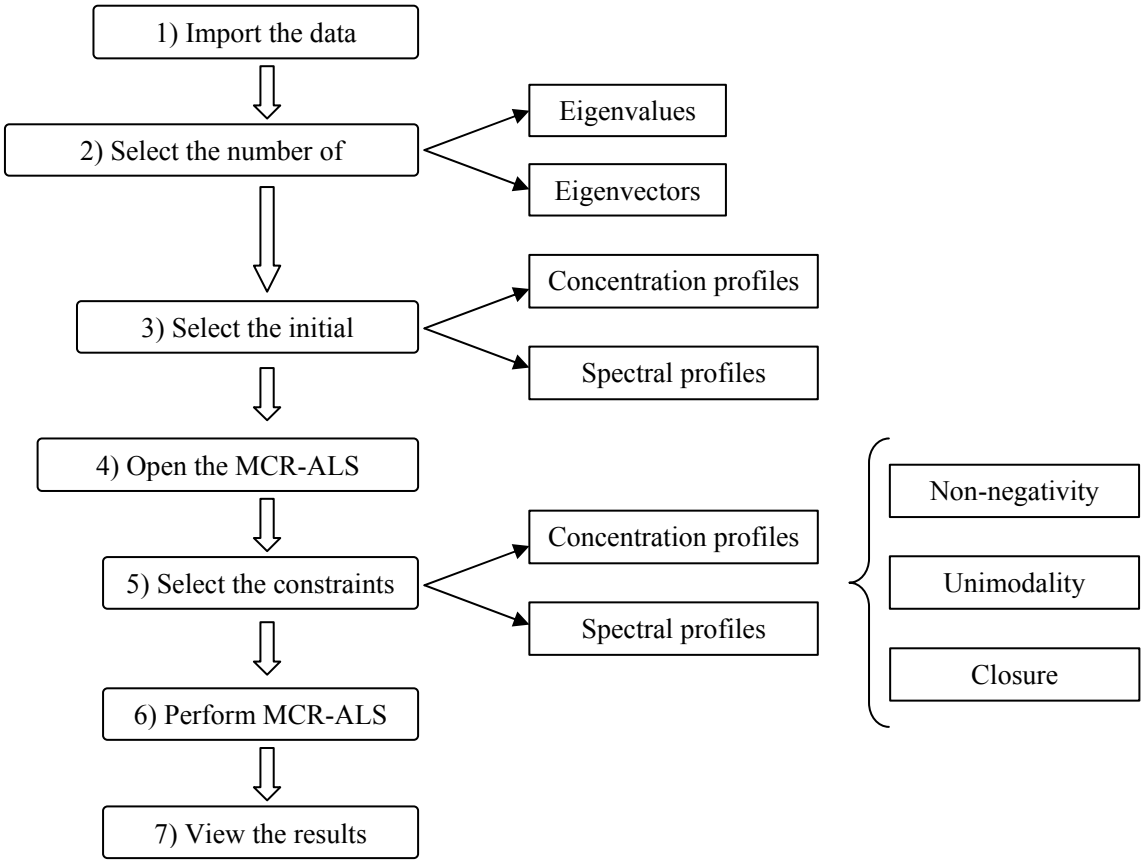
Select **File** ⟶ **Import Data** ⟶ *.xls or *.csv

This generates the **D (m × n)** matrix.

Vectors containing the wavelength and pH values can be easily created in MatLab workspace. You can copy this vector from Excel to MatLab and rename properly.

2) Run MCR-ALS algorithm

Main steps to apply MCR-ALS algorithms are shown in the next flow-chart. For a more detailed explanation look at the **demonstration** at the end of the Lab Documentation section.



INSTRUCTOR NOTES

Session 1: pH electrode calibration and 8HQS spectrophotometric titration

All reagents were used without further purification with the exception of sodium perchlorate which was recrystallized to eliminate the presence of trace metals following a well known procedure (1).

The ionic strength was kept at 0.5 M ($NaClO_4$). To avoid the precipitation of $KClO_4$ in the liquid junction porous plug of the pH electrode, the inner KCl solution was replaced by a 3M $NaClO_4$ solution saturated with AgCl. Commercial standard buffers solutions at this ionic strength are not usual. Consequently, pH electrode was calibrated by following the methodology proposed by *Gran* (2). This part of the experiment can be easily avoided by using a lower ionic strength (i.e the classical 0.1 M KCl) and calibrating the electrode by the usual standard buffer solutions.

Tris solution was placed in the solution to avoid a sharp shape of pH and to allow a more gradual spectra change.

Session 2: MCR-ALS computation

Spectra are recorded in a wide wavelength range (200-500 nm.). However, the use of a very high or low wavelength range could introduce non-linearity or noise to the system. So, only the wavelength range containing useful analytical information has to be used (in our case, 280-420 nm).

Software for MCR-ALS algorithms are freely available from different sources (3, 4), all of them working under MatLab code. In these references and in different published papers (5, 6) easy-to-follow instructions for new users can be also found

The software provided by Gemperline et al (4) and by Tauler et al (3) are not exactly equivalents. In Gemperline's the SVD and EFA algorithms are already implemented, which are not in Tauler's. However, we recommend the later because initial estimates of spectral profiles can be used in an easier way. The software by *Gemperline et al.* is very useful when initial estimates of concentration profiles are used (EFA analysis), but we think that this approach is less intuitive for a student. Both software have a graphical

interface and contain a very useful manual that should be consulted by the instructor before starting the experiment.

OUTLINES OF MCR-ALS THEORY

Just a brief description of MCR-ALS is presented. More information about the algorithm can be found in references (7, 8).

Recorded spectra are arranged into a **D** matrix (*m*,*n*), where spectra are sorted in rows as a function of pH. The *m* rows correspond to the spectra obtained at *m* different pH and the *n* columns refer to the *n* different wavelengths. If the absorbance at each wavelength is assumed to be the sum of the absorbances of each component of the mixture,

$$\begin{aligned} A_{\lambda 1} &= a_{11}\cdot c_1 + a_{12}\cdot c_2 + \ldots\ldots\ldots\ldots + a_{1n}\cdot c_n + E_{\lambda 1} \\ A_{\lambda 2} &= a_{21}\cdot c_1 + a_{22}\cdot c_2 + \ldots\ldots\ldots\ldots + a_{2n}\cdot c_n + E_{\lambda 2} \\ &\vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ A_{\lambda m} &= a_{m1}\cdot c_1 + a_{m2}\cdot c_2 + \ldots\ldots\ldots\ldots + a_{mn}\cdot c_n + E_{\lambda m} \end{aligned} \qquad (2)$$

MCR-ALS decomposes the **D** matrix in the **C** ($m \times f$) and $\mathbf{S}^{\mathbf{T}}$ ($n \times f$), which contains the concentration and spectral profiles respectively. f is the number of absorbing chemical components. **E** (m x n) is the matrix of residuals (Equation 3).

$$D = CS^T + E \qquad (3)$$

The value of f has to be determinated before starting the MCR-ALS algorithm. In our system, we know beforehand that there are two protonation equilibria corresponding to three absorbing components, but many times this information is not available for real systems where the number of intermediates is not known. Several procedures have been described to solve this question (3, 9-11). Most of them are based on the principle that there are as many components as linearly independent elements in the **D** matrix (rank of the matrix). A very useful method is to get the eigenvectors and eigenvalues of **D** matrix by performing a Singular Value Decomposition (SVD). The chemical rank can be expressed as the number of eigenvalues higher than eigenvalues associated to the noise level. Also the shape of the eigenvector can be useful to estimate the correct number of absorbing components.

The great advantage of MCR-ALS is that the system can be solved without (almost) any prior assumption about the chemical system.

Obviously, the mathematical resolution of the product of two unknown variables (C and S^T) has not a unique solution. To avoid this ambiguity some constraints have to be applied to the concentration and/or to the spectral profiles. The most common constraints are:

Nonnegativity: Concentration profiles of the components are supposed to be always positives. This constraint can be extrapolated to spectral profiles when UV-Vis spectrophotometry is used because the absorbance has positive values.
Unimodality: When it can be assumed that the shape of the spectral and/or concentration profiles only have one maximum. It is usually applied for the concentration profiles.
Closure: When the total concentration of the components in the system is constant throughout the experiment.

Another important constraint that can help us to solve the ambiguity of the system is the *selectivity*. This constraint can be used in a certain area of the experiment where it is sure that the change in the absorbance is due only to one specific component. There are processes or reactions where it is possible to know the pure spectrum of some compounds (for instance, the spectra of the reagents at the beginning of the reaction). This information can be also used to minimize the ambiguity associated to the decomposition.

MCR-ALS requires initial estimates of the spectral or the concentration profiles for each component. Working with good initial estimates can help the algorithm to converge to a good solution avoiding local minima. There are a lot of methodologies to obtain initial estimates in an easy way. The more straightforward way to obtain initial estimates of the spectral profiles is to choose some of the spectra from the **D** data matrix. To obtain estimates about the concentration profiles, the most useful methodology is to perform an Evolving Factor Analysis (EFA). A good explanation of EFA may be found in reference (11).

