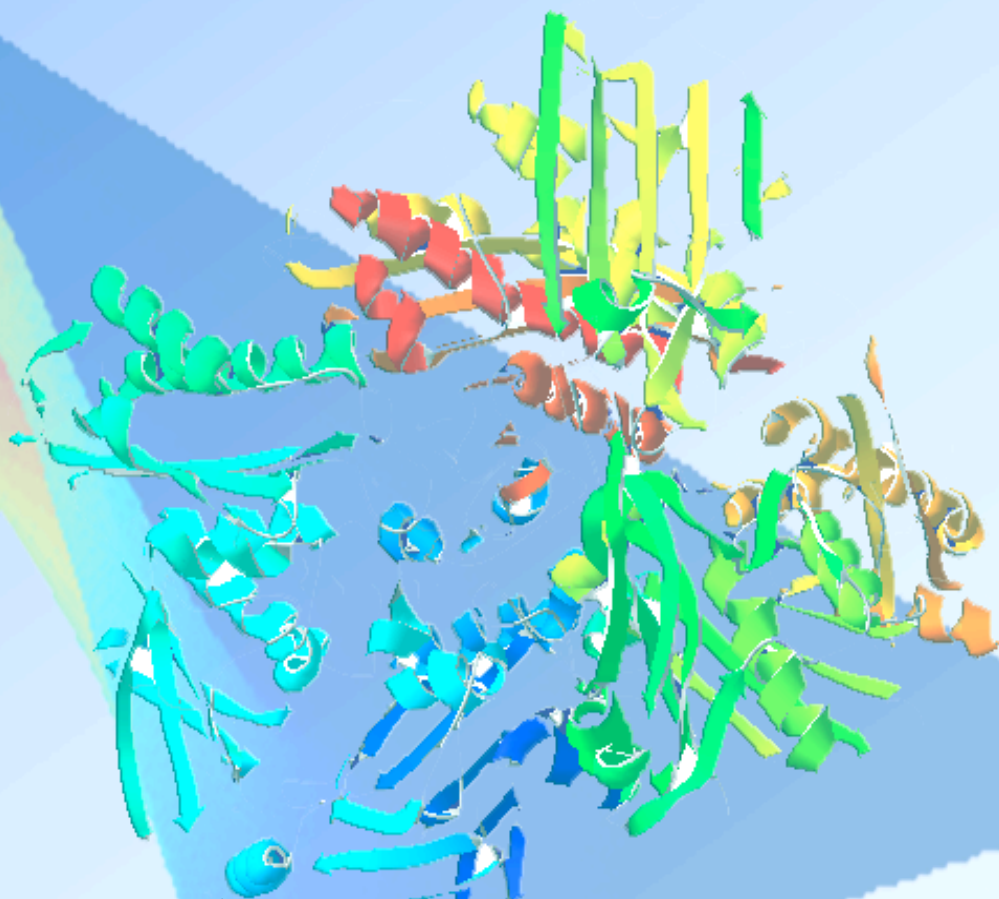


UAB

Universitat Autònoma de Barcelona

**DESARROLLO Y APLICACIÓN DE
MÉTODOS QUIMIOMÉTRICOS
MULTIDIMENSIONALES AL ESTUDIO DE
SISTEMAS ENZIMÁTICOS**



José Manuel Amigo Rubio

Directores:

Dr. Santiago Maspoch Andrés

Dr. Jordi Coello Bonilla

Tesis Doctoral

Julio 2007



Universitat Autònoma de Barcelona

**DESARROLLO Y APLICACIÓN DE
MÉTODOS QUIMIOMÉTRICOS
MULTIDIMENSIONALES AL ESTUDIO DE
SISTEMAS ENZIMÁTICOS**

José Manuel Amigo Rubio

Memoria presentada para optar al grado de Doctor.

Julio 2007

El Dr. Santiago Maspoch Andrés y el Dr. Jordi Coello Bonilla, catedráticos de Química Analítica de la Facultad de Ciencias de la Universidad Autónoma de Barcelona.

CERTIFICAN: Que el presente trabajo de investigación titulado

“Desarrollo y aplicación de métodos quimiométricos multidimensionales al estudio de sistemas enzimáticos”

que constituye la memoria presentada por **José Manuel Amigo Rubio** para optar al grado de Doctor, ha estado realizado bajo nuestra dirección y recoge, desde nuestro punto de vista, los requisitos exigidos para la consecución del citado título.

Para que quede constancia, firmamos el presente certificado a catorce de mayo de 2007

Dr. Santiago Maspoch Andrés

Dr. Jordi Coello Bonilla

AGRADECIMIENTOS

Hoy es un día perfecto para escribir este apartado de la tesis doctoral. Quizás porque estoy a más de 2200 kilómetros de la gente que me quiere. Escribir estas líneas conlleva mucho más de lo que me pensaba. Conlleva tener que mirar hacia atrás en el tiempo. A un 2 de Septiembre de 2002 a las 11 de la noche cuando Idoia y yo cogimos el tren en la estación de Abando con destino a Barcelona. Conlleva repasar todos los buenos (y malos) momentos que han pasado en estos 4 años...y pico.

Los momentos y las personas que me rodean forman parte esencial de mí. Con ellos he reído, llorado, ido de fiesta, sufrido y crecido como persona y como investigador. Por esa razón quisiera dedicar estas cuatro líneas...y pico a todas las personas y a todos los momentos que han colaborado, directa o indirectamente, en esta aventura que se llama “tesis doctoral”.

PERSONAS

En primer lugar, quisiera agradecer al Dr. Blanco la oportunidad de formar parte del grupo de Quimiometría Aplicada. Aquí me he formado como investigador y como quimiométrico. Espero seguir manteniendo buenas conversaciones con usted sobre dónde colocar el MCR en una tabla o cómo calcular modelos PARAFAC. Sinceramente, gracias.

Al Dr. Maspoch y al Dr. Coello. Santi, Jordi, gracias por confiar en mí hace 4 años...y pico y en mis locas ideas que al final han ido dando sus frutos poco a poco. Espero no haberos defraudado. Hortensia, gracias por saber cuidar de nosotros.

A Juanma, Luís Angel, Gorka y la gente del seminario de la UPV-EHU. Luís, tú me metiste en esto de la investigación y posibilitaste el contacto con Jordi. Juanma, todas las conversaciones y sabios consejos que me fuiste dando los tengo muy presentes todavía. Gorka, gracias por enseñarme a saber llevar una investigación. Si algún día soy jefe, te tendré de ejemplo.

Al Dr. Romà Tauler y, en especial, a la Dra. Anna de Juan. Gracias Anna por escucharme y por todos los consejos, quimiométricos y no tanto, que me has dado.

Conocer a grandes investigadores en tu campo es difícil. Si además vienen acompañados de una calidad humana impresionante, puede que te encuentres muy poquitos casos. Uno de ellos es el Profesor Rasmus Bro. Rasmus, thanks you very much for all. For this question...Jeg vil gerne takke Gilda, Frans, Thomas og især Rasmus for al opmærksomheden og hjælpen gennem mit ophold i Danmark gennem mit ophold ved Kvalitet og Teknologi gruppen.

Toda la gente que he tenido a mí alrededor. Mis herman@s, amig@s, Alumn@s, laborantes, compañeros de laboratorio (los que fueron y los que son) y anexos, de piso, de karate, de cervezas y shisha. La gente del IIQAB-CSIC, personas que me dieron todo su amor y toda su amistad. A todos ellos infinitas gracias. De cada uno de vosotr@s he aprendido cosas. Gracias a las fuerzas que me habéis sabido dar cuando me han hecho falta. Seguro que alguna vez os he fallado. Os pido mil perdones. Ya sabéis que soy un pequeño gran desastre.

Gracias a mi familia y, en especial, a mis padres, **Manuel Amigo** y **Rosa Rubio**, que, aunque no sin esfuerzo, han sabido comprenderme y sé que han estado en todo momento a mi lado. Eskerrik asko bihotz, bihotzetik, aita eta ama.

MOMENTOS

Gracias a las noches de cerveza, guitarra y shisha con las que tanto nos hemos reído. A los momentos en el Golfo Norte y en la Terraza. A esos “pitis” furtivos con el café. A los cafés en el Dowells de Portugalete. Gracias hermanitos.

Un día, en las escaleras de una playa, un ángel vino y me hizo ver que no me estaba comportando demasiado bien con la gente que me rodeaba. Me hizo abrir los ojos. Eternamente agradecido pequeño angelito.

Gracias por los abrazo espontáneos que llegaron sin que yo los pidiese pero que tanto necesitaba. Supiste ver a través de mis ojos como papel transparente.

Hace unos años me regalaste una brújula en un bar irlandés de Barcelona. Lo hiciste para ayudarme a encauzar mi rumbo. Lástima que ese rumbo al final de este viaje se perdió por mi loca cabeza. ¿Sabes qué? Estoy volviendo a hacer que esa brújula indique el Norte poco a poco. Gracias, pequeña Amelie. Sigues siendo muy importante para mí.

Un lugar, una situación y un objeto. Por un momento creí estar soñando. Cerré los ojos y cuando los volví a abrir miré a mi lado y vi que no era un sueño. Apreté fuerte por si acaso...

Idoia, corre, despierta!! Ya estamos en Barcelona!!

Jose!! Joder!! Cállate y déjame dormir en paz!!

*“Mai no és massa tard per tornar a començar,
per sortir a buscar el teu tresor”*

Camins – Sopa de Cabra

*“Fear is the path to the Dark Side.
Fear leads to anger, anger leads to Hate
Hate leads to Suffering”*

Yoda Master

ÍNDICE GENERAL

INDICE GENERAL.....	i
INDICE DE TABLAS.....	vii
INDICE DE FIGURAS.....	viii
LISTADO DE ABREVIATURAS.....	xi
CONTEXTUALIZACIÓN DE LA TESIS Y ESTRUCTURACIÓN DE LA MEMORIA.....	xiii
ABSTRACT, CONTEXTUALIZATION AND STRUCTURE OF THE THESIS.....	xxi

OBJETIVOS

OBJETIVOS.....	xxix
OBJECTIVES.....	xxxiii

REACIONES ENZIMÁTICAS

1.1. INTRODUCCIÓN.....	3
1.2. DEFINICIÓN DE ENZIMA.....	3
1.3. COMPOSICIÓN Y ESTRUCTURA DE LOS ENZIMAS.....	4
1.4. CLASIFICACIÓN.....	7
1.5. CATÁLISIS ENZIMÁTICA. ECUACIÓN DE MICHAELIS-MENTEN.....	9
1.5.1. Inhibición enzimática.....	12
1.5.2. Reacciones multisustrato.....	14
1.5.3. Efecto del pH, y la temperatura sobre la actividad enzimática.....	15
1.6. ENZIMAS COMERCIALES.....	17
1.6.1. Introducción.....	17
1.6.2. Uso de enzimas en la industria.....	18
1.6.3. Aplicaciones médicas y farmacéuticas de los enzimas.....	18
1.6.4. Aplicaciones analíticas de los enzimas.....	20
1.6.5. Uso de enzimas de grupo para la determinación de mezclas de analitos en matrices complejas... 21	
1.6.6. Otros campos de aplicación de los enzimas.....	22
1.6.7. Procedencia de los enzimas comerciales.....	22

MONITORIZACIÓN DE REACCIONES ENZIMÁTICAS

2.1. INTRODUCCIÓN.....	27
2.2. ADQUISICIÓN DE DATOS CINÉTICOS. INSTRUMENTACIÓN.....	28

2.2.1.	ESPECTROSCOPIA UV-VISIBLE-DAD ACOPLADA A MÓDULOS DE FLUJO INTERRUMPIDO.....	28
2.2.2.	FLUORESCENCIA MULTIDIMENSIONAL.....	31
2.3.	TRATAMIENTO DE LOS DATOS REGISTRADOS.....	33
2.3.1.	MÉTODOS GRÁFICOS DE CÁLCULO DE CONSTANTES.....	33
2.3.2.	METODOLOGÍAS DE MODELADO DURO (<i>HARD-MODELLING METHODS</i>)	35
2.3.3.	METODOLOGÍAS DE MODELADO BLANDO (<i>SOFT-MODELLING METHODS</i>). INTRODUCCIÓN AL ANÁLISIS MULTIVARIABLE.....	37

LA QUIMIOMETRIA COMO HERRAMIENTA PARA MODELAR REACCIONES ENZIMÁTICAS

3.1.	INTRODUCCIÓN.....	41
3.2.	DEFINICIÓN. CLASIFICACIÓN DE LOS MÉTODOS QUIMIOMÉTRICOS.....	42
3.3.	ANÁLISIS QUIMIOMÉTRICO MULTIDIMENSIONAL DE DATOS EVOLUTIVOS MULTIVARIABLE.....	43
3.3.1.	Datos evolutivos. Datos de segundo y tercer orden.....	45
3.3.2.	Conceptos de linearidad y rango de una matriz multidimensional.	45
3.3.2.1.	Linealidad, bilinearidad y trilinealidad.....	45
3.3.2.2.	Rango matemático–rango químico. Pseudo-rango. Fuentes químicas de variación.	47
3.4.	TRANSFORMACIÓN DE UNA MATRIZ TRIDIMENSIONAL EN BIDIMENSIONAL. TÉCNICAS <i>UNFOLDING</i>.....	49
3.4.1.	<i>Unfolding</i> -PCA y <i>Unfolding</i> -PLS.....	50
3.5.	3W-PCAY 3W-PLS.....	52
3.5.1.	3W-PCA.....	52
3.5.2.	3W-PLS.....	54
3.6.	ANÁLISIS PARALELO DE FACTORES. <i>PARALLEL FACTOR ANALYSIS. PARAFAC</i>.....	55
3.6.1.	Introducción.....	55
3.6.2.	Aplicación de PARAFAC.....	57
3.6.2.1.	Número de componentes en PARAFAC.....	58
3.6.2.2.	Iniciando PARAFAC. Estimaciones iniciales y restricciones.....	59
3.6.2.3.	Algoritmo ALS en PARAFAC.....	59
3.7.	RESOLUCIÓN MULTIVARIABLE DE CURVAS. <i>MULTIVARIATE CURVE RESOLUTION (MCR-ALS)</i>.....	60
3.7.1.	Introducción.....	60
3.7.2.	Paso 1: Determinación del número de componentes. Sistemas deficientes de rango.....	62
3.7.2.1.	Análisis en Componentes Principales. <i>Principal Components Analysis</i> . PCA.....	64

3.7.2.2.	Descomposición en valores singulares. <i>Singular Value Decomposition</i> . SVD.....	65
3.7.2.3.	Análisis de Factores Evolutivos. <i>Evolving Factor Analysis</i> , EFA.....	67
3.7.2.4.	Sistemas deficientes de rango.....	68
3.7.3.	Paso 2: Necesidad de estimaciones iniciales.....	69
3.7.4.	Paso 3: Optimización ALS. Ambigüedades.....	70
3.7.4.1.	Minimización de las ambigüedades. Uso de restricciones.....	71
3.7.4.2.	Restricciones de modelo. <i>Hard-Soft</i> MCR-ALS HS-MCR-ALS.....	73
3.7.5.	Criterio de convergencia y parámetros de calidad del ajuste en MCR-ALS.....	75
3.7.6.	MCR-ALS aplicado a sistemas tridimensionales. 3W-MCR-ALS.....	76
3.7.6.1.	Encadenamiento por columnas.....	77
3.7.6.2.	Encadenamiento por filas.....	78
3.7.7.	Metodología de trabajo en 3W-MCR-ALS.....	79
3.8.	IMPLEMENTACIÓN ON-LINE DE PARAFAC PARA EL CONTROL EN TIEMPO REAL DE PROCESOS.....	79
3.8.1.	Introducción.....	79
3.8.2.	Construcción del modelo PARAFAC de calibración.....	81
3.8.3.	Monitorización del bioproceso en tiempo real. Proyección de una muestra en el modelo PARAFAC de calibración.....	83
3.8.4.	Estadístico residual, Q_n , como parámetro de control.....	84
3.9.	PROGRAMAS DISPONIBLES EN EL ANÁLISIS MULTIVARIABLE MULTIDIMENSIONAL DE DATOS EVOLUTIVOS.....	86
3.9.1.	<i>Unfold</i> PCA y PLS.....	87
3.9.2.	3W-PCA y 3W-PLS.....	90
3.9.3.	SVD, EFA.....	90
3.9.4.	MCR-ALS y 3W-MCR-ALS.....	91
3.9.5.	HS-MCR-ALS.....	92
3.9.6.	PARAFAC.....	93