The ALS optimization stops when the relative difference in lack of fit (%LOF) values (Equation 4) between consecutive iterations is below than a threshold value. LOF value is defined as the difference between the input data **D** (*m* x *n*) and the data reproduced from the $\mathbf{CS}^{\mathbf{T}}$ product (9).

$$\% \text{ LOF} = 100 \times \sqrt{\frac{\sum_{i}^{I}\sum_{j}^{J} e_{mn}^{\ 2}}{\sum_{i}^{I}\sum_{j}^{J} d_{mn}^{\ 2}}} \qquad (4)$$

where e is each *mn*th element of the residual matrix **E**, and d is each *mn*th element of the **D** matrix. The threshold value of LOF should be selected according to the noise in the spectra. A very used criterion is to stop the iterations when the difference between two consecutive LOF values is below 0.1%.

References

1. *"Some Laboratory Methods"*, Royal Institute of Technology, Sweden, **1959**.

2. Boiani, J.A J.Chem Educ., **1986**, 63(8), 724-726.

3. http://www.ub.es/gesq/mcr.htm

4. http://personal.ecu.edu/gemperlinep/

5. Jaunot, J.; Gargallo, R.; de Juan, A.; Tauler, R. Chemom. Intell. Lab. Syst. 2005, 76, 101-110.

6. Gemperline, P.J.; Cash, E. Anal. Chem. 2003, 75, 4236-4243.

7. Tauler, R.; Smilde, A.K.; Kowalski, B.J. J. Chemom. **1995**, 9, 31-58

8. Tauler, R. Chemom. Intell. Lab. Syst. **1995**, 30, 133-146.

9. Jaumot, J.; Gargallo, R.; de Juan, A.; Tauler, R. Chemom. Intell. Lab. Syst. **2005**, 76, 101-110

10. Malinowsky, E.R. and Howery, D.G. *Factor Analysis in Chemistry*, Wiley, New York, **1980**.

11. Maeder, M. Anal. Chem. **1987**, 59, 527-530.

DEMONSTRATION:

Attached in JCE additional material you will find *8hqs.mat* file.

Open **MatLab**

Open *8hqs.mat* file in the current directory

Double click on *8hqs.mat*

Seven matrices will appear in the workspace:

D (63 × 70): Data matrix that contains spectra recorded at 63 pH values in 70 wavelengths.

ph (63 × 1): A vector that contains the pH values obtained for each recorded spectra. This is an important vector to visualize correctly the concentration profiles due to the fact that pH values are not equidistant (1.95-10.56).

w (70 × 1): A vector that contains the wavelength range used to monitor the spectra at each pH value (282-420 nm every 2 nm).

S1 to S4 (70 × 3): Matrices that contain spectra recorded at different pH values (see table 1 of Lab Summary section). These matrices will be used as initial estimates of the spectral profiles of the components.

You can visualise the data matrix by typing in the command window:

Type *plot(w,D);* **[enter]**

Press **[close]** or file ⟶ save as

If you want to visualise, for example, the spectrum recorded at the first pH value:

Type *plot(w,D(1,:));* **[enter]**

Press **[close]** or file ⟶ save as

Find out the number of absorbing components of the system

1) Apply the SVD algorithm implemented in MatLab to the **D** matrix:

Type *[u,s,v]=svd(D);* **[enter]**

By doing this, three matrices, **u**, **s** and **v** will be calculated. **u (63 × 63)** and **v (70 × 70)** contain the eigenvectors for concentration and spectral profiles, respectively. **s (63 × 70)** contains the eigenvalues.

2) Look at the eigenvalues:

Double click on **s** in the workspace

From table 1 it can be observed that the eigenvalues decrease sharply with the number of components up to three.

Table 1: Eigenvalues obtained for a SVD decomposition performed on the D matrix.

components	eigenvalues
1	10.9840
2	3.4647
3	0.3820
4	0.0480
5	0.0123
6	0.0058
7	0.0028

Press **[close]** the matrix

3) Look at the eigenvectors of the concentration profile. The real important eigenvectors are those related with the first four eigenvalues. Plot only the first four eigenvectors (MatLab colour code: first, blue; second, green; third, red; fourth, cyan)

Plot eigenvectors for concentration profiles:

Type *plot(ph,u(:,1:4))*; **[enter]**

Press **[close]** or file ⟶ save as

The first three eigenvectors (figure 1) show a smooth shape; meanwhile the fourth eigenvector clearly contains an important amount of noise. Both items (eigenvalues and eigenvectors) suggest that the system is formed by three absorbing species. This result agrees with the previous knowledge of the system. The number of components of our system is three.

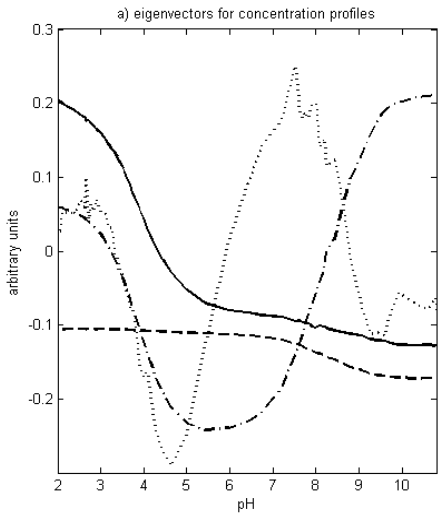


Figure 1: Eigenvectors obtained for the first four components (first, solid; second, dash-dotted; third, medium dash; fourth, dotted) in the concentration profile.

Choosing the initial estimates

As it has been mentioned in the Theory section, the more straightforward way to obtain initial estimates is to select three of the spectra from the **D** data matrix. As an example, different initial estimates will be used (S1 to S4). These matrices are also supplied by the authors. If you want to obtain by yourself:

1) Spectra obtained at pH 1.95, 5.98, 10.96 (the first, middle and last obtained spectra):

Type *S1=[D(:,1), D(:,37), D(:,63)];* **[enter]**

This generates a matrix on the workspace, **S1 (70 × 3)**. It contains the three spectra recorded at the mentioned pH to use them as initial estimates of the spectral profiles of the three components. To show the spectra:

Type *plot(w,S1)*; **[enter]**

Press **[close]** or file ⟶ save as

2) Spectra obtained at pH 2.54, 6.42, 9.04. **S2 (70 × 3)**

Type *S2=[D(:,7), D(:,38), D(:,54)];* **[enter]**

Type *plot(w,S2)*; **[enter]**

Press **[close]** or file ⟶ save as

3) Spectra obtained at pH 1.95, 3.10, 11.56. **S3 (70 × 3)**

Type *S3=[D(:,1), D(:,16), D(:,63)];* **[enter]**

Type *plot(w,S3)*; **[enter]**

Press **[close]** or file ⟶ save as

4) Spectra obtained at pH 1.95, 3.10, 11.56. **S4 (70 × 3)**

Type *S4=[D(:,1), D(:,52), D(:,63)];* **[enter]**

Type *plot(w,S4)*; **[enter]**

Press **[close]** or file ⟶ save as

als2004. Software by *Tauler et al (3)*.

1) Open the program

Type *als2004* **[enter]**

2) Selection the data set. Import the data matrices and initial estimates. As an example the S1 matrix will be used as initial estimates of the spectral profiles:

Data matrix ⟶ Press **[select a variable]** ⟶ Select *D*

Initial estimate ⟶ Press **[select a variable]** ⟶ Select *S1*

Nr. of matrices ⟶ Type *1*

Press **[continue]**

3) Selection of ALS constraints. The constraints to apply in our system are the logical ones in a spectrophotometric study: nonnegativity to concentration and spectral profiles; whereas unimodality and closure only to the concentration profiles.

No-negativity

Select *Yes?*

Select *Conc & spec.*

Implementation for conc ⟶ Select *fnnls*

Nr. of species with non-neg conc ⟶ Select *3*

Implementation for spec ⟶ Select *fnnls*

Nr. of species with non-neg spec ⟶ Select *3*

Unimodality

Select *Yes?*

Select *Conc.*

Implementation of the unimodality constraint ⟶ Select *average*

Nr. of species with unimodal conc ⟶ Select *3*

Constraint tolerance for conc ⟶ Type *1.05*

Closure. The total concentration of 8HQS in our system is 8.0×10^{-5} mol L^{-1}

Select *Yes?*

Select *Conc.*

No. of closure constraints to be included? ⟶ Select *1*

First closure constant Equal to ⟶ Type *8E-05*

Closure condition ⟶ Select *Equal Than*

Which species are in 1st closure? ⟶ Select *All*

4) Optimization parameters

No. of iterations ⟶ Type *50*

Convergence criterion ⟶ Type *0.1*

Select *Graphical output*

5) Output

Concentration ⟶ Type *copt*

Spectra ⟶ Type *sopt*

6) Perform MCR-ALS

Press **[optimize]**

When convergence is achieved %LOF and explained variance is shown

Press **[close]**

7) View the result (MatLab colour code: first, blue; second, green; third, red):

The program generates three matrices:

als_opt: Arranged matrix with the parameters used for MCR-ALS decomposition

copt (63 × 3): It contains the concentration profiles obtained.

sopt (3 × 70): It contains the spectral profiles obtained.

Plot the concentration profiles:

Type *plot(w,sopt);* **[enter]**

Press **[close]** or file ⟶ save as

Plot the spectral profiles:

Type *plot(ph,copt);* **[enter]**

Press **[close]** or file ⟶ save as

If you want, you can plot both figures at the same time:

Type *subplot(1,2,1);plot(ph,copt);subplot(1,2,2);plot(w,sopt);* **[enter]**

Press **[close]** or file ⟶ save as

On-Line Parallel Factor Analysis. A new stepforward in the monitoring of bioprocesses in real time.

José Manuel Amigo, Anna Surribas, Jordi Coello, Santiago Maspoch, José Luís Montesinos, Francisco Valero.

Sent to *Talanta*. May 2007

Anexos / Annexes

ON-LINE PARALLEL FACTOR ANALYSIS. A STEPFORWARD IN THE MONITORING OF BIOPROCESSES IN REAL TIME

José Manuel Amigo[a*], Anna Surribas[b], Jordi Coello[a], José Luís Montesinos[b], Santiago Maspoch[a], Francisco Valero[b]

[a] Unitat de Química Analítica, Departament de Química, Universitat Autònoma de Barcelona, E-08193 Bellaterra (Barcelona), Spain.

[b] Departament d'Enginyeria Química. ETSE. Universitat Autònoma de Barcelona, E-08193 Bellaterra (Barcelona), Spain.

* Corresponding author:

Mr. José Manuel Amigo

Unitat de Química Analítica. Departament de Química.

Universitat Autònoma de Barcelona.

08193 Bellaterra (Barcelona). Spain.

Tel. +34-93-581 1712

Fax +34-93-581 2379

e-mail address: josemanuel.amigo@uab.es

Abstract

There is a widely growing interest in obtaining robust and rapid methodologies capable of monitoring fermentation processes in real time. Different analytical methods have been adapted to measure cell density evolution throughout a culture, and fluorescence spectroscopy is a very promising technique due to its sensitivity, selectivity towards important chemical analytes and to its easy implementation as a non-invasive procedure.

This work is focused on showing the advantages of coupling the trilinear algorithm Parallel Factor Analysis (PARAFAC) to the Multivariate Statistical Process Control (MSPC) as a monitoring and real-time control tool for bioprocesses. In this context, two cultures, both involving the *Pichia pastoris* microorganism, were monitored by multiwavelength-fluorescence. In the first one methanol was used as substrate; whereas glycerol was used in the second one. Taking advantages of the mathematical properties of PARAFAC, batches of a bioprocess measured under normal operating conditions (NOC) were used to develop a calibration model; whereas the residuals of the model in combination with MSPC were used to create control limits. The control limits were used for new batches in real time.

Fluorescence spectroscopy combined with PARAFAC and MSPC is a feasible approach for controlling and performing fault diagnosis of bioprocesses offering the opportunity of doing real-time process surveillance based on relevant quality measurements.

Keywords: fluorescence spectroscopy; PARAFAC; MSPC; *Pichia pastoris*; bioprocess control; real-time monitoring.

1 INTRODUCTION

Real time monitoring and control of various quality parameters are necessary to optimise a bioprocess and to obtain a reproducible production. For this purpose, simple physical and chemical parameters have traditionally been measured. However, monitoring of biological and biochemical parameters such as biomass, metabolic activity, substrate consumption and product formation are needed in order to obtain the level of detail in understanding required to truly control the process efficiently (Olsson, 1997; Schügerl, 2001; Rhee, 2004; Blanco, 2004; Haack, 2004). In that sense**,** fluorimetric sensors are becoming a powerful tool due to their intrinsic advantages as the high specificity and sensitivity and their easy miniaturization (Skibsted, 2001; Kermis, 2002). Moreover, since fluorimetric probes can be implemented in non-invasively, they do not require sample withdrawal and thus the process sterility is not compromised. Most importantly, fluorescence often reflects important quality parameters either directly or indirectly.

The first on-line fluorescence sensor was based on the use of just one excitation and emission wavelength to monitor only a specific fluorophore (Zabriskie, 1978). Fluorometric spectrophotometers are now able to record the fluorescence intensity measured at a wider excitation and emission spectral range within a few seconds. In this context, multi-wavelength fluorescence implemented with an optical probe has turned into a very powerful non-invasive methodology (Lindemann, 1998).

Pichia pastoris is becoming an increasingly employed host system for the production of a wide variety of heterologous proteins, both in academic and industrial biotechnology. The application of fluorometric measurements to the monitoring of *P. pastoris* cultures is still in a preliminary stage. It has been observed that the appearance of some fluorophores such as tryptophan, NAD(P)H and riboflavin during the bioprocess is closely linked to the generation of biomass and consequently can be used for its indirect estimation (Surribas, 2006a). A procedure to estimate *P. pastoris* biomass concentration growing in glycerol or methanol from off-line fluorescence measurement of tryptophan has been reported (Surribas, 2006b). Hisiger and Jolicoeur (2005) used a multi-wavelength fluorescence probe to prove that NAD(P)H and riboflavin fluorescence emission could also be used for biomass estimation.

Multi-wavelength fluorescence spectroscopy applied to any process produces data sets with an intrinsically three-dimensional structure (time, excitation wavelength,

emission wavelength). Such complex signals require advanced chemometric procedures to extract the relevant information especially because the measurements are performed directly in the process and hence give highly overlapping contributions from many chemical compounds. (Li ., 1991; Lindemann ., 1998; Hagerdon , 2003; Boehl , 2003; Surribas , 2006c).

It has been demonstrated in a previous work that a multi-wavelength on-line fluorescence sensor combined with unfolding-PLS regression provided a quantitative prediction of biomass and substrates (glycerol and methanol) in the heterologous production of *Rhizopus oryzae* lipase. Heterologous lipase was satisfactory estimated in the exponential growth phase but not in the stationary phase when proteolytic degradation occurred (Surribas, 2006a).

This procedure has two main drawbacks: the difficulty to interpret the scores and loadings plots of unfold-PLS and the introduction of a non-existent correlation between the unfolded spectral profiles. Furthermore, reference concentrations of the different products are needed to construct the regression model. An alternative was proposed by Lopes and Menezes (Lopes, 2003); they pointed out that the control in real time of a bioprocess could be followed by means of multilinear-PLS (N-PLS) (Bro, 1996). The main drawbacks of the PLS-based methodologies are the need of reference measurements in order to develop the calibration model. Another drawback of the PLS-based methodologies is the time consuming steps of model development and validation.

Harshman (Harshman, 1970) and Carrol and Chang (Carrol and Chang, 1970) developed an algorithm that was able to handle three-way data systems (Bro, 1997a; Bro, 1997b; Andersen 2003): Parallel Factor Analysis (PARAFAC), which can be considered as a generalization of the well known Principal Components Analysis (PCA) to three-way systems to obtain important qualitative information avoiding the unfolding of the data-set. This algorithm has demonstrated to be a powerful tool to obtain qualitative information about the behaviour of the fluorophores during cultures evolution, being monitored by means of multi-wavelength fluorometry. Recently, several papers have been published showing the utility of linking PARAFAC with a multivariate regression method as PLS to obtain both qualitative and quantitative information about the bioprocess behaviour (Esteves da Silva 1999; Haack , 2004; Surribas , 2006c).

The main advantage of PARAFAC is that the true underlying phenomena (pure excitation and emission spectra of each fluorophore and their relative concentration

profiles) can be obtained from the model scores often eliminating the need for developing a calibration model. The qualitative information obtained with PARAFAC algorithm can be directly related to the evolution of different target biochemical parameters such as biomass, metabolic activity, substrate consumption and product formation. If the cultures used for the development of PARAFAC model are working in normal operating conditions (NOC), the model can be used thereafter to predict the behaviour of the fluorophores in new batches (Surribas 2006c).

Process Analytical Technology (PAT) is becoming an important tool to understand and control bioprocesses. The use of Multivariate Statistical Process Control (MSPC) can provide effective and efficient means for acquiring information to facilitate process understanding, to develop risk-mitigation strategies, to achieve continuous improvement, and to share information and knowledge (Wise, 1996; Louwerse , 2000). In this context, Louwerse and Smilde (2000) exposed an interesting study about the use of Multivariate Statistical Process Control (MSPC) applied to unfinished batches of simulated semi-batch emulsion polymerisation of styrene-butadiene using PARAFAC.

The main purpose of this work is to develop an approach for coupling the PARAFAC algorithm to MSPC as a real-time monitoring and control tool for biotechnology without the need of reference values. PARAFAC calibration models will be obtained from batches of a process working under normal operating conditions (NOC) and without the need for off-line measurements of target parameters. This information and MSPC will be used to create different control limits for taking action in terms out-of-control and end-of-batch for monitoring new batches. Real-time graphical control chart will be obtained. Two situations will be studied: in the first one, a bioprocess in which methanol was used as substrate for the growing of *P. Pastoris*. In the second situation, the substrate for the bioprocess will be glycerol and the calibration model will be developed with two batches working under normal operating conditions (NOC). Off-line measurements of biomass, methanol, glycerol and lipase were also carried out with the purpose of checking the results obtained by means of on-line PARAFAC-MSPC.