METODOLOGÍA

PARTE I: DETERMINACIÓN SIMULTÁNEA DE HIPOXANTINA, XANTINA Y ÁCIDO ÚRICO EN ORINA HUMANA.		
4.1.	REACTIVOS	97
4.2.	CONDICIONES ÓPTIMAS DE CATÁLISIS ENZIMÁTICA.....	98
4.3.	APARATOS Y PROGRAMAS.....	98
4.4.	PROCEDIMIENTO EXPERIMENTAL PARA LA MONITORIZACIÓN DE LAS CINÉTICAS.....	98
4.5.	TRATAMIENTO DE LOS DATOS UV-VISIBLE.....	100

**PARTE II: CONTROL EN TIEMPO REAL DE PROCESOS DE PRODUCCIÓN DE ENZIMAS.
PRODUCCIÓN DE LIPASAS.**

4.6.	REACTIVOS.....	100
4.7.	CONDICIONES DEL BIOPROCESO.....	101
4.8.	MEDIDAS DE FLUORESCENCIA MULTICANAL ON-LINE.....	102
4.9.	TRATAMIENTO DE DATOS DE FLUORESCENCIA.....	103
4.10.	MEDIDAS <i>OFF-LINE</i>	103

**PARTE III: ESTUDIO ESPECTROFOTOMÉTRICO DE LOS EQUILIBRIOS ÁCIDO BASE DEL
ÁCIDO 8-HIDROXIQUINOLINA-5-SULFÓNICO. INTRODUCCIÓN AL MCR-ALS.**

4.11.	REACTIVOS E INSTRUMENTAL.....	105
4.12.	MONTAJE DEL EXPERIMENTO.....	106
4.12.1.	Primera sesión: Valoración ácido base del 8HQS.....	106
4.12.2.	Segunda sesión: Tratamiento quimiométrico de los espectros registrados.....	107

RESULTADOS

**PARTE I: DETERMINACIÓN SIMULTÁNEA DE HIPOXANTINA Y XANTINA EN ORINA
HUMANA.**

5.1.	INTRODUCCIÓN.....	111
5.2.	SISTEMA ENZIMÁTICO.....	112
5.3.	DETERMINACIÓN SIMULTÁNEA DE HIPOXANTINA Y XANTINA EN ORINA HUMANA A TRAVÉS DE 3W-PCA Y 3W-PLS. COMPARACIÓN CON TÉCNICAS DE DESDOBLAMIENTO <i>unfolding-PCA</i> Y <i>unfolding-PLS</i>	115
5.3.1.	Objetivos.....	115
5.3.2.	Diseño del conjunto de calibración, validación y muestras de orina.....	115
5.3.3.	Procesado de los datos.....	118
5.3.4.	Resultados y discusión.....	120
5.3.4.1.	Comparación entre las muestras de calibración, validación y las muestras de orina.....	120
5.3.4.2.	Determinación de hipoxantina y xantina.....	123
5.3.5.	Conclusiones a este primer trabajo.....	126
5.4.	HS-MCR-ALS APLICADO AL ESTUDIO Y MONITORIZACIÓN DE SISTEMAS ENZIMÁTICOS EN FLUIDOS BIOLÓGICOS.....	129
5.4.1.	Objetivos.....	129
5.4.2.	El sistema enzimático: Descripción y monitorización experimental.....	129
5.4.2.1.	Descripción detallada del sistema enzimático. Modelo de Michaelis-Menten.....	129
5.4.3.	Conjuntos de datos.....	132

5.4.4.	Algoritmo HS-MCR-ALS. Introducción de la restricción de modelo enzimático en MCR-ALS	132
5.4.5.	Resultados y discusión.....	134
5.4.5.1.	Estrategia para la elucidación del modelo enzimático.....	134
5.4.5.2.	Modelado del proceso en presencia de fluidos biológicos.....	139
5.4.6.	Conclusiones a este trabajo.....	143
5.5.	HS-MCR-ALS APLICADO A LA DETERMINACIÓN SIMULTÁNEA DE OXIPURINAS Y ÁCIDO ÚRICO EN ORINA HUMANA.....	145
5.5.1.	Objetivos.....	145
5.5.2.	Conjunto de datos.....	145
5.5.3.	Tratamiento de los datos.....	147
5.5.3.1.	Aplicación de HS-MCR-ALS para la cuantificación de los analitos.....	147
5.5.3.2.	Validación de los resultados cuantitativos.....	150
5.5.4.	Resultados y discusión.....	150
5.5.5.	Comparación de los resultados obtenidos con distintas metodologías.....	156
5.5.6.	Conclusiones a este trabajo.....	157
PARTE II: CONTROL EN TIEMPO REAL DE PROCESOS DE PRODUCCIÓN DE ENZIMAS.		
PRODUCCIÓN DE LIPASAS		
5.6.	INTRODUCCIÓN.....	159
5.7.	OBJETIVOS.....	160
5.8.	RESULTADOS Y DISCUSIÓN.....	160
5.8.1.	Modelo PARAFAC de calibración.....	160
5.8.2.	Monitorización en tiempo real de un bioproceso.....	162
5.9.	CONCLUSIONES DE ESTE TRABAJO.....	165
5.10.	MONITORIZACIÓN USANDO GLICEROL COMO SUSTRATO.....	166
PARTE III: ESTUDIO ESPECTROFOTOMÉTRICO DE LOS EQUILIBRIOS ÁCIDO BASE DEL ÁCIDO 8-HIDROXIQUINOLINA-5-SULFÓNICO. INTRODUCCIÓN AL MCR-ALS.		
5.11.	INTRODUCCIÓN.....	167
5.12.	OBJETIVOS.....	168
5.13.	RESULTADOS Y DISCUSIÓN.....	169
5.13.1.	Aplicación de MCR-ALS.....	169
5.13.2.	Resultados y discusión.....	169
5.14.	CONCLUSIONES A ESTE TRABAJO.....	171

CONCLUSIONES GENERALES

6.1.	CONCLUSIONES.....	173
6.2.	CONCLUSSIONS.....	179
7.	REFERENCIAS / REFERENCES.....	185

ANEXOS / ANNEXES

A1.	ARTICULOS PRODUCIDOS EN LA TESIS. PAPERS.....	195
A.1.1.	THREE-WAY PARTIAL LEAST-SQUARES REGRESSION FOR THE SIMULTANEOUS KINETIC-ENZYMATIC DETERMINATION OF XANTHINE AND HYPOXANTHINE IN HUMAN URINE.....	197
A.1.2.	A MIXED HARD- AND SOFT-MODELING APPROACH TO STUDY AND MONITOR ENZYMATIC SYSTEMS IN BIOLOGICAL FLUIDS.....	209
A.1.3.	A MIXED HARD- AND SOFT-MODELING APPROACH FOR THE QUANTITATIVE DETERMINATION OF OXIPURINES AND URIC ACID IN HUMAN URINE.....	221
A.1.4.	AN INTRODUCTION TO MULTIVARIATE CURVE RESOLUTION-ALTERNATING LEAST SQUARES: SPECTROPHOTOMETRIC STUDY OF THE ACID-BASE EQUILIBRIA OF 8-HYDROXYQUINOLINE-5-SULFONIC ACID.....	233
A.1.5.	ON-LINE PARALLEL FACTOR ANALYSIS. A STEP FORWARD IN THE MONITORING OF BIOPROCESSES IN REAL TIME.....	241
A2.	APORTACIONES A CONGRESOS Y REUNIONES NACIONALES E INTERNACIONALES. CONFERENCES AND CONGRESSES.....	271

ÍNDICE DE TABLAS

Capítulo 1

1.1.	Evolución histórica en la investigación enzimática.....	6
1.2.	Clasificación de los enzimas.....	8
1.3.	Principales campos de aplicación de los enzimas en la industria.....	19

Capítulo 3

3.1.	Clasificación de los métodos analíticos en función de los datos disponibles por muestra.....	43
3.2.	Comparación de los diferentes algoritmos en función de su aplicabilidad.....	87
3.3.	Comparación de algoritmos para el tratamiento de datos multivariable multidimensionales evolutivos.....	89

Capítulo 5

5.1.	Composición de las muestras de calibración, validación y orina dopada ($\mu\text{mol mL}^{-1}$ de orina)..	117
5.2.	Porcentaje de varianza explicada por el análisis 3W-PCA de cada uno de los modos.....	120
5.3.	Resultados para la determinación de hipoxantina.....	123
5.4.	Resultados para la determinación de xantina.....	124
5.5.	Estadísticos de la regresión del modelo 3W-PLS2 para la cuantificación de hipoxantina y xantina.....	126
5.6.	Concentraciones iniciales ($\mu\text{mol L}^{-1}$) para la hipoxantina, xantina, ácido úrico y porcentaje de orina en los experimentos usados para cada conjunto de datos A y D.....	131
5.7.	Características de los métodos aplicados.....	137
5.8.	Constantes enzimáticas obtenidas aplicando las diferentes metodologías	139
5.9.	Composición de los dos estándares (D1 y D2) y de las 19 mezclas (D3-D21) que contenían orina (Expresadas en % de volumen).....	146
5.10.	Aplicación de las metodologías propuestas para el análisis del sistema enzimático.....	151
5.11.	Comparación de los parámetros estadísticos obtenidos	155
5.12.	Comparación de los resultados obtenidos con otros trabajos.....	157
5.13.	Resultados obtenidos en la aplicación del MCR-ALS a la valoración ácido base del 8HQS.....	170

ÍNDICE DE FIGURAS

Capítulo 1

1.1.	Estructura tridimensional de la xantina oxidasa (E.C. 1.17.3.2) [Web 1.4].....	5
1.2.	Representación gráfica del modelo de una reacción de catálisis enzimática.....	10
1.3.	Hipótesis del estado estacionario postulada por Briggs-Haldane.....	10
1.4.	Representación gráfica de la ecuación de Michaelis-Menten.....	11
1.5.	Tipos más comunes de inhibición enzimática y su representación gráfica.....	13
1.6.	Ejemplo de reacción multisustrato siguiendo un mecanismo ping-pong.....	15
1.7.	Efecto del pH y la temperatura en la actividad enzimática.....	16
1.8.	Principales usos comerciales de los enzimas.....	18

Capítulo 2

2.1.	Esquema del sistema de flujo interrumpido (<i>stopped-flow</i>) acoplado al espectrofotómetro.....	30
2.2.	Representación gráfica del registro espectral con el tiempo que se puede obtener con un espectrofotómetro UV-Vis-DAD acoplado a un módulo de flujo interrumpido (<i>stopped-flow</i>).....	30
2.3.	Esquema del equipo formado por el fluorímetro multidimensional acoplado a un bioreactor estanco.....	31
2.4.	Representación gráfica del registro espectral con el tiempo que se obtiene con un fluorímetro multidimensional acoplado a un bioreactor.....	32
2.5.	Métodos de linealización de la ecuación de Michaelis-Menten.....	34
2.6.	Metodología de trabajo de modelado duro (<i>Hard Modelling</i>).....	36

Capítulo 3

3.1.	Organización de datos de segundo y tercer orden.....	44
3.2.	Desdoblamiento de una matriz bidimensional en un tensor de primer orden, o vector.....	50
3.3.	Obtención de una matriz bidimensional desdoblada a partir de una matriz tridimensional.....	50
3.4.	Representación gráfica de expresión matricial de una descomposición PCA.....	51
3.5.	Aplicación del 3W-PCA a la matriz \mathbf{M}	53
3.6.	Descomposición 3W-PLS de una matriz \mathbf{M} ($I \times J \times K$).....	54
3.7.	Descomposición PARAFAC de una matriz \mathbf{M} ($I \times J \times K$).....	56
3.8.	Resolución Multivariable de Curvas MCR-ALS aplicada a la descomposición de \mathbf{M}	60
3.9.	Pasos en una descomposición MCR-ALS de la matriz \mathbf{M} ($J \times K$).....	61
3.10.	Evolución del espectro ácido-base en un sistema diprótico [Rodríguez-Rodríguez 2007].....	63
3.11.	Perfiles de <i>scores</i> y <i>loadings</i> y porcentaje de varianza explicada por factor.....	65

3.12.	Análisis SVD del ejemplo propuesto.....	66
3.13.	Análisis EFA del ejemplo propuesto.....	68
3.14.	Sistema deficiente de rango.....	69
3.15.	Restricciones naturales.....	73
3.16.	Aplicación del modelo enzimático a los perfiles de concentración obtenidos en un análisis MCR-ALS.....	74
3.17.	Encadenamiento 3W-MCR-ALS por columna.....	78
3.18.	Encadenamiento 3W-MCR-ALS por fila.....	78
3.19.	a) Creación del modelo PARAFAC de calibración con un bioproceso monitorizado en condiciones normales de operación (NOC). B) Control a tiempo real de un nuevo bioproceso.....	82
3.20.	Estadístico residual para una muestra P monitorizada a un tiempo determinado.....	84
3.21.	Diagrama de control de un nuevo bioproceso.....	85

Capítulo 4

4.1.	a) Sistema de flujo interrumpido (<i>stopped-flow</i>). b) Espectrofotómetro UV-Vis-DAD acoplado al módulo de flujo interrumpido y al módulo <i>Peltier</i>	99
4.2.	Evoluciones cinéticas obtenidas para tres de las muestras de calibración a) C2, cinética de hipoxantina; b) C3, cinética de xantina; c) C4, cinética del ácido úrico.....	100
4.3.	a) Reactor esterilizado donde se llevaron a cabo los bioprocesos. b) Detalle de la sonda de fluorescencia multicanal. c) Acoplamiento de la sonda multicanal con el bioreactor.....	102
4.4.	Evolución de la señal de fluorescencia con el tiempo para el primer bioproceso.....	104
4.5.	Diagrama del montaje de la valoración ácido-base.....	107

Capítulo 5

5.1.	Proceso enzimático en dos etapas de la catálisis de la hipoxantina (H) por la xantina oxidasa (XO) y espectros UV-Vis de la hipoxantina (H, azul), xantina (X, rojo) y el ácido úrico (U, verde)...	113
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).....	116
5.3.	Gráficos de <i>scores</i> obtenidos para los modos de a) concentración, b) cinético y c) espectral.....	121
5.4.	Gráfico de <i>scores</i> del primer factor frente al tercer factor para el modo de concentración.....	122
5.5.	Diferencia espectral entre muestras que contenían orina (rojo) y muestras sin orina (azul).....	122
5.6.	<i>Loadings</i> obtenidos en el análisis 3W-PLS2	125
5.7.	a) Monitorización UV-Visible de una muestra que contiene 30 $\mu\text{mol L}^{-1}$ de hipoxantina y xantina, respectivamente (experimento A ₃). b) Evolución de la absorbancia en el máximo de absorbancia del ácido úrico (290 nm.).....	135
5.8.	Perfiles cinéticos y espectrales obtenidos para el experimento A ₃	136

5.9.	Perfiles cinéticos y espectro puros obtenidos en el análisis del conjunto de datos A por medio del análisis <i>hard</i> aplicando el nuevo modelo enzimático propuesto.....	138
5.10.	Análisis de factores evolutivos, EFA, para los experimentos D ₁ y D ₃	140
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.....	142
5.12.	Concatenación en columna de las dos muestras estándar y las 19 muestras de orina para la aplicación de HS-MCR-ALS.....	147
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).....	148
5.14.	Perfiles de concentración (C3-C20) y espectros puros (S ^T) obtenidos en el experimento 2 para de las muestras con orina.....	154
5.15.	Rectas de correlación entre los valores predichos con HS-MCR-ALS y los valores de referencia para los tres analitos.....	156
5.16.	<i>Loadings</i> espectrales de excitación y emisión obtenidos para el modelo de calibración.....	161
5.17.	<i>Loadings</i> de concentración obtenidos para el modelo de calibración.....	162
5.18.	<i>Loadings</i> de concentración y valores de <i>Q_n</i> obtenidos con PARAFAC en el control a tiempo real del Segundo bioproceso.....	163
5.19.	Comparación de los <i>loadings</i> de concentración obtenidos en el bioproceso controlado a tiempo real con las medidas <i>off-line</i>	164
5.20.	Equilibrio ácido-base del 8HQS. pK _{a1} = 3.80 y pK _{a2} = 8.20 [Smith 1975].....	168
5.21.	Perfiles de concentración y espectrales obtenidos mediante MCR-ALS.....	170

LISTADO DE ABREVIATURAS

3W-PCA *3 Way-Principal Component Analysis* Análisis en Componentes Principales tridimensional

3W-PLS *3 Way-Partial Least Square Regression* Regresión por Mínimos cuadrados Parciales tridimensional

A, B, C Matrices de loadings del primer, segundo y tercer modo en PARAFAC, respectivamente.

a, b, c loadings del primer, segundo y tercer modo en PARAFAC, respectivamente.