2 THEORETICAL BACKGROUND: PARAFAC algorithm as on-line method.

2.1 Building the PARAFAC calibration model

The proposed algorithm, Parallel Factor Analysis, PARAFAC, performes three dimensional models in a similar way that PCA algorithm in bilinear systems. In order to obtain good models, they must be built up from batches working at normal operating conditions (NOC). When a batch culture is monitored throughout the time by means of a multiwavelength fluorimetric probe, the excitation (λ_{ex}) and emission spectra (λ_{em}) are recorded at *n* times. The three-way matrix obtained, $\underline{M}$, has dimensions ($n \times \lambda_{ex} \times \lambda_{em}$) (Fig. 1a). This model will explain the behaviour of the fluorophores during the culture, that can be related to the biomass growth, the decrease of substrate and the formation of the product throughout the bioprocess and will constitute the "calibration stage". PARAFAC is an algorithm capable of decomposing three-dimensional systems into three bi-dimensional matrices which will model the data in a least squares sense (Fig.1a). Two of these matrices, named "spectral loading matrices", are related to the excitation (loading matrix B, $f \times \lambda_{ex}$) and emission (loading matrix C, $f \times \lambda_{em}$) spectral profiles of the *f* components in the system. The third matrix represents the scores matrix A ($f \times n$) and is related to the variation of the *f* components signal during the bioprocess. Mathematically, each m_{ijk} of $\underline{M}$ ($n \times \lambda_{ex} \times \lambda_{em}$) can be expressed as follow:

$$m_{n\lambda_{ex}\lambda_{em}} = \sum_{f=1}^{F} a_{nf} b_{\lambda_{ex}f} c_{\lambda_{em}f} + e_{n\lambda_{ex}\lambda_{em}} \qquad (1)$$

where *a*, *b* and *c* represent each element of matrices A, B and C, respectively; whereas *e* represents each element of the residual matrix E (Louwerse , 2000; Smilde , 2004). Another way of expressing this decomposition is as shown in Eq. 2:

$$M = A(C \circ B)^T + E \qquad (2)$$

where $\circ$ representates the Khatri-Rao product (Rao , 1971)

PARAFAC algorithm solves the system as a function of the number of *f* components. The election of the proper number of component is an important question, since each component can be related to each chemical and/or physical effect that makes the signal significantly different from the noise. The previous chemical knowledge about the system get ideas for what reasonable number of components can be (Haack , 2004). However, there can be unexpected events causing significant signal variations.

Therefore, mathematical and statistical methods are mostly used to assess the appropriate number of components. Several procedures such as the core consistency or the explained variance, have been proposed (Bro, 1997a) and will be used together with visual evaluation of the chemical interpretability in this application. In a previous paper (Surribas 2006c) several methods are suggested to calculate the number of components.

2.2 Residual statistic, Q*n*, as control parameters

A basic idea in MSPC is to compare the residuals of new batches with those obtained from the NOC data. In order to quantify possible significant changes the Residual statistic (Q*n*) is often used (Wise, 1996; Louwerse , 2000). The *Qn* statistic indicates the amount of variation in each sample not captured by the *f* components of the model (Wise, 1996). In this work we focus on the residual statistic *Qn* as the control parameter, in such a way as to indicate how well each sample conforms to the PARAFAC model. This parameter is defined as the sum of squared residuals obtained at time *i*

$$Qn = \sum_{j=1}^{J}\sum_{k=1}^{K} e_i^2 \qquad (3)$$

where e_i is the residual value defined in Eq. 4. The "normality" of this value can be assessed using confidence levels (Louwerse , 2000), estimated from the residuals from the calibration PARAFAC model. Assuming a normal distribution of the residuals in the calibration model, the 95 and 99% confidence interval are usually used as warning and alarm limits respectively (Figure 1b). These values are roughly equal to the mean residual plus 2 and 3 times the standard deviation, respectively. Alternative control chats may also be developed but this will not be pursued here.

2.3 Predicting and controlling new batch cultures

Once the model has been built from the calibration NOC batches, the fluorescence signal of new batches, P, can be monitored over the same excitation and emission wavelength ranges. The signal variation is expected to be caused by the evolution of the same fluorophores present in the calibration set, and, consequently, the spectral profiles obtained in the loading matrices B and C are assumed to remain

appropriate for new batches of the same bioprocess. Should a batch show variation in new unmodelled fluorophores this will show up in that the samples of that batch will get higher residuals for the scores matrix A.

The real-time control methodology proposed in this work has to be handled as follow. At any specific *i* time, a new matrix $\underline{P_i}$, of dimensions ($1 \times \lambda_{ex} \times \lambda_{em}$) is obtained. The new scores matrix a_i ($f \times 1$) for $\underline{P_i}$ can be predicted by using the loading matrices found in the PARAFAC model obtained from the batches at NOC (calibration stage). Smilde , (2004) defined this projection mathematically as follow:

$$a_i = [(C \circ B)^T (C \circ B)]^{-1} (C \circ B)^T p_i \qquad (4)$$

where p_i is the experimental value corresponding to each element of P_i matrix. The obtained a_i offers us information about the behaviour of each component at time *i*. This fitting is repeated for each new measured sample, so the scores, and thus the fluorophores behaviour through the process, can be obtained practically at real time.

The residuals of the sample e_i can be calculated as follow:

$$e_i = p_i - (C \circ B) a_i \qquad (5)$$

The *Qn* statistical parameter can be calculated with the e_i value calculated for each *i* sample (eq. 4). These values will be compared with the confidence limits imposed with the calibration model. If the value of the *Qn* at a given time is below the warning level, the process may be assumed to be under controlled conditions. If it is found within 95% and 99% it may indicate that some anomaly occurs with respect to the NOC; and outside 99% it is clearly out of control.

This working methodology can be applied during the whole fermentation process. All the steps needed for a new process monitoring are described in the flow chart of Figure 2.

3. MATERIALS AND METHODS

3.1 Strain

The wild type *P. pastoris* X-33 strain containing the vector pPICαA_ROL was used for heterologous expression of a *Rhizopus oryzae* lipase (ROL) under the control of the *AOX1* promoter (Minning , 2001).

3.2 Cultivation conditions

The medium used for starter cultures was YPD medium containing 1 % (w/v) yeast extract, 2 % (w/v) peptone and 2 % (w/v) dextrose.

Batch cultivations were carried out using a basal salt synthetic medium containing 0.02 M H_3PO_4, 6.8 × 10^{-3} M $CaSO_4$, 0.1 M K_2SO_4, 0.06 M $MgSO_4 \cdot 7H_2O$, 0.073 M KOH in a final volume of 1 l. Also 0.1 mL of Mazu DF 7960 antifoam (Mazer Chemicals, PPG Industries Inc., Guarnee, Ill., USA) was added at the beginning of the culture, 2 mL of a biotin solution (200 mg L^{-1}) and 4.35 ml of PTM1 solution. All solutions were prepared with distilled water.

In the first situation, two batch cultures were carried out using 10 $g \cdot L^{-1}$ of methanol as the carbon source. In the second situation three batch cultures were carried out using 40 $g \cdot L^{-1}$ of glycerol.

PTM1 contained per litre of distilled water: $CuSO_4 \cdot 5\ H_2O$ 6.0 g, NaI 0.08 g, $MnSO4 \cdot\ H_2O$ 3.0 g, $Na_2MoO_4 \cdot 2H_2O$ 0.2 g, H_3BO_3 0.02 g, $CoCl_2$ 0.5 g, $ZnCl_2$ 20.0 g, $FeSO_4 \cdot 7H_2O$ 65.0 g, biotin 0.2 g, H_2SO_4 concentrated 5 ml. Biotin and trace salts components were filtered-sterilized.

All cultivations were performed in a 2 L bioreactor Biostat B unit (Sartorious Biotech, Germany) at 30 ºC and pH 5.5 (controlled by adding NH_4OH 30 % v/v) under continuous stirring at 1000 rpm and air flow rate of 2.5 $L \cdot min^{-1}$.

3.3 On-line multi-wavelength fluorescence measurements and data treatment

Multiway fluorescence measurements were carried out using the BioView® sensor (DELTA Light & Optics, Lyngby, Denmark) (Lindemann , 1998) connected through a 25 mm port to a quartz window in the bioreactor.

The excitation wavelength range was from 270 to 550 nm in steps of 20 nm, and the emission wavelength range was from 310 to 590 nm, also in steps of 20 nm. The instrument was set-up to collect the spectra every 10 min.

Spectral matrices were imported into ASCII files for further treatments. The area containing no excitation – emission information was filled with zeros and missing values according to ref (Surribas 2006c, Thygesen , 2004). In this case missing values were set for emission below the excitation wavelength plus 40 nm and for emission below the excitation wavelength minus 40 nm, the intensities were set to zero. The PARAFAC algorithm used is implemented into PLS-Toolbox 3.5 (Eigenvector Research, WA, USA) working under MatLab Version 7.0 (MathWorks, Natick, MA, USA).

3.4 Off-line measurements

For biomass determination, 10 mL samples were filtered through a glass microfibre filter (Whatman GF/F; Maidstone, UK), washed with distilled water and dried at 105 °C until constant weight. Residual standard deviation (R.S.D.) was about 5%.

Methanol and glycerol were quantified by a Shimadzu GC-14B gas chromatograph with FID-Detector (Shimadzu Deutschland GmbH, Duisburg, Germany,) using a packed column Carbograph with 5% Carbowax 20 M, AT "steel 1/8" o.d., length 6 ft. (Alltech GmbH, Unterhaching, Germany). The operating conditions were 180 and 220 °C for injector and detector, respectively. Oven-temperature profile 70 °C, 10 °C min^{-1} to 170 °C. Analysis time was 7 min. Nitrogen was used as carrier gas with a flow rate of 30 ml·min^{-1}. As internal standard isopropanol (4 g·l^{-1}) mixed at 50% (v/v) with samples was utilized. Data were quantified by Shimadzu Class VP software (Shimadzu Deutschland GmbH, Duisburg, Germany). R.S.D. was estimated about 3.5%.

Lipolytic activity of the supernatant was determined by a modified method based on a colorimetric commercial kit assay (LIP kit from Roche Diagnostics, Mannheim, Germany), as described elsewhere (Cos , 2005). R.S.D. was estimated about 10%.

4. RESULTS AND DISCUSSION

4.1. Process using methanol as substrate

4.1.1 PARAFAC model.

Data from a batch, M, known to be under NOC was used to build the PARAFAC model. The explained variance and the core consistency obtained from PARAFAC calculations upon varying the number of components indicated that a suitable number of components is three (more than 99.7 % for the explained variance and 84% for the core consistency). Excitation and emission spectral profiles for these three components are shown in Figure 3a and 3b. They correspond to the excitation and emission spectra of the fluorophores involved in the process, which were identified as tryptophan, NADH and riboflavin.

Figure 3c shows the obtained score vectors together with the off line values of biomass, substrate and lipolytic activity. As it can be seen from the comparison of these values and profiles the scores can be used to follow the time course of the batch bioprocess and taking their shape into account, three different zones can be distinguished:

1) Latent period: During the first 4 hours of the process, the fluorophores scores profiles remain constant. The microorganism is adapting to the culture conditions in the bioreactor media and environment. If the scores evolution are compared to the biomass and substrate profiles, neither growth nor methanol consumption can be clearly observed.

2) Exponential phase: After the latent period, the score vectors intensities start to increase. The signal grows due to the formation of biomass and its related fluorophores. In this stage, the yeast is exponentially metabolising methanol to biomass, energy and other products.

3) Stationary phase: After the exponential growth phase, the stationary phase begins. There is a sudden NADH decreasing and, thereafter, the three fluorophores profiles remain constant throughout the time. This evolution of the scores profiles was compared to the off-line measurements of biomass, methanol and lipase production to asses to which extent the PARAFAC scores could be taken to be indicative of these fundamental quality characteristics. Figure 3c shows that methanol exhaustion and the maximum in the lipase production match with the decreasing in the NADH signal. For this reason, the decreasing of NADH signal can be used as an objective parameter to decide whether the culture is ending.

The obtained PARAFAC model of the fluorescence matrix has demonstrated its usefulness in the elucidation of the time course of a batch bioprocess simply analysing the shape of the scores related to the corresponding spectral loadings and with no need for off-line values for any calculation except from a visual comparison.

As it has been explained in the Theoretical Background of this work, the 95 and 99% confidence interval of the residuals of the calibration model were calculated to establish the warning and alarm limits, respectively (figure 4b).

4.1.2 Monitoring and real-time control of a new batch.

After performing the calibration model and establishing the control limits, a new batch can be then run and its gathered fluorescence signal can be used to obtain the scores profiles for each individual sample as indicated in figure 1b. In Figure 4 the evolution of the scores is shown together with some off-line measurements performed to contrast our conclusions. As it has formerly been detailed in the Theoretical Background section, the loading matrices for the calibration PARAFAC model can be used to calculate the new score at any specific sampling time *i*.

Although Figure 4 indicates that the shape of the scores profiles is similar to the scores obtained from the NOC process, a question still remains to be solved: is the new bioprocess running under NOC conditions? To investigate this, the Residual Statistics (*Qn*) is calculated for each sample (Figure 4b). As it can be seen *Qn* value at the beginning of the process is slightly above of the 95% confidence line. This high value could indicate that some unexpected event is occurring. Nevertheless, it has to be pointed out that the culture was starting and the signal-to-noise ratio is still low. The latent period depends on many factors, such as inoculums culture time or biomass concentration. Although kept as constant as possible, these factors may slightly vary from batch to batch and therefore model prediction power can be lower during this stage. As seen in Figure 4a, the latent period is over 4 hours and the scores profiles are as expected.

Then, the bioprocess enters into the exponential phase. During this stage the scores profiles behave as expected. Moreover, residuals are under the critical confidence levels and indicate that the bioprocess is under controlled conditions.

The sample matrix collected after 34 hours ($\underline{P}_{34}$ in Figure 4a) shows a score value for NADH clearly lower than the previous one, while the score value for

tryptophan reaches a constant value. This fact may indicate that, at this point of the culture, substrate was exhausted. The time interval of data acquisition is decreased to verify the ending of the fermentation. The subsequent measurements assess the NADH signal decreases while, afterwards, both the score values for NADH and tryptophan remain constant. At this point the analysis of the Q*n* values becomes necessary. Given that its level remains under the critical value of 95%, it can be considered that the culture stays under controlled conditions. With this information it can be suggested that substrate is totally metabolized, having enough information for the possibility of re-feed the culture, or end the process (figure 2).

4.2. Batches using glycerol as substrate

4.2.1. PARAFAC model.

In this new situation, two NOC batches were used for the construction of the calibration model. Now, glycerol was the carbon source for the growth of *Pichia pastoris*.

Figure 5 shows the excitation and emission loadings and the scores profiles obtained with PARAFAC. This culture was thoughtfully studied in a previous work (Surribas 2006c). It was observed that a fourth component was necessary to explain correctly the behaviour of the culture (apart from the tryptophan, NADH and Riboflavin). This fourth component was associated to the slight precipitation of orthophosphate (HPO_4^{2-}) with Mg^{2+} and Ca^{2+}, and other polyvalent ions present in PTM1 salts solution at pH = 5.5.

The evolution of the scores profiles for the three main fluorophores is consistent with the three main phases of the culture (latent period, exponential growth and stationary phase) being the decreasing in the signal of NADH and the constant signal of tryptophan the two main parameters that indicate substrate depletion. As in the first situation, residuals of this model were used to obtain the two confidence levels and create the Q-chart for new batches. Off-line measurements for biomass, substrate and lipolytic activity on calibration batches are not necessaries and they were not used in order to establish the decision parameters for the evolution of the process.

4.2.2. Monitoring and real-time control

A new batch was started and measurements were recorded each 30 minutes. Measurements were individually modelled with the calibration model previously performed following the same methodology than in the first situation. Figure 6 shows the scores profiles and the *Qn* values obtained for each individual sample of a new bioprocess. The scores evolution for the four components indicates that the end of the process occurs 17 hours after the beginning of the culture (NADH decreasing, constant signal of tryptophan and depletion of the fourth component). Off-line measurements were used to confirm that the substrate was totally metabolized at this time. As it can be observed in figure 6, glycerol was totally metabolized after 17 hours, and biomass production reached the maximum.

Qn values indicate that the process was under control since all of them kept below the first confidence level. At the end of the process, *Qn* values are over the confidence levels. This situation may be a consequence of the depletion of the signal for the fourth component. Nevertheless, at this moment the culture has already finished.

5. CONCLUSSIONS

In this work a new working methodology using PARAFAC and MSPC has been applied for the monitoring and quality control of bioprocesses monitored by fluorescence. The main novelty of the method lays in a new consideration of the data sets collected by using multivariate fluorescence spectroscopy, since the 2-D matrices obtained for each sampling time ($\lambda_{ex} \times \lambda_{em}$) can be visualized as a 3-D matrix where the first dimension is the time ($1 \times \lambda_{ex} \times \lambda_{em}$). PARAFAC has proved to be a proper and easy-to-use tool to extract the relevant information of three-way systems. This algorithm allows the monitoring and potential fault diagnosis of a bioprocess in real time.

PARAFAC decomposition of the calibration batch ($\underline{M}$) has allowed us to identify important fluorophores involved in the microorganism growth. Three different culture phases have been identified during the batch evolution. The information contained in the spectral data has been used to build a model suited for real-time monitoring of different batches without needing reference concentrations of biomass, methanol and lipase.

In this laboratory work, culture conditions were highly controlled, with little variability, and only one or two NOC batches were used to develop the calibration model. This situation is not very common in the industry, were culture conditions may have significant variations from a batch to another; so in order to assure that the calibration model reflects all the variability of the process, more NOC cultures have to be used in building the calibration model.

The on-line determination of the Q*n* for any given sampling time in the new batch presents the advantage of assessing that the culture is under controlled conditions as assessed in terms of the chemical fingerprint obtained from a PARAFAC model of the fluorescence measurements. Any unexpected event that is either directly or indirectly reflected in the fluorescence will be recognized by an unusual scores profile together with a Q*n*-statistic above the critical level. The required time to obtain the $\underline{P}_i$ matrix and its PARAFAC decomposition do not exceeds more that 5 minutes, which, given the culture duration, is a suitable time interval to consider a real time actuation at the operator-level. Besides, the implementation of PARAFAC to the real time fluorescence measurements could give the operator the information (scores profiles and *Qn* values) in a user-friendly interface. Moreover, the proposed methodology offers objective decision parameters, - the NADH signal decreasing- to assess substrate depletion and to give the operator the possibility if either stopping the cultivation process or re-feeding the batch culture. In addition, the calculation of the *Qn* value at real time offers the advantage of assessing culture's controlled conditions and can be utilized to make a decision for further actions.

From a biotechnological viewpoint, a process monitoring as the above is relevant since a diagnosis of a bioprocess can be made without the need of any reference value. Off-line data have been used in this work only to facilitate the process interpretation but they have not been used to take any decision regarding normality of a new batch. The batch end is translated in a decay of the NADH associated fluorescence signal while the MSPC calculations have be used to assess process controlled conditions.

Moreover, the assessment of the batch end together with the confidence of a bioprocess under normal operational conditions, offers a new possibility for the operator: the batch could be ended or start a fed-batch strategy adding methanol. In case the *Qn* value was too high there would be an objective parameter to stop the culture or to identify what is unexpectedly running.

Acknowledgments

The authors want to thank to the Spanish programme of Chemical Processes Technologies (projects CTQ 2004-02013 and CTQ 2004-00300) and Dursi (Generalitat de Catalunya) 2005-SGR-00698. Besides, we would like to thank Thomas Scheper group (Institut für Technische Chemie, Universität Hannover) for the support received with all the on-line fluorescence data presented in this work.