ADN Ácido desoxirribonucleico

C Matriz de perfiles de concentración en MCR-ALS y HS-MCR-ALS

c índice de perfiles de concentración en MCR-ALS y HS-MCR-ALS

CANDECOMP *Canonical Decomposition* Descomposición Canónica

λ Longitud de onda

DAD-UV-VIS *Diode Array Detector Ultraviolet Visible* Espectrofotómetro Ultravioleta Visible de Matriz de Diodos

E Enzima

ELISA *Enzyme-Linked Immuno Sorbent Assay*

EFA *Evolving Factor Analysis* Análisis de Factores Evolutivos

ES Complejo enzima-sustrato.

EI Complejo enzima-inhibidor.

ESI Complejo enzima-sustrato-inhibidor.

F Número de Componentes Principales, factores o variables latentes

f índice de Componentes Principales, factores o variables latentes

H Hipoxantina

HS-MCR-ALS *Hard-Soft Multivariate Curve Resolution-Alternating Least Squares* Resolución

Multivariable de Curvas con el algoritmo de Mínimos Cuadrados Alternos aplicado al modelado mixto Duro-Blando

i, j, k, l, ... Índices

I, J, K, L, ... Máximo de los índices *i, j, k, l, ...*

ILS *Inverse Least Squares Regression* Regresión Inversa por Mínimos Cuadrados

ITTFA *Iterative Target Factor Analysis*

isp Matriz de correspondencia entre species.

K_M Constante de Michaelis-Menten

L *Loading*

m vector. Tensor de primer orden

M matriz bidimensional. Tensor de segundo orden

M matriz tridimensional. Tensor de tercer orden

MCR-ALS *Multivariate Curve Resolution-Alternating Least Squares* Resolución Multivariable de Curvas con el algoritmo de Mínimos Cuadrados Alternos

MIR *Mid Infra Red Spectroscopy* Espectroscopia en el Infrarrojo Medio

MLR *Multi Linear Regression* Regresión Linear Múltiple

MSPC *Multivariate Statistical Process Control* Estadística Multivariable del Control de Proceso

NIPALS *Nonlinear Iterative Partial Least Squares*

NIR *Near Infra Red Spectroscopy* Espectroscopia en el Infrarrojo Cercano

NOC *normal operating conditions* condiciones normales de trabajo

O Orina

P Producto de reacción y matriz de *loadings*

p Índice de la matriz de *loadings*

PARAFAC *Parallel Factor Analysis* Análisis Paralelo de Factores

PARAFAC2 *Parallel Factor Analysis 2* Análisis Paralelo de Factores 2

PAT *Process Analytical Technology* Tecnología Analítica de Proceso

PCA *Principal Component Analysis* Análisis en Componentes Principales

PLS-R *Partial Least Square Regression* Regresión por Mínimos cuadrados Parciales

R Número de componentes en PARAFAC, MCR-ALS y HS-MCR-ALS

r índice de número de componentes en PARAFAC, MCR-ALS y HS-MCR-ALS

SVD *Singular Value Decomposition* Descomposición en Valores Singulares

SIMPLISMA *Simple-to-use interactive self-modeling mixture analysis*

T Matriz de *scores*

t Índice de la matriz de *scores*

U Ácido úrico

U-PCA *Unfolding* Análisis en Componentes Principales desdoblado

U-PLS *Unfolding Partial Least Square Regression* Regresión por Mínimos cuadrados Parciales

V_{max} Velocidad máxima de catálisis

x parámetros

X Xantina

XO Xantina oxidasa

y variables

**CONTEXTUALIZACIÓN DE LA
TESIS DOCTORAL Y ESTRUCTURA
DE LA MEMORIA**

CONTEXTUALIZACIÓN

Tal vez, la mejor manera de presentar este trabajo es comenzar diciendo que los enzimas forman parte esencial de la vida. Los enzimas catalizan miles de reacciones necesarias para el correcto funcionamiento de nuestro metabolismo y de los seres vivos en general. Controlan la actividad celular, la digestión, los ciclos metabólicos, la respiración, etc. Por ser el aspecto más importante, la bioquímica de los enzimas es ampliamente estudiada y, hoy en día, se conoce perfectamente muchas de las rutas catalíticas que se producen en los seres vivos.

Pero la importancia de los enzimas trasciende mucho más allá del aspecto puramente biológico. Se sabe que el hombre ha utilizado los enzimas en procesos de fermentación desde el comienzo de la humanidad. Probablemente los primeros procesos de fermentación fueron la elaboración de pan, queso, cerveza y vino. Estos alimentos se producían sin tener consciencia de que lo que hacía posible su elaboración eran las reacciones enzimáticas producidas por los microorganismos presentes.

A medida que la tecnología fue avanzando, también lo hizo el estudio y el conocimiento de estos procesos. Se puede decir que la explosión del uso de enzimas en procesos industriales se produjo a mediados del siglo XX. Se vio que el uso de enzimas en diferentes procesos de manufacturación mejoraba el rendimiento de las reacciones y la calidad del producto final. De esta manera, los enzimas se fueron introduciendo en diversos campos de la industria, de los cuales los más importantes son el campo agroalimentario, textil, papel, detergentes y cosméticos. Hoy en día, el trabajo en estas industrias se centra en conseguir mejoras en:

- La calidad del producto acabado.
- El rendimiento de las reacciones enzimáticas.
- La búsqueda de nuevas enzimas y nuevos procesos enzimáticos.
- La búsqueda de tecnologías de seguimiento a tiempo real del proceso

Otro campo donde los enzimas tienen un papel relevante es el analítico. Cada vez son más utilizadas las reacciones enzimáticas para el análisis químico de diferentes analitos, sobre todo en matrices biológicas. Puntos clave de esta aplicación son:

- La gran selectividad de los enzimas hacia diversos sustratos o determinados grupos funcionales.
- El desarrollo de metodologías relativamente sencillas de monitorización de la reacción enzimática.

El punto de unión entre la aplicación industrial y analítica de los enzimas lo encontramos en la necesidad de monitorizar y controlar las diferentes reacciones a tiempo real y usando metodologías de control relativamente sencillas y de fácil manejo por el usuario. En este aspecto están empezando a tener importancia la introducción de sondas dentro de los biorreactores o en el análisis *on-line* que nos aporta gran información, generalmente de carácter espectrofotométrico y fluorimétrico, sobre la evolución de la reacción.

Todo este volumen de datos registrados, generalmente de carácter multivariable, necesitan ser tratado para extraer la información relevante del sistema. En este punto es

donde inciden los métodos quimiométricos de análisis de datos. Podríamos decir, sin lugar a dudas, que una buena monitorización de un proceso tiene que ir acompañada de un adecuado tratamiento quimiométrico de los datos para la obtención de los mejores resultados. Estos procedimientos, de carácter matemático, tienen que cumplir dos requisitos indispensables:

- Ser lo suficientemente robustas para proporcionar “información real” en cada situación.
- Estar implementadas en programas de control de fácil manejo por el operario.

En esta tesis doctoral se ha abordado el desarrollo y aplicación de métodos quimiométricos multidimensionales que permitan extraer el máximo de información en sistemas enzimáticos, siguiendo la evolución de la cinética enzimática con el tiempo de manera espectrofotométrica. Se plantea el uso de diferentes metodologías quimiométricas para cada caso estudiado y se estudia la capacidad de cada una de ellas para la obtención de la información deseada de cada sistema.

ESTRUCTURA DE LA MEMORIA

Para facilitar su lectura, esta memoria se ha estructurado en los siguientes capítulos:

- El **capítulo cero** expone y define los objetivos que se quieren conseguir.
- El **primer capítulo** trata de dar una perspectiva general sobre los enzimas, la catálisis enzimática y sobre la aplicación de los enzimas en diferentes aspectos de la vida cotidiana y en las diferentes industrias.
- El **segundo capítulo** se centra en las nuevas técnicas de monitorización de reacciones enzimáticas y en las nuevas metodologías de adquisición de datos multivariable multidimensionales que se empiezan a implantar, tanto en la industria como en los laboratorios de química analítica.

· En el **tercer capítulo** se presentan los algoritmos que existen para el análisis de datos multivariable multidimensionales. Esta presentación se hace de una manera totalmente aplicada, intentado ofrecer una visión directa de la información que se puede extraer de cada uno de ellos. Se compararán las ventajas e inconvenientes que presentan, así como los diferentes programas disponibles.

· El **cuarto capítulo** contiene un resumen global de las metodologías empleadas en los diferentes trabajos. Los trabajos presentados se agrupan en tres partes. En la primera se presentan tres trabajos relacionados con el uso de los enzimas como elementos diferenciadores en una determinación simultánea de varios analitos en muestras biológicas. En la segunda parte se presenta un trabajo dirigido al control en tiempo real de procesos de producción de enzimas. En la tercera parte de esta memoria se desarrolla una práctica de laboratorio dedicada al estudio del MCR-ALS por parte de los estudiantes de cursos superiores de la licenciatura de Ciencias Químicas. Este trabajo no está directamente ligado al estudio de sistemas enzimáticos, pero se ha incluido en esta memoria como un ejemplo detallado de cómo se puede implementar el aprendizaje de metodologías quimiométricas en los planes de estudio mediante la realización de una práctica sencilla. En este trabajo se describe exhaustivamente la metodología de trabajo para la aplicación del MCR-ALS.

· El **quinto capítulo** recoge los resultados más relevantes obtenidos en los trabajos realizados.

· El **capítulo seis** engloba las conclusiones generales obtenidas en esta memoria.

· El **séptimo capítulo** recoge las referencias, tanto bibliográficas como vía Web, que se han utilizado para la elaboración de la memoria. La manera de consultar las referencias es la siguiente:

[apellido del primer autor año de publicación] Inicial del nombre y apellido de todos los autores, *revista*, número de la revista (año de publicación) páginas de la revista

[apellido del primer autor año de publicación] Inicial del nombre y apellido de todos los autores, “Título del libro”, número de edición, Editorial, año.

[Web capítulo.número de referencia] enlace Web.

· Por último los **anexos** recogen las aportaciones a congresos y reuniones nacionales e internacionales y los trabajos producidos en esta tesis.

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

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

A mixed hard- 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

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

On-Line Parallel Factor Analysis. A stepforward in the monitoring of bioprocesses in real time. José Manuel Amigo, Anna Surribas, Jordi Coello, Santiago MasPOCH, José Luís Montesinos, Francisco Valero. Send to *Talanta*.

**ABSTRACT,
CONTEXTUALIZATION AND
STRUCTURE OF THE THESIS**

ABSTRACT AND CONTEXTUALIZATION

Enzymes are an essential part of life. They catalyze thousand of reactions that are necessary for the correct development of the metabolic paths. They control the cellular activity, digestion, etc. The biological aspects of enzymes have been widely studied and nowadays, a wide range of catalytic paths are well established.

Nevertheless, the importance of enzymes extends from the biological field. It is known that enzymes have been used in fermentation processes since the beginning of humanity. Probably, the first fermentation processes were those related to the manufacturing of bread, wine, beer and cheese. All these foods were produced without knowing the involved enzymatic reactions produced by microorganisms, but as technology advanced, the knowledge of the fermentation processes was increasing in order to improve the quality of the product. The big explosion of the industrial application of enzymes was in the middle of XX century. It was observed that the use of enzymes in different industrial processes improved the yield of a great amount of

reactions and the quality of the final product. In this way, enzymes were introduced in different fields of the industry (animal feed, paper, detergents, textile, cosmetics...). Nowadays, these industries are trying to improve on the enzymatic processes in several aspects:

- improving the quality of the products.
- improving the yield of enzymatic reactions.
- finding new enzymes.
- improving the technologies for monitoring and controlling fermentation processes in real time.

Another field where enzymes begin to be important is in analytical chemistry. Enzymatic reactions become very useful in the chemical analysis of mixtures of different analytes in biological matrices. Two main aspects are essential in this type of analysis:

- The high selectivity of enzymes with several substrates.
- The development of easy-to-use methodologies for monitoring enzymatic reactions.

The inclusion of probes in the reactors or the on-line analyses of reactions are two of the most commonly used alternatives for monitoring enzymatic reactions. Generally, these two methods record the spectral evolution of the system with time. For this question, it is necessary to develop different methods to extract the relevant information for the system. It is in this point where Chemometrics takes importance. It has to be pointed out that a good monitoring of a process is directly linked to an adequate chemometric treatment of the data to obtain the best results.

This thesis is focused in the development and application of multidimensional methodologies that allow extracting the maximum of information in enzymatic systems, following photometrically the enzymatic evolution with time.

STRUCTURE OF THE THESIS

This thesis has been divided into the following chapters:

- **Chapter zero** exposes the objectives of the dissertation.

- The **first chapter** offers a general perspective about enzymes, enzymatic catalysis and about the application of enzymes in different aspects of life and industry.

- The **second chapter** is focused in the new multivariate multidimensional methodologies.

- **Chapter three** presents the most relevant algorithms for the analysis of multidimensional data. This chapter is focused in the practical application of these algorithms to show the kind of information that can be obtained. An exhaustive comparison between different software will be developed.

- The **fourth chapter** shows a scheme of the working methodologies used in the different works presented in this thesis. This section is divided into three parts.
Part I: Three works related with the use of enzymes in the determination of mixtures of several analytes in biological samples are presented.
Part II: A work focused in real time control of a reaction for enzymes production is presented.
Part III: In this part of the thesis, a laboratory experiment directed to students of the last courses of Chemistry Degree is presented. The objective is the divulgation of curve resolution methods. Although this work is not directly related with the study of enzymatic systems, it is a detailed example of the application of Multivariate curve resolution with alternating least squares (MCR-ALS).

· The **fifth chapter** shows the results obtained in each one of the parts described in chapter four.

· **Chapter six** resumes all the conclusions obtained in this thesis.