The Department of Chemical Engineering (UAB) is a member of the Unit of Biochemical Engineering of the Centre de Referència en Biotecnologia de Catalunya (Generalitat de Catalunya).

José Manuel Amigo thanks the Generalitat de Catalunya for his four-month pre-doctoral stay and Prof. Rasmus Bro for his comments and interest. Anna Surribas thanks the Ministerio de Educación y Ciencia of Spain for her predoctoral fellowship.

References

Andersen, C.M., Bro, R., 2003. Practical aspects of PARAFAC modeling of fluorescence excitation-emission data. J. Chemom. 17, 200-215

Blanco, M., Peinado, A., 2004. Analytical monitoring of alcoholic fermentation using NIR spectroscopy. J. Mas. Biotechnol. Bioeng. 88-4, 536-542

Boehl, D., Solle, D., Hitzmann, B., Scheper, T., 2003. Chemometric modeling with two-dimensional fluorescence data for *Claviceps purpurea* bioprocess characterization. J. Biotechnol. 105:1-2, 179-188

Bro, R., 1996. Multiway calibration. Multilinear PLS. J. Chemom. 10:1, 47-62.

Bro, R., Workman Jr., J.J., Mobley, P.R., Kowalski, B.R., 1997. Review of chemometrics applied to spectroscopy: 1985-95, Part 3 - multi-way analysis Appl. Spectrosc. Review, 32:3, 237-261.

Bro, R., 1997 PARAFAC. Tutorial and applications. Chemom. Intell .Lab. Syst. 38, 149-171.

Carrol, J.D., Chang, J., 1970. Analysis of individual differences in multidimensional scaling via an N-way generalization of and Eckart-Young decomposition. Psychometrika, 35, 283.

Cos, O., Serrano, A., Montesinos, J.L., Ferrer, P., Cregg, J.M., Valero, F., 2005. Combined effect of the methanol utilization (Mut) phenotype and gene dosage on recombinant protein production in *Pichia pastoris* fed-batch cultures, J. Biotechnol. 116, 321–335.

Esteves da Silva, J.C.G., Oliveira, C.J.S., 1999. Parafac decomposition of three-way kinetic-spectrophotometric spectral matrices corresponding to mixtures of heavy metal ions. Talanta, 49, 889-897.

Haack M.B., Eliasson A., Olsson, L., 2004. On-line cell mass monitoring of *Saccharomyces cerevisiae* cultivations by multi-wavelength fluorescence. J. Biotechnol. 114, 199-208.

Hagerdon, A., Legge, R.L., Budman, H., 2003. Evaluation of spectrofluorometry as a tool for estimation in fed-batch fermentations. Biotechnol. Bioeng. 83, 104-111.

Harshman, R.A., 1970. Foundations of the PARAFAC procedure: Model and conditions for an 'explanatory' multi-mode factor analysis. UCLA Working papers in phonetics. 16, 1.

Hisiger, S., Jolicoeur, M., 2005. A multiwavelength fluorescence probe: Is one probe capable for on-line monitoring of recombinant protein production and biomass activity. J. Biotechnol. 117, 325-326.

Kermis, H.R., Kostov, Y., Harms, P., Rao, K., 2002. Dual excitation ratiometric fluorescent pH sensor for noninvasive bioprocess monitoring: Development and application. Biotechnol. Prog. 18, 1047-1053.

Li, J.K., Asali, E.C., Humphrey, A.E., Horvath, J.J., 1991. Monitoring cell concentration and activity by multiple excitation fluorometry. Biotechnol. Prog. 7, 21-27.

Lindemann, C., Marose, S., Nielsen, H.O., Scheper, T., 1998. 2-Dimensional fluorescence spectroscopy for on-line bioprocess monitoring. Sens. Actuators. B 51, 273-277.

Lopes, J.A., Menezes, J.C., 2003. Industrial fermentation end-product modelling with multilinear PLS. Chemom. Intell. Lab. Sys. 68:1-2, 75-81.

Louwerse, D.J., Smilde, A.K., 2000. Multivariate statistical process control of batch processes based on three-way models. Chem. Eng. Sci. 55, 1225-1235.

Minning, S., Serrano, A., Ferrer, P., Solà, C., Schmid, R.D., Valero, F., 2001. Optimization of the high-level production of *Rhizopus oryzae* lipase in *Pichia pastoris*. J Biotechnol. 86, 59-70.

Olsson, L., Nielsen, J., 1997. Online and in situ monitoring of biomass in submerged cultivations. Trends Biotechnol. 15, 517-522.

Rao, C.R., Mitra, S., 1971. Generalized inverse of matrices and its applications. New York: Willey, es J. Willey & Sons comprobar como se citan libros 12-13.

Rhee, JI., Ritzka, A., Scheper, T., 2004. On-line monitoring and control of substrate concentrations in biological processes by flow injection analysis system. Biotechnol. Bioproc. Eng. 9, 156-165.

Schügerl, K., 2001. Progress in monitoring and control of bioprocesses during the last 20 years. J. Biotechnol. 85, 149-173.

Skibsted, E., Lindemann, C., Roca, C., Olsson, L., 2001. On-line bioprocess monitoring with a multi-wavelength fluorescence sensor using multivariate calibration. J. Biotechnol. 88:1, 47-57.

Smilde, A.K., Bro, R., Geladi, P., 2004. Multi-Way Analysis. Applications in the chemical sciences. J. Wiley & Sons

Surribas, A., Geissler, G., Gierse, A., Scheper, T., Hitzmann, B., Montesinos, J.L., Valero, F., 2006. State variables monitoring by in situ multi-wavelength fluorescence spectroscopy in heterologous protein production by *Pichia pastoris* J. Biotechnol. 124:2, 412-419.

Surribas, A., Montesinos, J.L., Valero, F., 2006. Biomass estimation using fluorescence measurements in *Pichia pastoris* bioprocess. J. Chem. Technol. Biotechnol. 81, 23-28.

Surribas, A., Amigo, J.M., Coello, J., Montesinos, J.L., Valero, F., Maspoch, S., 2006. Parallel Factor Analysis combined with PLS regression applied to the on-line monitoring of *Pichia pastoris* cultures. Anal. Bioanal. Chem. 385, 1618-2642.

Thygesen, L.G., Rinnan, A., Barsberg, S., Møller, J.K.S., 2004. Stabilizing the PARAFAC decomposition of fluorescence spectra by insertion of zeros outside the data area. Chemom. Intell. Lab. Sys. 71:2, 97-106.

Wise, B.M., Gallagher, N.B., 1996. The process chemometrics approach to process monitoring and fault detection. J. Proc. Cont. 6, 329-348.

Zabriskie, D. W., Humphrey, A. E., 1978. Estimation of fermentation biomass concentration by measuring culture fluorescence. Appl. Eur. Microbiol. 35, 446–450.

Figure Captions

Figure 1: Application of the on-line PARAFAC methodology

Figure 2: Decision flow-chart for the control of the bioprocess.

Figure 3a) excitation loadings, **b)** emission loadings and **c)** scores values obtained with PARAFAC model applied to the calibration batch with methanol sustrate. Solid, tryptophan; dashed, NADH; dotted, riboflavin. Dotted lines in Fig. 3c correspond to the off-line measurements (circle, methanol; square, biomass; triangle, heterologous product). The left axis corresponds to the scores values obtained with PARAFAC; whereas the right axis corresponds to the off-line measurements. The heterologous product concentration is expressed in Units of Activity · mL^{-1} ($UA \cdot mL^{-1}$).

Figure 4: Scores values and Q*n* values obtained with PARAFAC model applied to each $\underline{P}_n$ batch (empty circle, tryptophan; empty square, NADH; empty triangle, riboflavin) in methanol cultures. Dotted lines correspond to the off-line measurements (solid circle, methanol; Solid Square, biomass; solid triangle, heterologous product) The left axis corresponds to the scores values obtained with PARAFAC. Right axis corresponds to the off-line measurements. Heterologous product concentration is expressed in Units of lipolytic activity·ml^{-1}.

Figure 5a) Excitation loadings, **b)** emission loadings and **c)** scores values obtained with PARAFAC model applied to the calibration batches with glycerol substrate. Solid, tryptophan; dashed, NADH; dotted, riboflavin. Dash-dotted, fourth-component.

Figure 6: Scores values and Q*n* values obtained with PARAFAC model applied to each $\underline{P}_n$ batch (empty circle, tryptophan; empty square, NADH; empty triangle, riboflavin; cross, fourth component) in glycerol cultures. Dotted lines correspond to the off-line measurements (solid circle, glycerol; Solid Square, biomass). Left axis corresponds to the scores values obtained with PARAFAC. Right axis corresponds to the off-line measurements.