· The **seventh chapter** includes all the references cited in this dissertation using the format:

[Surname of the first author year of publication] First letter of the name. surname of all authors, *journal*, number (year of publication) pages.

[Surname of the first author year of publication] First letter of the name. surname of all authors, “Title of the book”, edition number, editor, year of publication.

[Web chapter.reference number] Web site

· Finally, the **annexes** compile the different posters and oral expositions presented in national and international congresses and conferences and the papers produced in the thesis.

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

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

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

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.* Accepted for publication, May 2007

On-Line Parallel Factor Analysis. A stepforward in the monitoring of bioprocesses in real time. José Manuel Amigo, Anna Surribas, Jordi Coello, Santiago MasPOCH, José Luís Montesinos, Francisco Valero. Submitted to *Talanta*, May 2007.

OBJETIVOS

El objetivo general de esta tesis doctoral es la **aplicación de metodologías quimiométricas multidimensionales al análisis de reacciones enzimáticas**. Las reacciones enzimáticas estudiadas tienen aplicación en los campos del análisis químico en matrices biológicas y en los procesos de fermentación a escala industrial. De una manera más concreta, podemos desglosar los objetivos en los siguientes:

- Estudio del mecanismo de reacción en matrices biológicas y cálculo de las constantes enzimáticas en reacciones complejas. Además, se examinará la influencia del medio de reacción en el mecanismo enzimático. Las reacciones en estudio serán la catálisis de mezclas de hipoxantina, xantina y ácido úrico con xantina oxidasa. Estas reacciones se estudiarán en un medio controlado y en orina humana.

- Determinación cuantitativa de los analitos anteriormente mencionados en orina humana a través de su diferente comportamiento cinético con la xantina oxidasa. Se comparan diferentes algoritmos multidimensionales aplicados a la determinación cuantitativa y se estudian las ventajas e inconvenientes de su aplicación.

· Establecimiento de una metodología para el control a tiempo real de procesos de fermentación. En este sentido se aplicarán técnicas de factorización multidimensional a datos espectrofluorimétricos para crear un modelo de calibración robusto y se estudiará la posibilidad de su implementación para el control a tiempo real. El sistema estudiado es la producción de enzimas lipasas por el microorganismo *Pichia pastoris*.

Un objetivo ligado con esta tesis doctoral, no relacionado con la monitorización de reacciones enzimáticas, es el desarrollo de una práctica de laboratorio dirigida a estudiantes de cursos superiores de la licenciatura de Ciencias Químicas para la introducción a las metodologías de resolución de curvas y a la quimiometría en general. En este sentido se desarrollará una práctica completa para la aplicación del MCR en un sistema evolutivo sencillo como es el estudio ácido-base del ácido 8-hidroxiquinolina-5-sulfónico.

OBJECTIVES

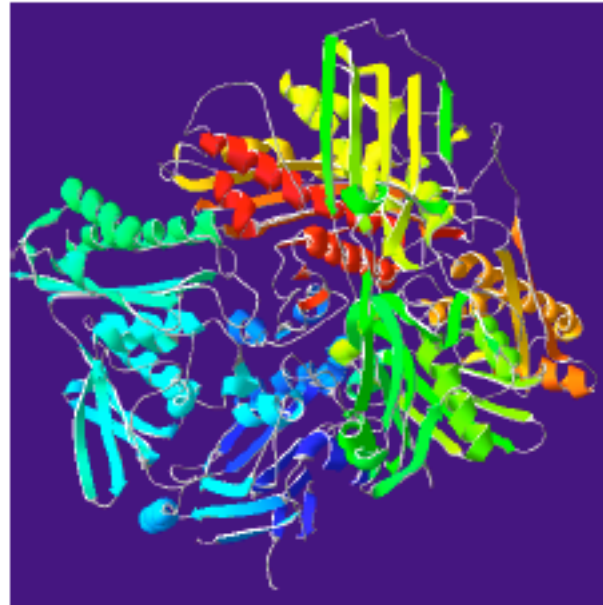
The main objective of this dissertation is the **application of multidimensional chemometric methodologies to the kinetic analysis of enzymatic reactions**. These enzymatic reactions have importance in analytical chemistry and in fermentation processes at industrial scale. This principal objective may be split into several partial objectives:

- The study of reactions related with the catalysis of mixtures of hypoxanthine, xanthine and uric acid with xanthine oxidase with the aim of studying the reaction mechanism in biological matrices and to calculate the enzymatic constants in complex reactions. In addition, the influence of the reaction media in the enzymatic mechanism will be studied in a controlled media and in human urine.

- Quantitative determination of the above mentioned analytes in urine samples by means of their catalytic reaction with xanthine oxidase. In this work different algorithms will be applied and their advantages and disadvantages will be compared.

- Establishment of a methodology for the real time control of fermentation processes. In this sense, multidimensional tensor models will be applied to multidimensional spectrofluorimetry. The studied system will be the production of enzymes lipases throughout the microorganism *Pichia pastoris*.

Another objective linked to this thesis is the divulgation of curve resolution methods. As an example of their use, a laboratory experiment directed to students of the last courses of Chemistry Degree is presented. Multivariate curve resolution with alternating least squares (MCR-ALS) will be applied to the spectrophotometric titration of 8-hydroxyquinoline-5-sulfonic acid and the evolution of the system will be discussed.



Capítulo 1

REACCIONES ENZIMÁTICAS

1.1. INTRODUCCIÓN

La tecnología enzimática es un campo en auge dada la gran importancia que, en los últimos años, han adquirido los enzimas en disciplinas tan diversas como la bioquímica, la medicina y en diferentes tipos de industria (como la agroalimentaria, textil y detergentes) [Web 1.1, Web 1.2, Web 1.3, Gacesa 1990, Gella 2003]. Esto ha generado un gran avance en técnicas que nos ayudan a purificar, extraer, clasificar enzimas así como a monitorizar y estudiar los diferentes mecanismos por los que transcurre una reacción enzimática. Por este motivo, se hace necesario presentar en esta Memoria una visión global sobre aspectos generales de los enzimas, como son su estructura, clasificación y los diferentes tipos de catálisis.

1.2. DEFINICIÓN DE ENZIMA

La palabra “enzima” deriva del griego *έν ζυμη* y significa “en la levadura” [Web 1.4]. Según la Real Academia de la Lengua Española [Web 1.5], un enzima es una proteína que cataliza específicamente cada una de las reacciones bioquímicas del metabolismo. Esta definición es de carácter general y se debe hacer una definición más completa de enzima diciendo que es una proteína compleja que produce un cambio químico específico sobre otras sustancias sin que exista un cambio sobre sí misma. Como

Capítulo 1

se ha dicho en la definición todas las enzimas son proteínas, siendo la única excepción las moléculas catalíticas no proteicas implicadas en la generación del ácido ribonucleico [Nelson 2004].

Los enzimas se pueden encontrar en todos los órganos del cuerpo y su función es vital para el correcto desarrollo de la actividad corporal, de tal manera que una determinada enfermedad puede estar provocada tanto por defecto como por exceso de actividad enzimática. Podríamos decir que la función principal de un enzima es posibilitar reacciones químicas que, en las condiciones de pH, temperatura y fuerza iónica del medio fisiológico, no se podrían dar o se producirían en un tiempo extremadamente largo.

Gran parte de la historia de la Bioquímica es la historia del desarrollo de la investigación enzimática. Establecer una fecha en la cual se descubrió la actividad enzimática es complicado. Los primeros catalizadores biológicos fueron descritos y clasificados a finales del siglo XVIII en estudios sobre la digestión [Nelson 2004]. A partir de este momento, muchos investigadores han catalogado e investigado diferentes tipos de reacciones enzimáticas (Tabla 1.1) [Segel 1993, Nelson 2004]. Desde finales del siglo XX la investigación enzimática ha sido intensa. Hoy en día se purifican y se sintetizan cientos de enzimas para muy diversos usos. Este aspecto se tratará con más profundidad en el apartado 1.6.

1.3. COMPOSICIÓN Y ESTRUCTURA DE LOS ENZIMAS

Los enzimas poseen la misma conformación que las proteínas. Poseen estructura primaria, secundaria, terciaria y cuaternaria (figura 1.1). Su poder catalítico depende en gran manera de su conformación proteica. Si un enzima se desnaturaliza o se disocia en sub-unidades la capacidad catalítica se pierde o se anula totalmente. El peso molecular de un enzima puede oscilar entre 12 000 y más de 1 000 000.

Muchos enzimas sólo requieren de los grupos funcionales de sus aminoácidos para realizar las funciones catalíticas. Muchos otros enzimas requieren de otros grupos

químicos, llamados **cofactores**, que normalmente son metales a nivel de trazas como Fe^{2+} , Mg^{2+} , Mn^{2+} o Zn^{2+} . También hay otros enzimas que requieren de una molécula orgánica o metalo-orgánica compleja llamada **coenzima**. A este coenzima o cofactor unidos covalentemente al enzima se les denomina **grupo prostético**.

Al enzima completo, es decir con grupo prostético se le denomina **holoenzima**. De tal manera que a la parte proteica de una holoenzima se le denomina **apoenzima** o **apoproteína** [Nelson 2004].

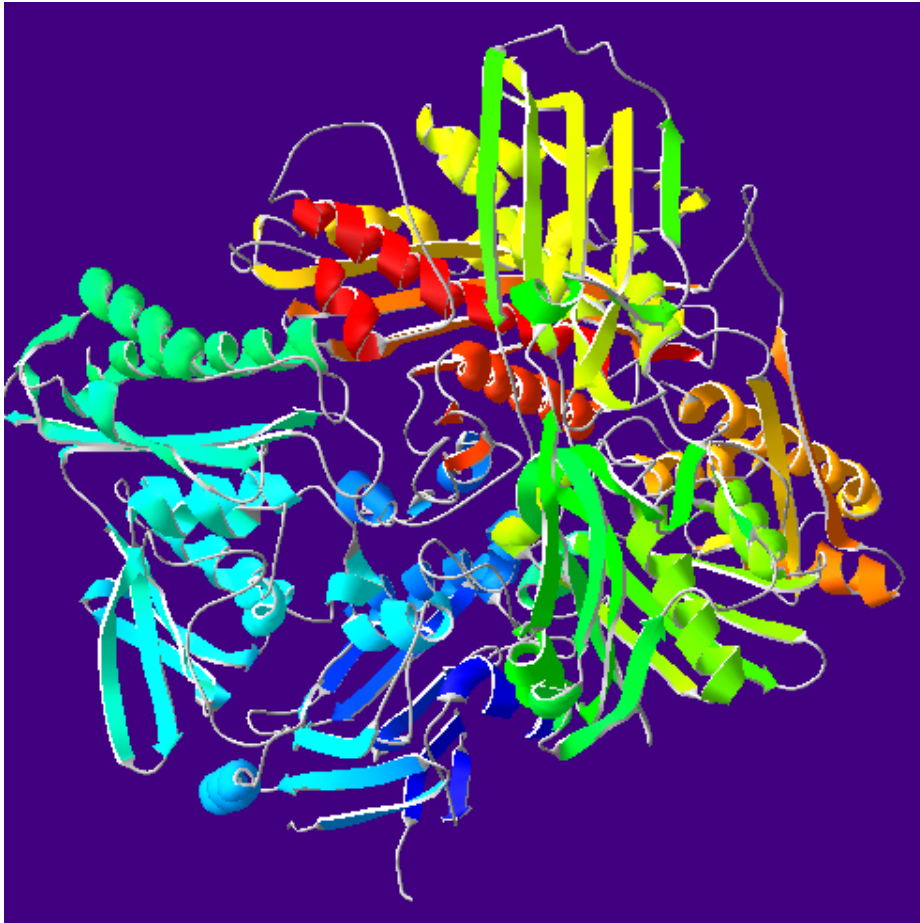


Figura 1.1: Estructura tridimensional de la xantina oxidasa (E.C. 1.17.3.2) [Web 1.4]

Tabla 1.1: Evolución histórica en la investigación enzimática.

Año	Investigador	Suceso	Año	Investigador	Suceso
1783	Spallanzani	La carne era digerida por el jugo gástrico de halcones	1902	Ernst y Brown	Postularon el modelo Enzimas-enzimato
1814	Kirchoff	Un componente glicoproteico del trigo convierte el almidón en azúcar	1913	Michaelis y Menten	Postularon el modelo enzimático general
1830	Robíquet, Boudrou y Chebnd	Hidrólisis de amilglucosa por almidones amargos	1925	Briggs y Haldane	Introdujeron el concepto de estado estacionario en cinética enzimática.
1834	Schwann	Descripción de la papaína	Finales década de 1920	Sumner	Cristalización de enzimas como
1835-1837	Berzelius	Formuló la primera definición de biocatalizador	1930	Nordtop	Cristalización de la papaína, tripsina y quimiotripsina
1858-1871	Pasteur	Concluyó que los cambios químicos en una fermentación están producidos por microorganismos	1940-1950		Descripción y cristalización de ciertos de enzimas y elucidación de sus mecanismos catalíticos
1858-1871	Salky y Liebig	Diferenciación entre "fermentos organizados" y "fermentos no organizados"	1963	Chazan	Presentación del primer modelo enzimático multienzimático
1878	Ktze	Sugirió el nombre de enzimas	1965	Michael, Wyman y Chaugera	Presentación de un modelo de catálisis para enzimas alostéricas
1900		Reconocimiento de que la catálisis de la oxidación celular es producida por acción enzimática	Hoy en día		Producción a escala industrial de miles de enzimas y uso en medicina, biología y diferentes tipos de industria

1.4. CLASIFICACIÓN

Muchos enzimas se nombran añadiendo el sufijo “-asa” al nombre del sustrato o con una palabra o frase que describa su actividad. Por ejemplo, la ureasa cataliza la hidrólisis de la urea y la ADN polimerasa cataliza la polimerización del ADN. Esta forma de nombrar a los enzimas genera cierta ambigüedad. Muchas veces un mismo enzima se denominaba de diferentes maneras o, incluso, dos enzimas podían llamarse igual. Debido a esta ambigüedad y a que cada vez se descubrían nuevos enzimas los bioquímicos, por consenso internacional, adoptaron una normativa para nombrarlos y clasificarlos. El Comité de Nomenclatura de Enzimas de la Unión Internacional de Bioquímica y Biología Molecular (NC-IUBMB) [Web 1.6] clasifica los enzimas en 6 clases, de acuerdo con el tipo de reacción que catalizan [Nelson 2006, Palmer 1985, Gray 1971] (tabla 1.6).

A cada enzima se le asigna un número con 4 cifras. Por ejemplo, el número sistemático de la ATP-glucosa fosfotrasnferasa (hexokinasa) es EC. 2.7.1.1. La primera cifra (2) hace referencia al tipo de reacción que cataliza (transferasa). La segunda cifra (7) denota la subclase (fosfotransferasa). La tercera (1) indica el grupo aceptor (grupo hidroxilo). Y la última cifra (1) indica la D-glucosa como grupo fosforilo aceptor. Más información sobre esta clasificación se puede encontrar en las referencias [web 1.6].

Tabla 1.2: Clasificación de los enzimas

Nombre genérico	Cataliza	Esquema de catalisis	Grupos que catalizan	Ejemplos de enzimas
oxidoreductasas	Oxidoreducción. Transferencia de e ⁻ o H ⁺ de un sustrato a otro	$AH_2 + B \rightleftharpoons A + BH_2$ $A_{ox} + B_{red} \rightleftharpoons A_{red} + B_{ox}$	C=O C=O C-CH ₂ -H C=N, N ₂ C=C-NH ⁺	Xenina oxidasa Sacarato deshidrogenasa Alcohol oxidasa Glucosa oxidasa
transferasas	Transferencia de un grupo químico (diferente de H) de un sustrato a otro	$AB + C \rightleftharpoons A + BC$	Aldehído carbona Amino con S glucosídico fosfatos	Glucosaquinasa Hexoquinasa Transaminasa
Hidrolasas	Reacciones de hidrólisis	$AB + H_2O \rightleftharpoons AH + BOH$	Ésteres C-N Triéster glucosídico Ésteres peptídico Anilídidos de ácido	Lactasa Fosfatasa alcalina Lipasas Pepsina
Ligasas	Reacciones de ruptura o síntesis de sustratos	$AB \rightleftharpoons A + B$	Adición de dobles enlaces C=C O=O C=N-	Acetato desacetilasa Hexoánhidrasa
Isomerasas	Reacciones de interconversión de isómeros	$\text{Isómero A} \rightleftharpoons \text{Isómero B}$	Isómeros	Fosfofructosa isomerasa Fosfogluconato isomerasa
Ligasas	Reacciones de unión de dos sustratos con hidrólisis simultánea de un grupo trifosfato	$A + B + XTP \rightleftharpoons AB + XDP + P_i$	C-O C-N C-S C-C	Glutamina sintetasa Cauronina sintetasa

1.5. CATÁLISIS ENZIMÁTICA. ECUACIÓN DE MICHAELIS-MENTEN

La bioquímica asociada a la catálisis enzimática a menudo es compleja y conlleva un gran número de reacciones simultáneas. Los modelos enzimáticos se caracterizan porque en su gran mayoría conducen a establecer relaciones no lineales entre la evolución cinética de las especies implicadas y su concentración. Estas relaciones no-lineales deben ser válidas para todo el intervalo de concentración de los sustratos, productos y el enzima a lo largo del proceso. Es necesario establecer estas relaciones en base a estudios cinéticos para obtener una descripción cuantitativa de la cinética del sistema.

El modelo básico de cinética enzimática postulado por Michaelis y Menten en 1913 afirma que un sustrato (S) es transformado en un producto (P) por la acción de un enzima (E) de acuerdo con el esquema (1.1) (figura 1.2).

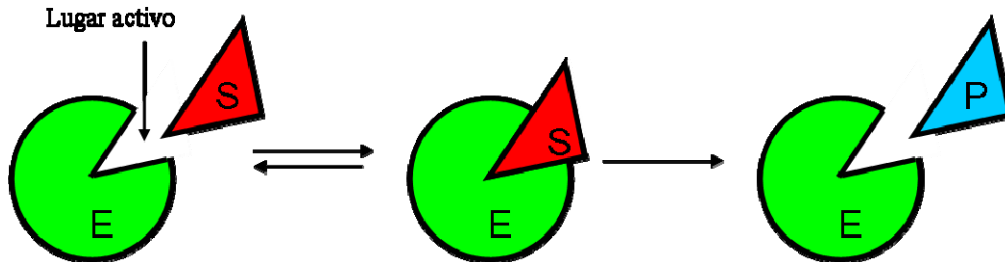
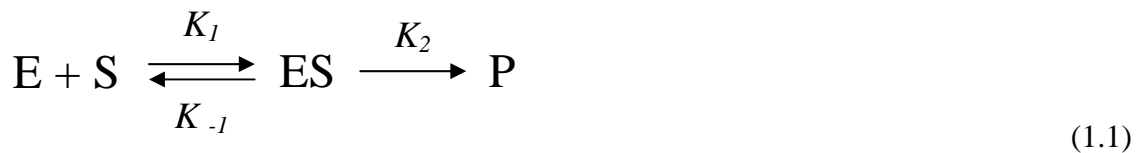


Figura 1.2: Representación gráfica del modelo de una reacción de catálisis enzimática

Esta reacción catalítica queda dividida en dos procesos. En primer lugar el enzima y el sustrato se combinan para formar el complejo enzima-sustrato (ES). Las presunciones que hicieron Michaelis y Menten para desarrollar su teoría son que esta

Capítulo 1

etapa es rápida y reversible y sin que tengan lugar cambios químicos. Enzima y sustrato se mantienen unidos por fuerzas físicas. Los procesos químicos tienen lugar en una segunda etapa caracterizada por una constante de velocidad de primer orden K_2 . Con la colaboración de Briggs-Haldane [Fersht 1999, Nuñez 2001] y aplicando el concepto de “estado estacionario” a la concentración de ES, se llega a la ecuación (1.2). La hipótesis de “estado estacionario” es esencial para el desarrollo del modelo. Viene a postular que la concentración del complejo ES se mantiene constante a lo largo de la reacción enzimática (figura 1.3).

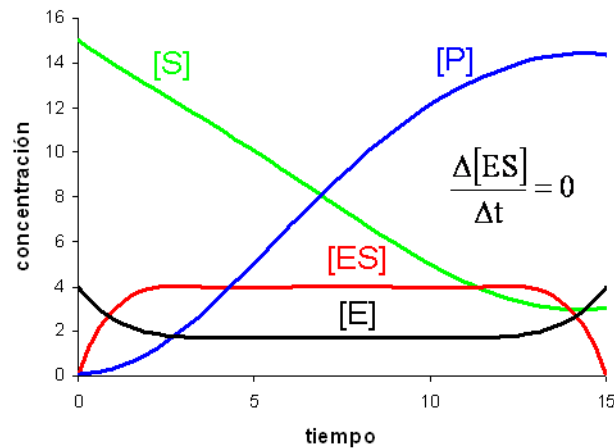


Figura 1.3: Hipótesis del estado estacionario postulada por Briggs-Haldane.

$$v = \frac{d[P]}{dt} = \frac{[E_0][S]K_2}{[S] + (K_2 + K_{-1})/K_1} \quad (1.2)$$

donde v viene definido como la velocidad de aparición de producto con el tiempo. A partir de esta ecuación y asumiendo que:

$$K_M = (K_2 + K_{-1})/K_1 \quad (1.3)$$

$$V_{\max} = K_2[E_0] \quad (1.4)$$

se llega a postular la ecuación básica de la cinética enzimática. La ecuación de Michaelis-Menten (1.5):

$$v = \frac{d[P]}{dt} = \frac{V_{\max}[S]}{[S] + K_M} \quad (1.5)$$

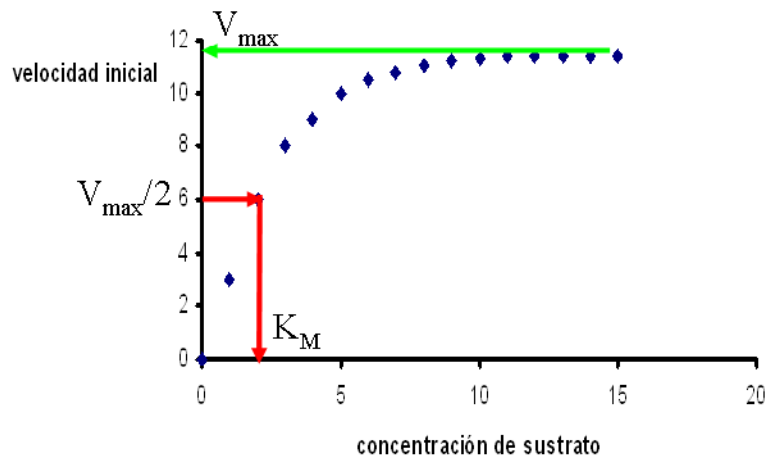


Figura 1.4: Representación gráfica de la ecuación de Michaelis-Menten.

El modelo viene definido por dos constantes. La constante de Michaelis, K_M , que es la concentración de sustrato que resulta en la mitad de la velocidad máxima. La velocidad máxima, V_{\max} se define como la velocidad de la reacción cuando la enzima está saturada con el sustrato (figura 1.4). Estas dos constantes aportan una información muy valiosa a la hora de estudiar el mecanismo que sigue un determinado sistema enzimático.

1.5.1. INHIBICIÓN ENZIMÁTICA

Los enzimas, además de ser inactivados irreversiblemente por temperaturas extremas o por reactivos químicos (por ejemplo, presencia de ciertos metales) pueden ser inhibidos reversiblemente por la unión no covalente de inhibidores. Se entiende por inhibidor una sustancia capaz de actuar sobre el enzima libre o sobre el complejo ES

Capítulo 1

alterando los parámetros cinéticos del enzima y haciendo disminuir o desaparecer su actividad enzimática.

Las reacciones enzimáticas se dan principalmente en las células del cuerpo, por lo que no resulta sorprendente que muchos inhibidores sean importantes agentes farmacológicos conocidos. Por ejemplo, el acetilsalicilato inhibe la enzima que cataliza la síntesis de las prostaglandinas. Incluso, muchas enzimas se ven inhibidas por el propio producto de la reacción catalítica, como es el caso del enzima xantina oxidasa cuando cataliza la oxidación de xantina en ácido úrico [**Massey 1969**].

El estudio de estos inhibidores es de vital importancia para elucidar el modelo de catálisis correcto y establecer rutas metabólicas complejas. En la figura 1.5 se presentan los tres modelos de inhibición más comunes [**Copeland 1996**]. La inhibición competitiva se caracteriza porque el inhibidor (I) se une al enzima de forma reversible, no covalente, pudiendo ser desplazada la reacción de inhibición. El grado de inhibición depende de la concentración de E, S y de P. En la inhibición acompetitiva el inhibidor se une de forma reversible al complejo formado por la enzima y el sustrato. Finalmente, en la inhibición no competitiva el inhibidor puede unirse tanto al enzima libre como al complejo ES.

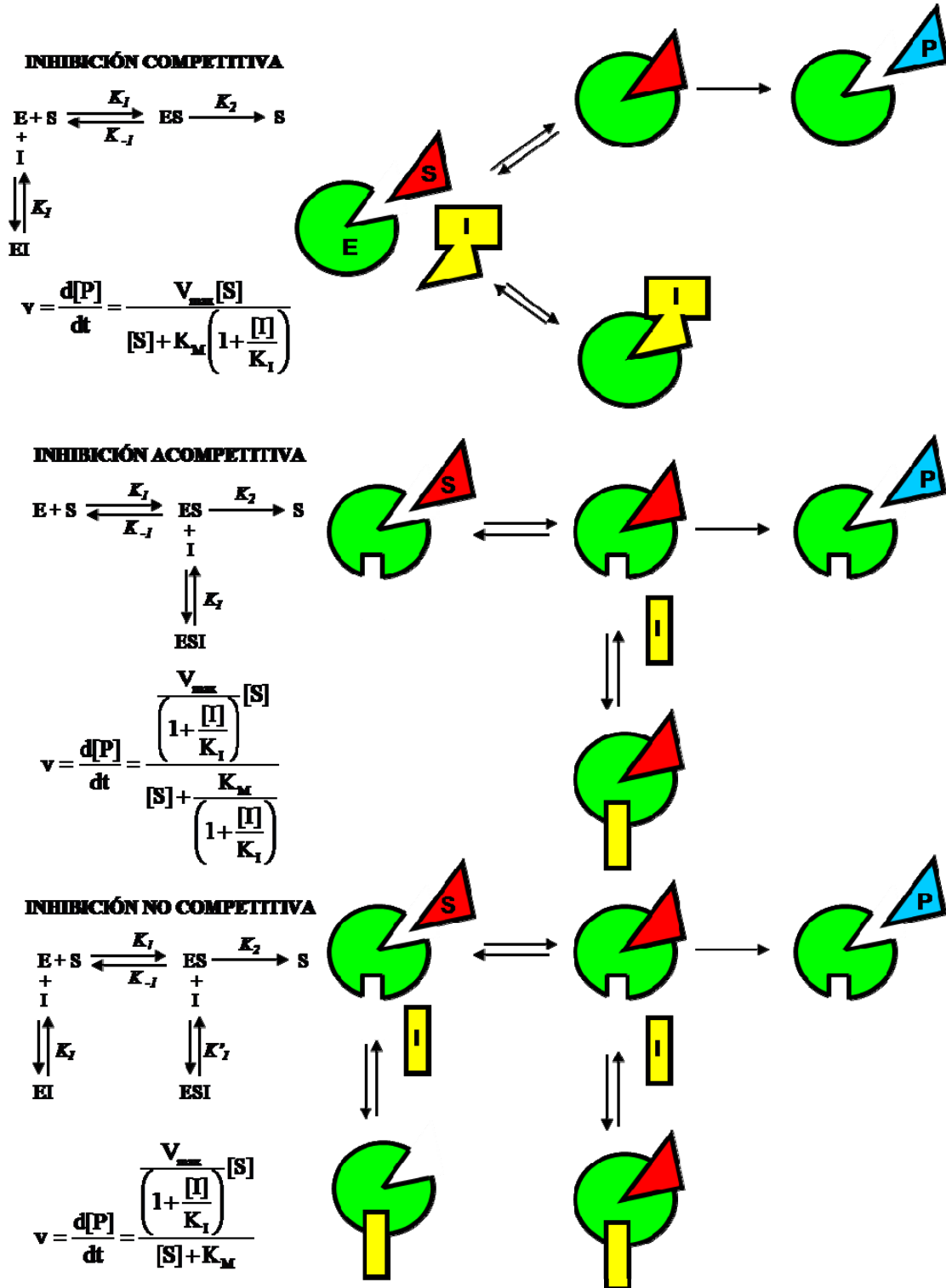


Figura 1.5: Tipos más comunes de inhibición enzimática y su representación gráfica.

1.5.2. REACCIONES MULTISUSTRATO

Hasta ahora se han considerado solamente los casos más sencillos de catálisis enzimática, en los que un sólo sustrato daba lugar a un sólo producto. Sin embargo, gran cantidad de reacciones enzimáticas conllevan la catálisis de varios sustratos simultáneamente. Estas reacciones son llevadas a cabo por los denominados “enzimas de grupo” los cuales no son específicas de una molécula sino que lo son de un determinado grupo funcional o estructura.

Los mecanismos por los que transcurre una reacción multisustrato son mucho más complicados que los mecanismos generales. Por ejemplo, en la figura 1.6 se presenta uno de los mecanismos más clásicos de reacciones multisustrato. Estos esquemas de reacción nos hacen plantearnos cuestiones del tipo: ¿Uno de los sustratos se une y se transforma antes de que el otro sustrato pueda unirse? ¿Es aleatorio el orden en el cual se unen los sustratos al enzima? Para resolver estas cuestiones se han formulado como generales tres mecanismos de reacción multisustrato: La catálisis en orden aleatorio, la catálisis en orden secuencial y el mecanismo ping-pong o de doble desplazamiento.

En el mecanismo de orden aleatorio los sustratos pueden encontrar indiferentemente el centro activo de la enzima. En el mecanismo secuencial, un sustrato presenta más especificidad por el centro activo que el otro sustrato, por lo que la unión del segundo con la enzima se realizará una vez se haya transformado el primero.

Quizás el más interesante por su complejidad sea el mecanismo de doble desplazamiento o mecanismo ping-pong (figura 1.6). Este mecanismo consiste en la formación de dos complejos binarios. Un sustrato entraría en el enzima, quedándose éste con parte del sustrato (un grupo funcional, electrones, etc.) y saliendo el resto como un primer producto. En ese momento se puede unir el segundo sustrato, que se transforma en

un segundo producto aprovechando la parte del primer sustrato que se quedó el enzima [Copeland 1996].

A menudo, el mayor logro de un análisis enzimático es el llegar a poder diferenciar cuál es el mecanismo de la reacción. Por eso se hace imprescindible el desarrollo de nuevas tecnologías que nos proporcionen información precisa sobre el comportamiento de sustratos y enzimas en el transcurso de una reacción.