Figure 1

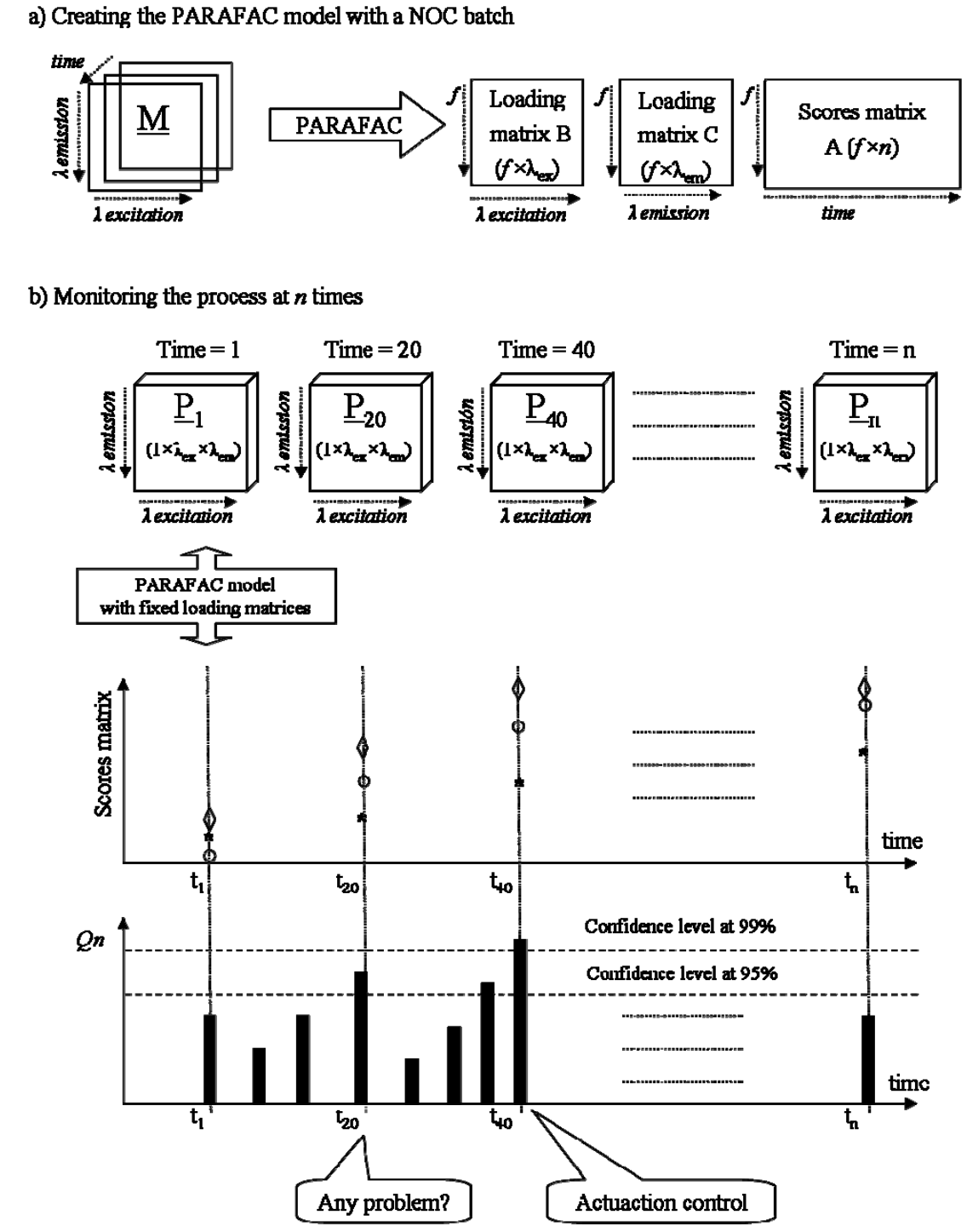

a) Creating the PARAFAC model with a NOC batch
time
λ emission
λ excitation
M
PARAFAC
f
Loading matrix B (f×λex)
Loading matrix C (f×λem)
Scores matrix A (f×n)
λ excitation
λ emission
time
b) Monitoring the process at n times
Time = 1
Time = 20
Time = 40
Time = n
P1
P20
P40
PII
(1×λex×λem)
λ emission
λ excitation
PARAFAC model with fixed loading matrices
Scores matrix
time
t1
t20
t40
tn
Qn
Confidence level at 99%
Confidence level at 95%
time
Any problem?
Actuaction control

Figure 2

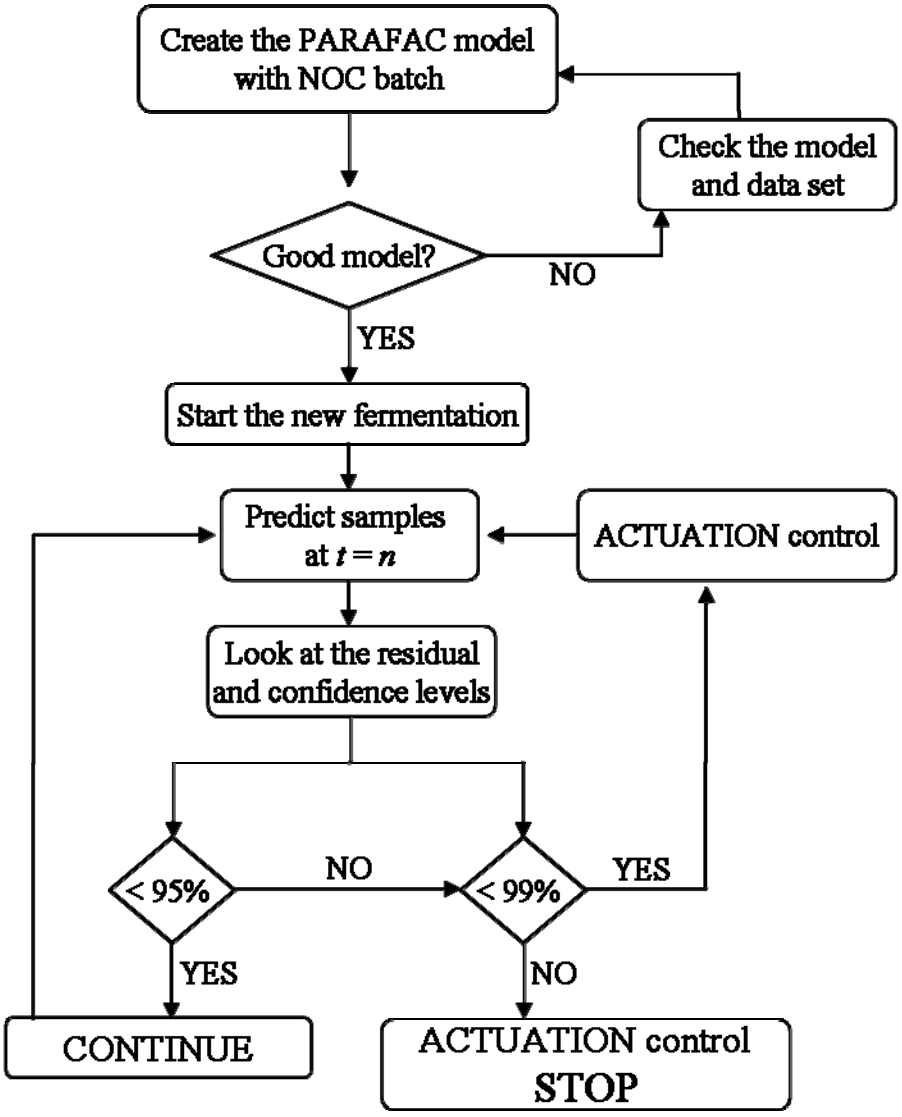

Create the PARAFAC model with NOC batch
Check the model and data set
Good model?
NO
YES
Start the new fermentation
Predict samples at $t = n$
ACTUATION control
Look at the residual and confidence levels
< 95%
NO
< 99%
YES
YES
NO
CONTINUE
ACTUATION control
STOP

Figure 3

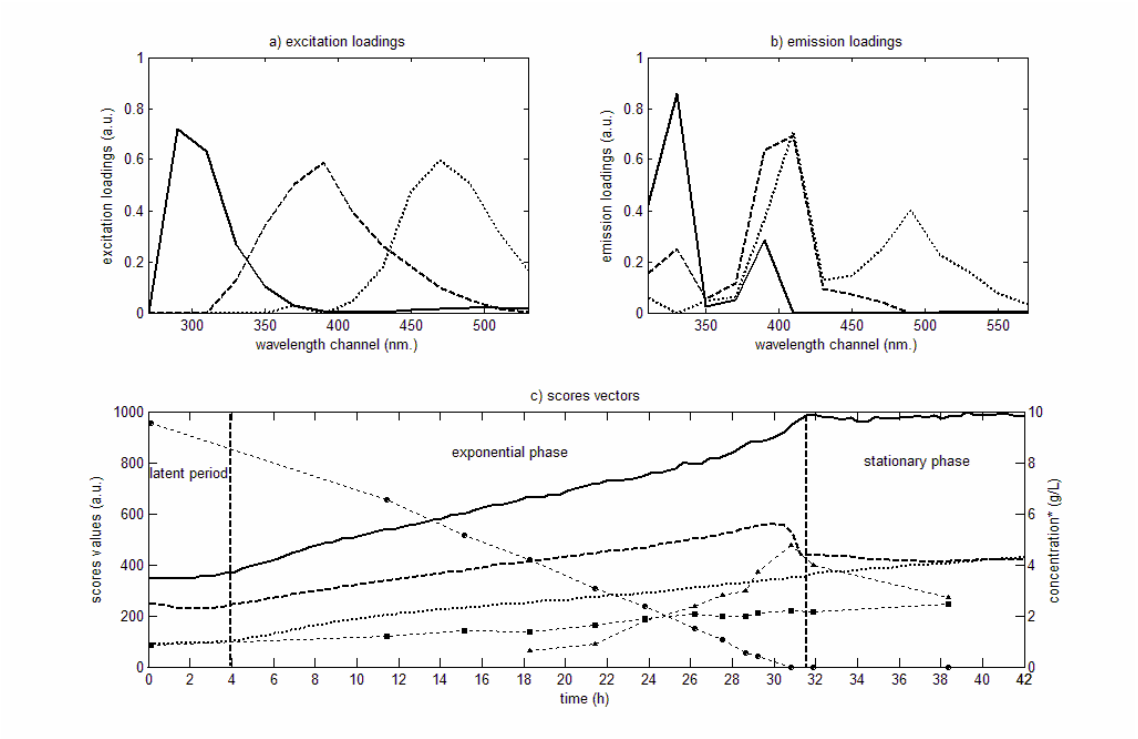

a) excitation loadings
excitation loadings (a.u.)
wavelength channel (nm.)
b) emission loadings
emission loadings (a.u.)
wavelength channel (nm.)
c) scores vectors
latent period
exponential phase
stationary phase
scores values (a.u.)
concentration* (g/L)
time (h)