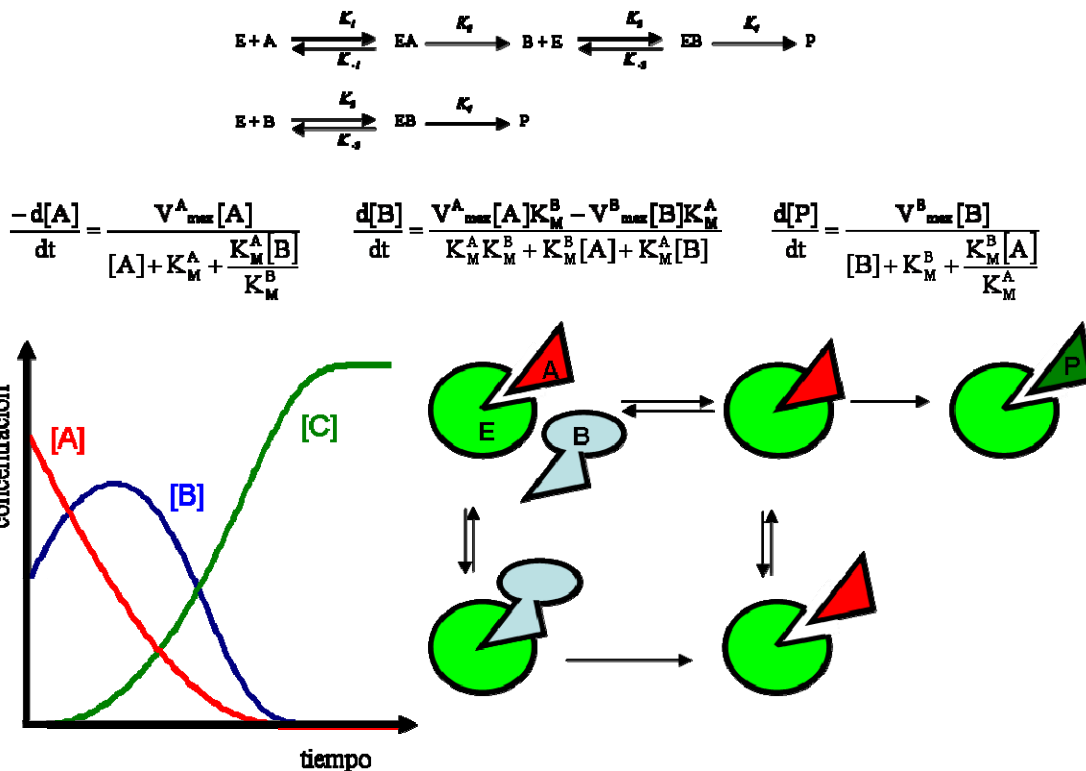


Figura 1.6: Ejemplo de reacción multisustrato siguiendo un mecanismo ping-pong. En esta reacción los sustratos A y B reaccionan simultáneamente con el enzima para producir P.

1.5.3. EFECTO DEL pH, Y LA TEMPERATURA SOBRE LA ACTIVIDAD ENZIMÁTICA

Las reacciones enzimáticas tienen lugar en el medio fisiológico. Este medio es de una complejidad enorme y es normal pensar que, dado el carácter proteico de los enzimas, las condiciones del medio puedan afectar a la actividad enzimática.

El pH afecta, principalmente, al estado de disociación de los grupos protonables / desprotonables de la proteína y, principalmente, del centro activo. La mayoría de enzimas tienen un pH óptimo de actividad, lo que se puede comprobar experimentalmente, dando como resultado la tendencia que se muestra en la figura 1.7. El pH puede afectar de varias maneras a la catálisis enzimática:

- A través de la ionización de los grupos amino o ácido carboxílico del centro activo.
- A través de la ionización de aminoácidos que no están en el centro activo, que puede provocar modificaciones en la conformación del enzima.

El pH también puede afectar al sustrato, protonando o desprotonando los grupos catalizables por el enzima.

La temperatura influye básicamente en la actividad. A medida que la temperatura va aumentando también lo hace la actividad enzimática. A partir de la temperatura óptima el enzima se empieza a desnaturalizar y a perder capacidad catalítica.

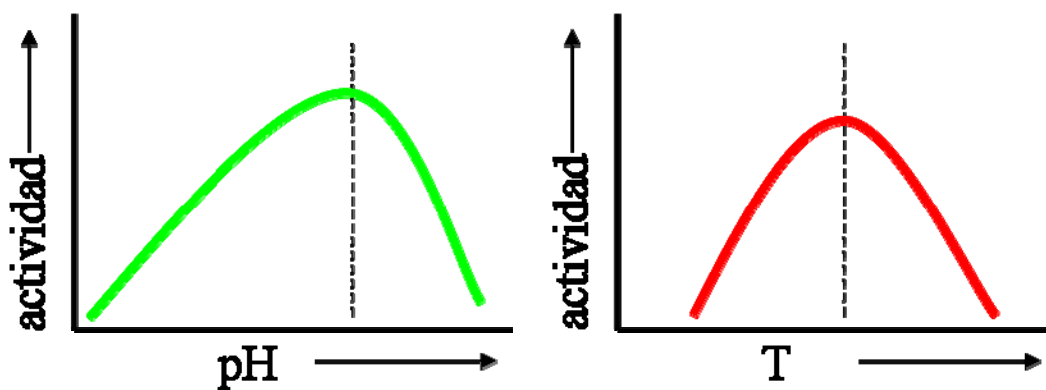


Figura 1.7: Efecto del pH y la temperatura en la actividad enzimática.

1.6. ENZIMAS COMERCIALES

1.6.1. INTRODUCCIÓN

El hombre ha utilizado indirectamente los enzimas desde el comienzo de la Historia de la humanidad. Uno de los primeros ejemplos del uso de enzimas procede probablemente de la observación fortuita de que la leche almacenada en los sacos estomacales de las cabras precipitaba y que la cuajada resultante era un material alimenticio adecuado [Gacesa 1990]. De esta misma manera, se han ido utilizando múltiples procesos biotecnológicos donde el uso de enzimas no requería el conocimiento del mecanismo de actuación de los mismos. Otros procesos naturales usados desde hace muchos años donde los enzimas juegan un papel esencial son el sazonado de carne y de quesos, y la liberación de azúcares fermentables de la cebada durante el proceso de malteado [Web 1.3].

Se puede decir que la explosión comercial de los enzimas comenzó a principios del siglo XX. En la industria de la alimentación los enzimas se empezaron a utilizar en 1930 en la manufactura de zumos de frutas. Se observaba que las pectinasas eran capaces de eliminar la turbidez del zumo. El primer enzima comercial fue el *Bacillus proteasa*. Este enzima fue catalogado en 1959 y comenzó a ser un gran negocio cuando la empresa danesa *Novozymes* la empezó a producir y muchas empresas de detergentes la empezaron a utilizar. Otra gran revolución en la producción de enzimas se dio en 1960 cuando se empezaron a usar enzimas alfa-amilasas para la hidrólisis del almidón, reemplazando por completo a la hidrólisis ácida para producir glucosa [Web 1.3].

En 2005 el valor estimado del mercado mundial de enzimas se situaba alrededor de los 2 billones de dólares americanos [Web 1.3]. Los detergentes, los textiles, el almidón, la manufacturación de pan y alimentos para animales son las principales industrias, las cuales abarcan un 75% de la producción mundial de enzimas (figura 1.8).

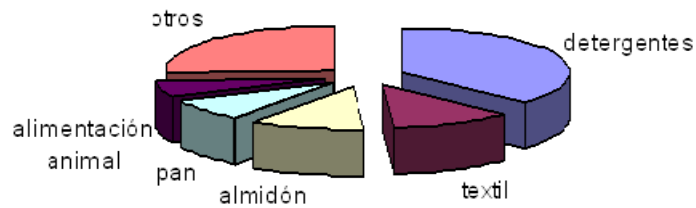


Figura 1.8: Principales usos comerciales de los enzimas.

A parte de la industria de la manufacturación, los enzimas comerciales son ampliamente utilizados en otros campos como son las aplicaciones agrícolas, médicas, farmacéuticas y analíticas.

1.6.2. USO DE ENZIMAS EN LA INDUSTRIA

El incremento del empleo de enzimas en las industrias de manufacturación y procesado ha tenido lugar, principalmente, en dos áreas:

- La mejora de los procesos tradicionales.
- El desarrollo de nuevas vías de manufacturación basados en el estudio de los mecanismos enzimáticos del proceso.

Los campos de la industria agroalimentaria donde la utilización de enzimas se ha asentado de una manera casi imprescindible son los que se indican en la tabla 1.3 [Web 1.3].

1.6.3. APLICACIONES MÉDICAS Y FARMACÉUTICAS DE LOS ENZIMAS

El uso de enzimas en la medicina y campos relacionados presenta un potencial inmenso de posibilidades. Sin embargo, así como en la industria agroalimentaria está muy extendido el uso de enzimas, en el campo de la medicina se está empezando a explotar este potencial.

Empieza a estar extendido el tratamiento terapéutico con enzimas. El fundamento de esta forma de terapia es simplemente la administración de un enzima concreto a un paciente, esperando que se produzca una progresiva mejoría del mismo. El problema de este método es la eliminación de la actividad enzimática por las respuestas defensivas del propio organismo. Otro aspecto a tener en cuenta son las consideraciones éticas que hacen más difícil el trabajo de experimentación y, en consecuencia, el desarrollo de nuevos productos [Gacesa 1990].

Tabla 1.3: Principales campos de aplicación de los enzimas en la industria.

INDUSTRIA	ENZIMAS	ACCIÓN
Detergentes	Proteinasa Lipasa Celulasa	Degradación de proteínas Eliminación de grasas Intensificación del brillo
Textil	Celulasa Lacasa	Eliminación de microfibras Intensificación del brillo
Alimentación animal	Xilasana Fitasa	Solubilidad de la fibra Aumento del fosfato
Almidón	Aamilasas Glucosa isomerasa	Formación de glucosa Formación de fructosa
Papel	Xilanasa	Bio-blanqueante
Zumos de frutas	Pectinasa Celulasa Xilanasa	Clarificación y extracción de zumo
Pan	Xilanasa Alfa-amilasa Glucosa oxidasa	Acondicionamiento de la masa Aumento del pan Calidad de la masa
Lacteos	Renina Lactasa	Coagulación de proteínas Hidrólisis de la lactosa
Cervecera	Glucanasa Papaina	Filtración Control de la calina

En este campo se trabaja intensamente en el tratamiento de enfermedades metabólicas hereditarias, solucionar defectos genéticos, sustituir enzimas naturales por la ingesta de las mismas cuando un órgano del cuerpo ha sido reemplazado por un órgano artificial, el control de la circulación sanguínea y la modificación del ADN [**Web 1.3**].

1.6.4. APLICACIONES ANALÍTICAS DE LOS ENZIMAS

Los enzimas son ampliamente utilizados dentro del análisis químico-clínico. Esto es debido a que los procedimientos enzimáticos tienden a mejorar los procedimientos analíticos clásicos. Muchos métodos químicos convencionales son de difícil utilización para valorar compuestos fisiológicos debido a la falta de sensibilidad y selectividad o de fácil manejo [**Gella 2003**].

Los tipos de aplicación de los enzimas al análisis químico se pueden dividir en dos grandes grupos. El primero de ellos recoge aquellos métodos que miden directamente un compuesto, mientras que el segundo lo hace con aquellos que utilizan un enzima para amplificar otra respuesta.

El ejemplo más característico de los métodos directos es, sin duda, la determinación cuantitativa de D-glucosa a través de su catálisis con el enzima glucosa oxidasa. Quizás sea éste el ejemplo más conocido. Utilizando el mismo fundamento catalítico con diferentes enzimas se han generado biosensores selectivos para una gran cantidad de analitos, como son el lactato, la lactosa, la sucrosa, el etanol, el metanol, el colesterol y algunos aminoácidos [**Gacesa 1990**].

Dentro de los ensayos que utilizan los enzimas para amplificar la respuesta destacan los métodos de inmunoensayo con enzimas acopladas. El más conocido es el denominado método ELISA (*Enzyme-Linked Immuno Sorbent Assay*). Este método se basa en la detección de un antígeno inmovilizado sobre una fase sólida mediante anticuerpos que directa o indirectamente producen una reacción cuyo producto, por ejemplo un colorante, puede ser medido espectrofotométricamente [**Web 1.4**]. Esta

metodología de ensayo ha tenido tanto éxito que la gran mayoría de laboratorios de análisis bioquímicos se apoyan en esta técnica. La versatilidad del sistema es prácticamente ilimitada, debido a la existencia de un número infinito de posibles interacciones antígeno-anticuerpo.

1.6.5. USO DE ENZIMAS DE GRUPO PARA LA DETERMINACIÓN DE MEZCLAS DE ANALITOS EN MATRICES COMPLEJAS

Un campo en el que los enzimas están empezando a jugar un papel importante es en el de la determinación simultánea de varios analitos en matrices complejas. Los enzimas de grupo pueden ser utilizados, con una instrumentación relativamente sencilla, como métodos rápidos de análisis cinético para varios analitos que no tengan la suficiente diferenciación espectral como para ser analizados por métodos espectroscópicos clásicos y que sean catalizables por la misma enzima.

Este tipo de análisis presenta una alternativa a los métodos de análisis clásicos, los cuales conllevan a menudo etapas previas de preconcentración y de separación cromatográfica. Ejemplos de aplicaciones son, por ejemplo, el uso de xantina oxidasa para la determinación simultánea de xantina, hipoxantina y ácido úrico en orina [**Amigo 2005, 2006 y 2006b**] y la determinación simultánea de metanol y etanol con el alcohol oxidasa [**Blanco 1999**].

Durante esta memoria se hará especial hincapié en estas metodologías de análisis enzimático multicomponente. Se hablará de los sistemas más adecuados de monitorización de este tipo de reacciones así como del tratamiento de los datos obtenidos.

1.6.6. OTROS CAMPOS DE APLICACIÓN DE LOS ENZIMAS

Los campos de aplicación expuestos hasta ahora son los más relevantes. Aún así, el empleo de enzimas sigue en pleno auge y cada vez son más las aplicaciones. De entre ellas, destacaremos las siguientes [Web 1.3]:

- Utilización en química fina
- Industria del cuero
- Cosmética
- Producción de aminoácidos quirales puros
- Química de las lipasas en disolventes orgánicos
- Síntesis de oligosacáridos

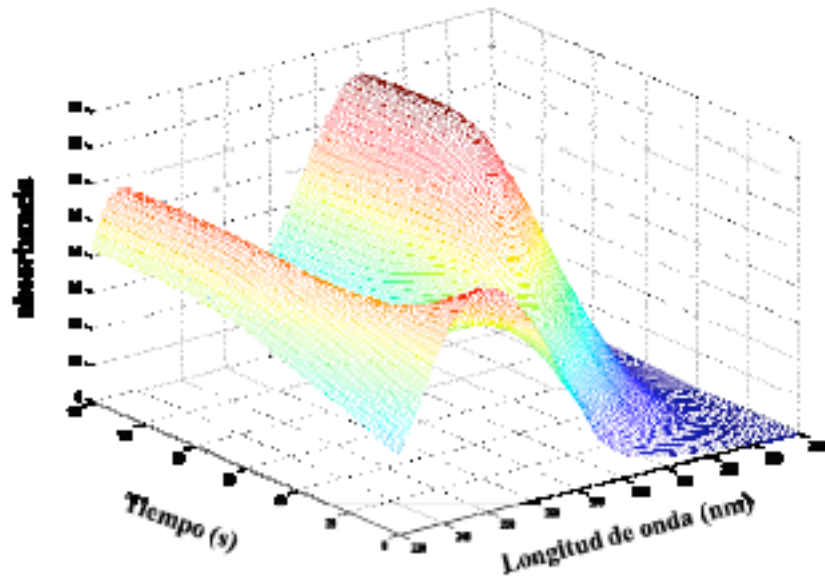
1.6.7. PROCEDENCIA DE LOS ENZIMAS COMERCIALES

En el mundo de la industria, normalmente, se hace necesario el uso de grandes cantidades de enzimas. Esto ha generado la aparición de una industria paralela dedicada a abastecer de enzimas. Son muchas las empresas que se dedican a la manufacturación de enzimas, tanto a escala de laboratorio como a escala industrial. De todas ellas, las dos más importantes son Novozymes ® [Web 1.1] y Genencor ® [Web 1.2].

La mayoría de enzimas que son comercialmente importantes se producen a partir de un número limitado de microorganismos. Los microorganismos son capaces de desarrollarse fácil y rápidamente en su medio de cultivo. Estos microorganismos producen los enzimas en grandes tanques de fermentación.

Tradicionalmente se han obtenido pocos enzimas de origen animal (lipasa pancreática y tripsina, por ejemplo) debido a que éstos pueden ser reemplazadas por otros similares derivados de microorganismos.

Las plantas también son una fuente muy importante de enzimas. Por ejemplo, a partir del látex producido por la papaya y otras especies se ha aislado, sobre todo, cistein-proteinasas. Pero, sin duda, la industria donde los enzimas de origen vegetal tienen más peso es la industria cervecera. La cebada malteada constituye la fuente principal de enzimas vegetales. De esta malta se extrae, por ejemplo, varios enzimas proteasas y glucosidrolasas [Gacesa 1990].



Capítulo 2

MONITORIZACIÓN DE REACCIONES ENZIMÁTICAS

2.1. INTRODUCCIÓN

El enfoque con que se puede abordar el estudio de sistemas enzimáticos dependerá de la finalidad del trabajo que se quiera realizar. Por ejemplo, el interés de un enzimólogo bioquímico está en estudiar los posibles puntos de anclaje enzima-substrato, la estructuración espacial del enzima y su procedencia. Los cinetistas estudian el posible modelo cinético enzimático en que se basa la reacción y posibles intermedios fugaces. En la industria, los enzimas son elegidos por su adecuación a la reacción de interés y al medio en el que se va a desarrollar, su robustez, eficacia y su posibilidad de inmovilización para llevar a cabo la reacción a gran escala sin importar demasiado el mecanismo por el cual transcurre. Desde un punto de vista analítico, la información que se busca es la especificidad de un determinado enzima para la resolución de una mezcla de analitos ya sea por diferenciación espectral o cinética.

Se han desarrollado metodologías de análisis enzimático muy diferentes, cada una de ellas adaptada al objetivo de la experimentación. En esta Memoria se destacarán las metodologías empleadas en la monitorización de la reacción enzimática para adquirir información de tres tipos:

- Información cinética: Determinar el mecanismo por el cual transcurre la reacción con un determinado sustrato, así como sus constantes enzimáticas.

- Información cualitativa: Diferenciar cinética o espectralmente mezclas de sustratos catalizables por un mismo enzima, así como determinar la existencia de intermedios fugaces.

- Información cuantitativa: Determinar y controlar la concentración de cada uno de los sustratos e intermedios a lo largo de la reacción.

2.2. ADQUISICIÓN DE DATOS CINÉTICOS. INSTRUMENTACIÓN

Los métodos cinéticos exigen el perfecto control y optimización de las variables experimentales. Éstas se refieren, fundamentalmente, a la temperatura, concentración de sustratos e intermedios, actividad enzimática, pH y fuerza iónica del medio. Por lo que la instrumentación a utilizar tiene que ser robusta, de fácil manejo y de actuación rápida.

Hay muchas maneras de obtener la información necesaria para poder estudiar un sistema enzimático, pero sin duda, cuando se quiere monitorizar la evolución de una reacción enzimática las metodologías más empleadas son las espectroscópicas. Estas técnicas se han extendido ampliamente debido a que la gran mayoría de sustratos y enzimas absorben en las regiones espectrales del ultravioleta, visible (absorción de radiación y fluorescencia), Infrarrojo medio, MIR (*Middle Infrared*) e Infrarrojo cercano, NIR (*Near Infrared*). En esta Memoria se utilizan, concretamente, dos de ellas, la espectroscopia UV-Visible y la espectrofluorimetría multidimensional.

2.2.1. ESPECTROSCOPIA UV-VISIBLE-DAD ACOPLADA A MÓDULOS DE FLUJO INTERRUMPIDO.

La espectroscopia Ultravioleta-Visible (UV-Vis) es una técnica ideal para la monitorización de sistemas enzimáticos debido a la gran evolución que han experimentado los espectrofotómetros en las posibilidades de registro. Hoy en día, resulta muy sencillo disponer de espectrofotómetros UV-Vis equipados con detectores de matriz

de diodos, DAD (*Diode Array Detector*), los cuales pueden registrar el espectro UV-Vis en un rango amplio de longitudes de onda en un corto intervalo de tiempo. Esta característica es importante, sobre todo si se tiene en cuenta que las reacciones enzimáticas acostumbran a ser rápidas (del orden de segundos). Por ejemplo, muchos espectrofotómetros equipados con detector de diodos son capaces de registrar un espectro UV-Vis en el rango completo de longitudes de onda 0.1 segundos.

Otra de las grandes ventajas que tienen los espectrofotómetros UV-Vis es la sencillez con la que se pueden acoplar sistemas para el control de la temperatura y sistemas de mezclado rápido. Uno de los sistemas de control de temperatura más extendidos son los módulos *Peltier*. Estos sistemas ofrecen una gran robustez a la hora de controlar la temperatura de una reacción con una variación de una décima de grado centígrado en la cubeta de reacción.

En 1934 Roughton [**Roughton 1934**] introdujo el método de flujo interrumpido (*stopped-flow*) como mejora del método de flujo continuo para mezclas de sustratos y enzimas. El principio se muestra en la figura 2.1. Mediante una cámara de mezcla se conectan dos jeringas a un tubo de desagüe. Se llena una jeringa con la disolución de enzima, mientras que la otra se llena con el sustrato. En un determinado momento se comprimen ambas jeringas simultáneamente proporcionando entre 50 y 200 μL cada una de ellas. Las dos soluciones se mezclan completamente en la cámara de mezcla (cubeta de flujo en la figura 2.1) siendo detenidas mecánicamente. Esta metodología de mezclado proporciona una velocidad de homogeneizado de milisegundos, por lo que si el sistema detector es adecuado, se podrán monitorizar reacciones enzimáticas que se lleven a cabo en un segundo, o menos.

El tipo de medidas que se obtienen cuando se utiliza un espectrofotómetro UV-Vis equipado con un detector de matriz de diodos y un módulo de flujo interrumpido son las que se muestran en la figura 2.2. Se obtiene un espectro UV-Vis en un rango de longitudes de onda definido para cada una de las unidades de tiempo que se registren.

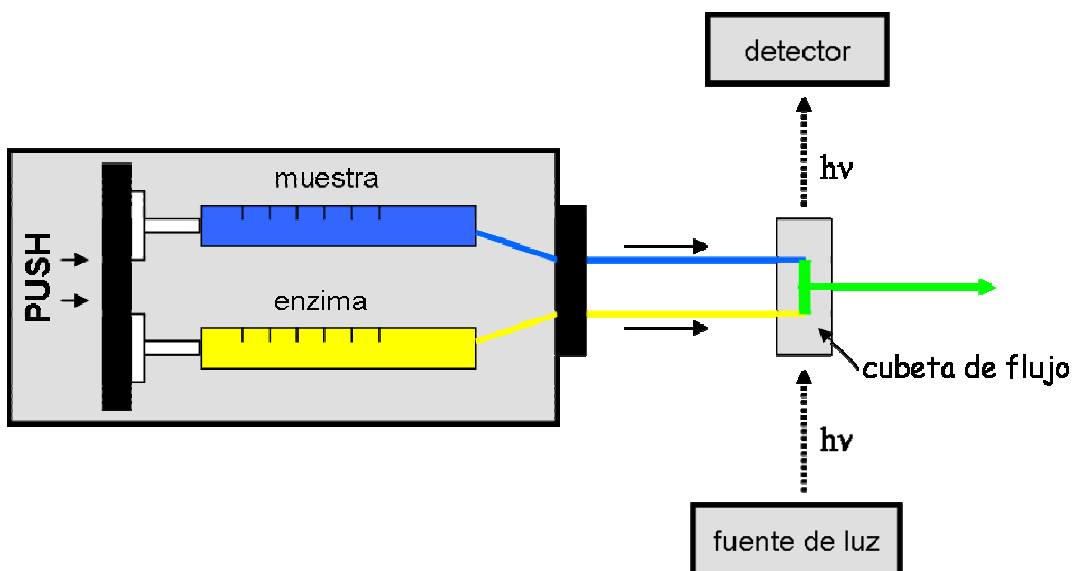


Figura 2.1: Esquema del sistema de flujo interrumpido (*stopped-flow*) acoplado al espectrofotómetro UV-Visible.

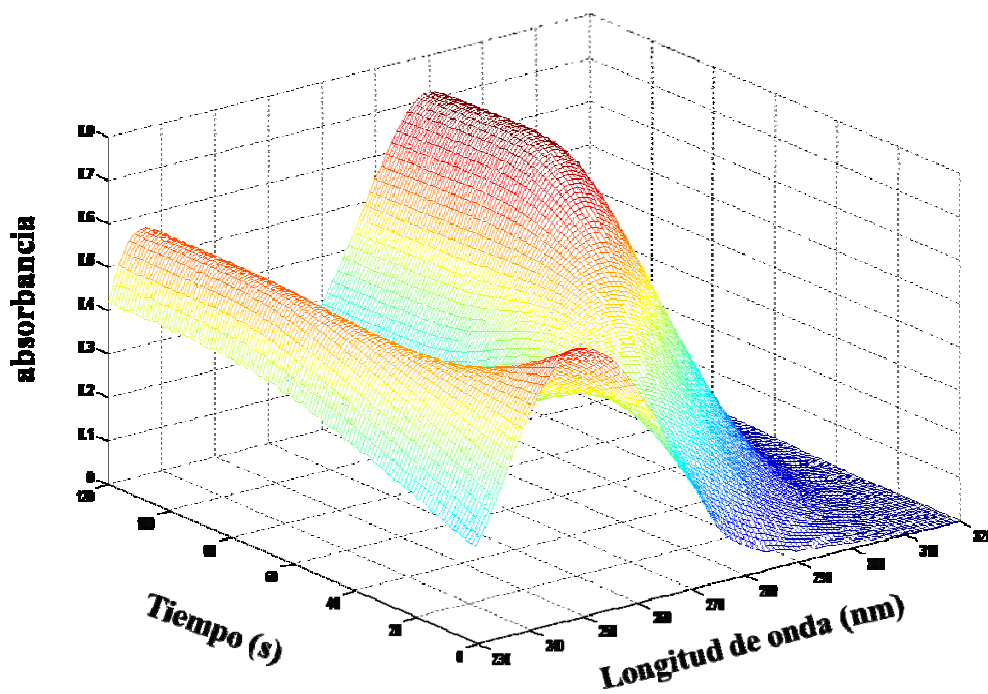


Figura 2.2: Representación gráfica del registro espectral con el tiempo que se puede obtener con un espectrofotómetro UV-Vis-DAD acoplado a un módulo de flujo interrumpido (*stopped-flow*).

Este método de monitorización aporta mucha información en un corto espacio de tiempo, por lo que es uno de los más empleados para:

- Conocer y estudiar el modelo enzimático.
- Determinar el número de especies absorbentes en el sistema.
- Determinar las constantes enzimáticas del sistema.

2.2.2. FLUORESCENCIA MULTIDIMENSIONAL.

Los espectrofotómetros de fluorescencia han experimentado un gran auge en los últimos años debido a su gran aplicabilidad en la monitorización de bioprocesos [Skibsted 2001, Surribas 2006, Kermis 2002]. Estos espectrofotómetros son capaces de registrar la señal de fluorescencia de un determinado sistema en un amplio intervalo espectral de excitación y emisión. Dotados de una sonda de fibra óptica, la fluorescencia multidimensional se ha convertido en una metodología no invasiva poderosa para estudiar y monitorizar reacciones **bioquímicas** [Lindemann 1998, Hisinger 2005] (figura 2.3).

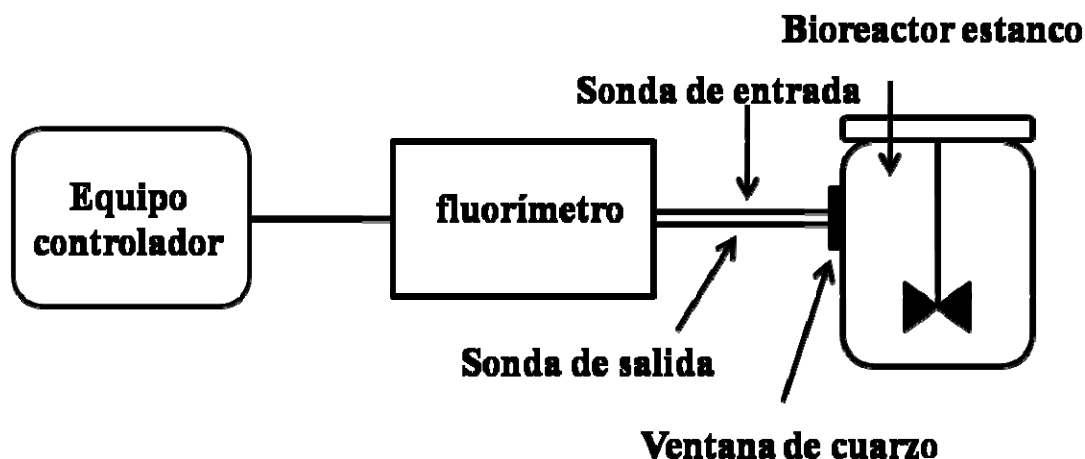


Figura 2.3: Esquema del equipo formado por el fluorímetro multidimensional acoplado a un bioreactor estanco.

Una de las principales ventajas de esta metodología de monitorización es la característica de técnica “no invasiva”. Los espectros de fluorescencia se registran conectando una sonda de fibra óptica al bioreactor a través de una ventana transparente, normalmente de cuarzo (figura 2.3). Esto hace posible registrar toda la información

necesaria sin necesidad de introducir ninguna sonda al bioreactor. De esta manera se conserva la estanqueidad del sistema evitando así posibles contaminaciones que se pueden producir a la hora de la recogida de muestra para el análisis.

Otra ventaja es la rapidez con que se registran los espectros. El espectrofluorímetro multidimensional es capaz de realizar la medida en pocos segundos. Esta metodología se ha adaptado al control de fermentaciones y bioprocesos. La duración de éstos normalmente es de horas, o incluso días, por lo que se puede considerar que el registro espectral se realiza en un tiempo relativamente despreciable.

El tipo de datos que se registran con esta metodología son de naturaleza multidimensional, tal y como se observa en la figura 2.4.