Figure 4

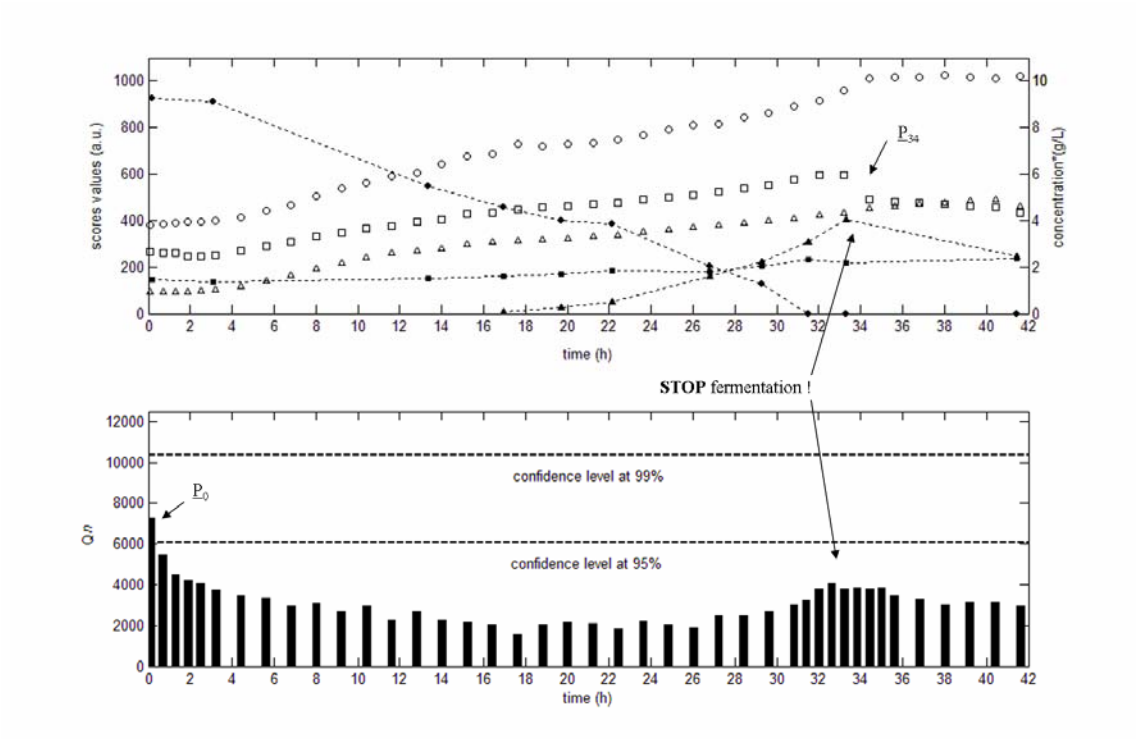

scores values (a.u.)
concentration*(g/L)
time (h)
P34
STOP fermentation !
Qn
P0
confidence level at 99%
confidence level at 95%
time (h)

Figure 5

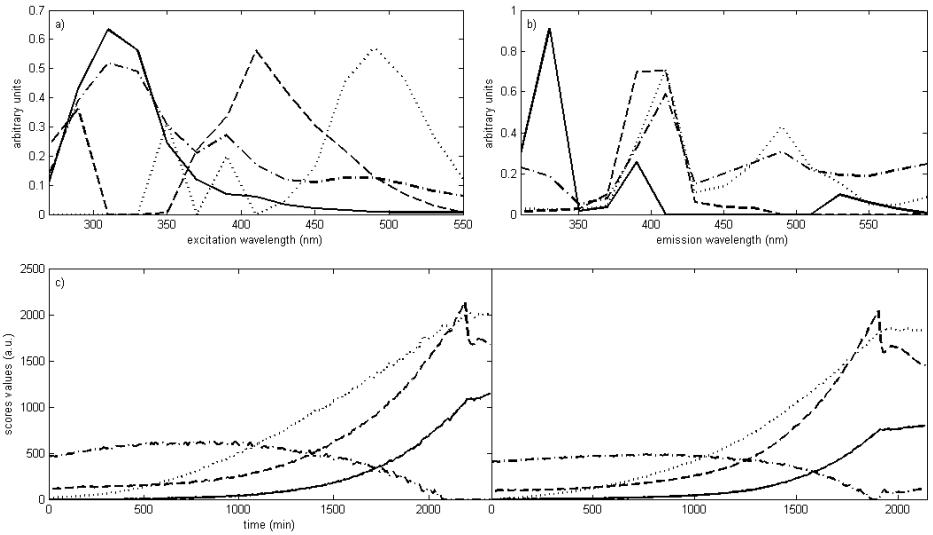
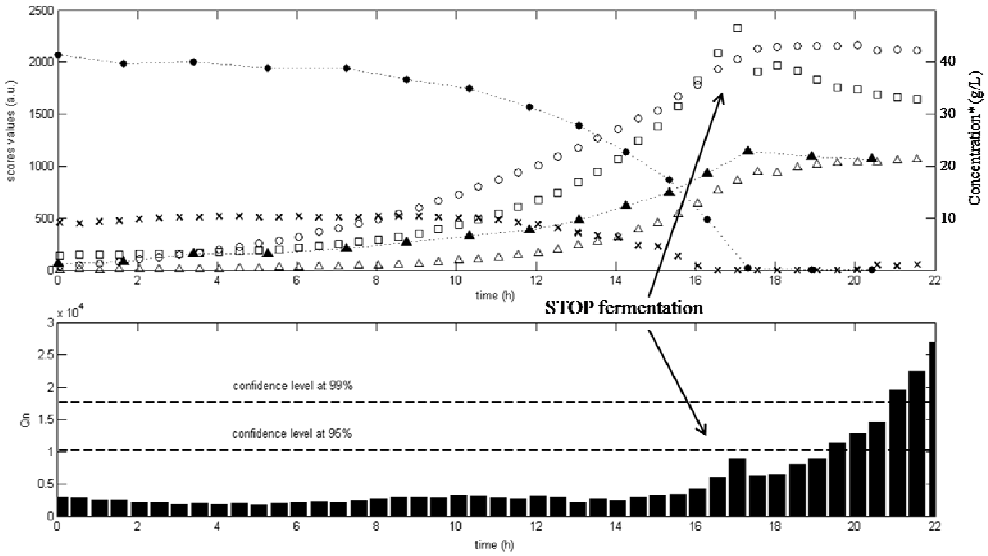

a)
arbitrary units
excitation wavelength (nm)
b)
arbitrary units
emission wavelength (nm)
c)
scores values (a.u.)
time (min)

Figure 6


scores values (a.u.)
Concentration* (g/L)
time (h)
STOP fermentation
x 10^4
confidence level at 99%
confidence level at 95%
Qn
time (h)

ANEXO 2 / ANNEX 2

Aportaciones a congresos y reuniones nacionales e internacionales
Conferences and congresses

Anexos / Annexes

1. 9th International Conference in Chemometrics in Analytical Chemistry.

Lisbon, Portugal. September 2004. Poster contributions:

"Simultaneous determination of xanthine and hypoxanthine in human urine in presence of uric acid."

"Three-way methods for the modelling of complex enzymatic systems. Comparison between PARAFAC, PARAFAC2 and MCR-ALS."

2. 1r. Workshop de la Xarxa Catalana de Quimiometria.

Barcelona, Spain. September 2005. Oral contribution (Spanish):

"Resolución Multivariable de Curvas restringida (HS-MCR) aplicada al análisis cinético de oxipurinas y ácido úrico en orina humana." (Hard-Soft Multivariate Curve Resolution applied to the kinetic analysis of oxipurines and uric acid in human urine)

3. 11as Jornadas de Análisis Instrumental.

Barcelona, Spain. November 2005. Posters contributions:

"Parallel Factor Analysis combined with Partial Least Squares Regresión applied to the on-line monitoring of *Pichia Pastoris* cultures."

"Hard-Soft Multivariate Curve Resolution applied to the quantitative determination of oxipurines and uric acid in human urine."

"An introduction to Multivariate Curve Resolution-Alternating Least Squares. Spectrophotometric study of acid-base equilibria of 8-hydroxyquinoline-5-sulfonic acid."

4. XX Reunión Nacional de Espectroscopía. IV Congreso Ibérico de Espectroscopía

Ciudad Real, Spain. September 2006 **.** Oral contribution (Spanish):

"Multidimensional fluorescence and three-way Chemometrics. A new Step forward in the monitoring of bioprocesses in Real Time."