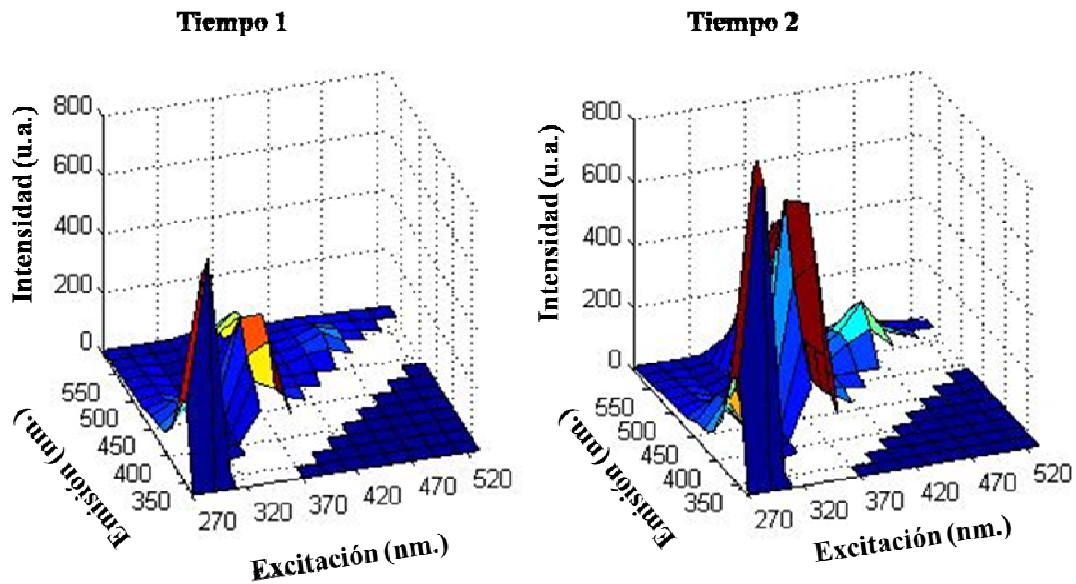


Figura 2.4: : Representación gráfica del registro espectral con el tiempo que se obtiene con un fluorímetro multidimensional acoplado a un bioreactor.

2.3. TRATAMIENTO DE LOS DATOS REGISTRADOS

Las técnicas de tratamientos de datos registrados en una evolución enzimática han ido avanzando al mismo ritmo que lo hacía la capacidad de cálculo de los ordenadores

personales y la necesidad de estudiar sistemas enzimáticos cada vez más complejos. De los primeros métodos gráficos de linearización de la ecuación de Michaelis-Menten mediante el uso de datos puntuales se han pasado a métodos que usan toda la información espectral y cinética para elucidar el modelo enzimático y calcular la concentración de cada uno de los analitos en cada momento. A continuación se expone un resumen de todas estas metodologías y se discute la aplicabilidad de cada una de ellas.

2.3.1. MÉTODOS GRÁFICOS DE CÁLCULO DE CONSTANTES.

Las medidas cinéticas de un enzima en el estado estacionario sólo proporcionan, normalmente, dos datos cinéticos, el valor de K_M y el valor de V_{max} . Hasta hace unos años, el cinetista sólo disponía de unas pocas herramientas para poder estudiar la evolución de una reacción enzimática, encontrándose limitado si la reacción era compleja (reacciones multisustrato, por ejemplo). Las metodologías más empleadas están basadas en la linearización de la ecuación de Michaelis-Menten utilizando un mínimo de datos experimentales, como por ejemplo, la variación de la velocidad inicial de reacción con diferente concentración inicial de sustrato, para el cálculo de las constantes enzimáticas.

La representación gráfica de la ecuación de Michaelis-Menten (velocidad inicial frente a concentración de sustrato, ecuación 1.5, figura 1.4) es una hipérbola. La V_{max} corresponde al valor máximo al que tiende la curva experimental, y la K_M corresponde a la concentración de sustrato a la cual la velocidad de la reacción es la mitad de la V_{max} . Para determinar gráficamente los valores de K_M y V_{max} se tiende a linearizar la ecuación de Michaelis-Menten. En este sentido, cabe destacar tres variaciones de la ecuación (figura 2.5):

- **Gráfico de Lineweaver-Burk:** Este tipo de representación gráfica también se conoce como gráfico de dobles recíprocos o de inversos. Es una de las maneras más conocidas de linearizar la ecuación 1.5 y una de las que aporta mejores resultados cuando el sistema enzimático es sencillo.

- **Gráfico de Hanes-Woolf**

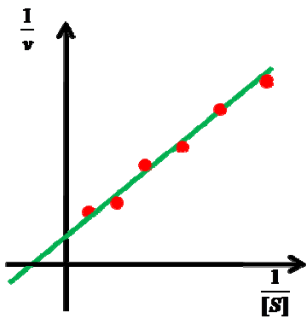
· Gráfico de Eadie-Hofstee

Estos métodos gráficos son muy útiles para la obtención de una estimación semi-cuantitativa de las constantes y cuando el sistema estudiado no es complejo, como por ejemplo sistemas multisustrato.

Michaelis-Menten

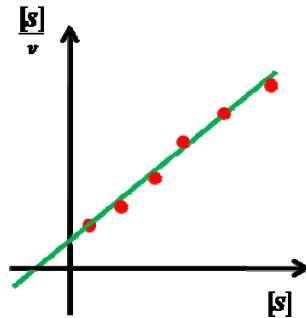
$$v = \frac{d[P]}{dt} = \frac{v_{\max}[S]}{[S] + K_M} \xrightarrow{\text{Linearización}} y = ax + b$$

Lineweaver-Burk



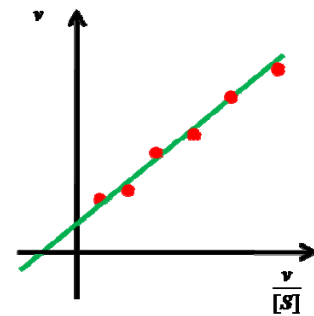
$$\frac{1}{v} = \frac{K_M}{V_{\max}} \frac{1}{[S]} + \frac{1}{V_{\max}} \begin{cases} a = \frac{K_M}{V_{\max}} \\ b = \frac{1}{V_{\max}} \end{cases}$$

Hanes-Woolf



$$\frac{[S]}{v} = \frac{1}{V_{\max}} [S] + \frac{K_M}{V_{\max}} \begin{cases} a = \frac{1}{V_{\max}} \\ b = \frac{K_M}{V_{\max}} \end{cases}$$

Eadie-Hofstee



$$v = -K_M \frac{v}{[S]} + V_{\max} \begin{cases} a = -K_M \\ b = V_{\max} \end{cases}$$

Figura 2.5: Métodos de linearización de la ecuación de Michaelis-Menten para el cálculo de las constantes enzimáticas.

2.3.2. METODOLOGÍAS DE MODELADO DURO (HARD-MODELLING METHODS).

Los métodos gráficos de linearización de la ecuación de Michaelis-Menten han quedado superados por los métodos de ajuste no-lineal con los que se cuenta hoy en día para calcular las constantes enzimáticas de un sistema. Son las denominadas metodologías de modelado duro, o metodologías *Hard*.

Estos métodos se basan en el ajuste de los perfiles de absorbancia obtenidos a determinadas longitudes de onda puntuales a lo largo del tiempo, transformados a concentraciones, con ecuaciones diferenciales presupuestas con anterioridad mediante algoritmos de ajuste no-lineal [**Escribano 1988**]. El algoritmo de ajuste no-lineal más empleado es el desarrollado por Marquardt en 1963 [**Marquardt 1963**]. Este proceso se observa en la figura 2.6 y consta de los siguientes pasos:

- 1) Elección del número de especies absorbentes que intervienen en el modelo enzimático.
- 2) Elección de las longitudes de onda más apropiadas para cada una de las especies.
- 3) Transformación de la evolución de la absorbancia en evolución de la concentración mediante el uso de las absorptividades molares de cada una de las especies a una determinada longitud de onda.
- 4) Presuposición de un modelo cinético-enzimático para cada una de las especies.
- 5) Ajuste de los perfiles de concentración al modelo presupuesto mediante algoritmos de ajuste no-lineal.
- 6) Estudio de los residuales del ajuste para verificar la adecuación del modelo cinético presupuesto al sistema enzimático estudiado.
- 7) Obtención de las constantes del sistema.

Las metodologías *Hard* de trabajo se ha utilizado dando buenos resultados en sistemas complicados, como en reacciones multisustrato, pero tiene diversos problemas:

- Es imprescindible conocer, de antemano, las absorptividades molares de cada uno de las especies en las longitudes de onda preseleccionadas. Por lo que es necesario un estudio previo del espectro de las especies absorbentes a lo largo de la reacción.

- En determinados sistemas enzimáticos es complicado presuponer un modelo cinético que defina perfectamente la evolución de cada una de las especies.
- Esta metodología es muy sensible a variaciones inesperadas que puedan sufrir las especies a lo largo de la reacción así como a la existencia de interferencias de absorbancia constante.

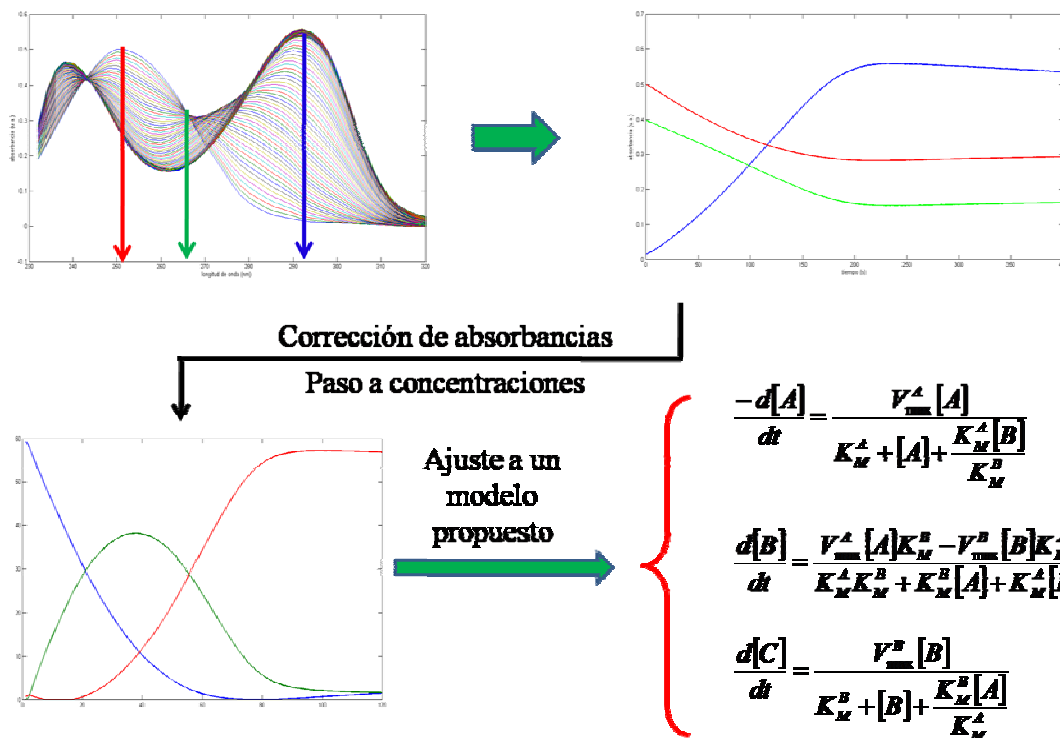
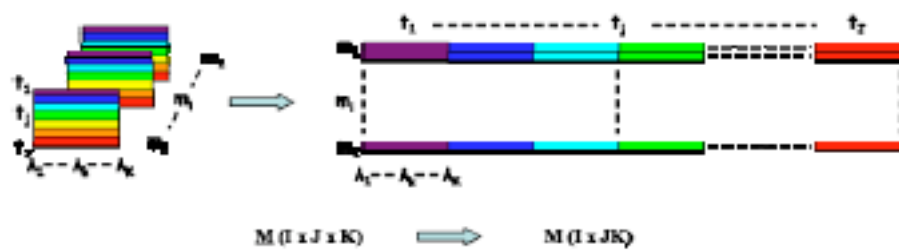


Figura 2.6: Metodología de trabajo de modelado duro (*Hard Modelling*).

2.3.3. METODOLOGÍAS DE MODELADO BLANDO (*SOFT-MODELLING METHODS*). INTRODUCCIÓN AL ANÁLISIS MULTIVARIABLE.

Las metodologías *Hard* y los métodos gráficos, tal y como se han definido, hacen uso de información puntual para el análisis cinético. Es decir, hacen uso de información univariable. Sin embargo, se han desarrollado metodologías de análisis de sistemas cinéticos que hacen uso de toda la información espectral registrada de manera simultánea. Son las denominadas metodologías multivariable de análisis.

Unas de las metodologías multivariable de modelado más útiles son las metodologías de modelado blando (*Soft Modelling*), las cuales son capaces de modelar de forma semi-cuantitativa un sistema enzimático sin necesidad de presuponer el modelo cinético. Su utilidad y el tipo de información que se puede obtener con cada una de ellas, se estudia con profundidad en el capítulo siguiente.



Capítulo 3

LA QUIMIOMETRIA COMO HERRAMIENTA PARA MODELAR REACCIONES ENZIMÁTICAS

3.1. INTRODUCCIÓN

En el capítulo anterior se ha expuesto cómo, hoy en día, es relativamente sencillo obtener una gran cantidad de información en un tiempo lo suficientemente rápido para la monitorización de un sistema enzimático. Con la generación de esta gran cantidad de información, generalmente el registro de la variación del espectro con el tiempo, la siguiente necesidad es el desarrollo de herramientas que nos permitan extraer la información estrictamente relevante de nuestro sistema y que nos ayude a conocer mejor su evolución y comportamiento. Ésta puede ser:

- **Información cualitativa:** Conocer el número de especies absorbentes en la reacción, la existencia de intermedios lo suficientemente estables y su evolución, tanto cinética como espectral, como paso previo a establecer un modelo empírico. Saber las posibles interferencias espectrales y cinéticas presentes en la reacción.

- **Información cuantitativa:** Conocer la concentración de cada especie en cada momento. Conocer el modelo enzimático por el que se rige el sistema así como sus constantes enzimáticas en ausencia o presencia de interferentes de tipo espectral o cinético.

3.2. DEFINICIÓN. CLASIFICACIÓN DE LOS MÉTODOS QUIMIOMÉTRICOS

La Quimiometría se ha definido de múltiples maneras [Wold 1972, Wold 1995] siendo quizás la más acertada, o la más global, la definición dada por el Prof. Massart [Massart 1997]: *“La Quimiometría es la parte de la Química que se sirve de las matemáticas, estadística y lógica formal para: diseñar o seleccionar procedimientos experimentales óptimos, proporcionar información química relevante a partir del análisis de señales analíticas y, finalmente, adquirir conocimiento de los sistemas químicos”*.

Los tratamientos quimiométricos se encuentran en una clara expansión en su utilización. Hace unos años la capacidad de cálculo de los computadores asequibles era un problema a la hora de obtener resultados en un breve espacio de tiempo. Hoy en día este problema se está solventando con la evolución de los ordenadores personales. Por otro lado, se está trabajando continuamente en la mejora de los algoritmos y en la implementación de éstos en interfases gráficas fáciles de manejar por cualquier usuario [Jaumot 2005].

Los campos donde la quimiometría se viene utilizando o se empieza a utilizar van en aumento debido a la gran cantidad de herramientas de análisis útiles que podemos adaptar a la resolución de los problemas individuales. El diseño de experimentos, el tratamiento de señales, la calibración multivariable, la resolución de sistemas dinámicos, el reconocimiento de pautas y el control de procesos son algunos de los campos más destacados que hacen uso de diferentes técnicas quimiométricas.

Los métodos analíticos pueden tener una base totalmente empírica o bien estar soportados por una base teórica que explica el fenómeno, físico o químico, responsable de la señal analítica. Boosksh y Kowalski [Boosksh 1994] clasificaron estos métodos desde un punto de vista matemático, en función de las dimensiones de los datos disponibles por muestra, los cuales han sido generados por instrumentos de diferente

dimensión (tabla 3.1). Cuando se tratan simultáneamente distintas muestras, el orden de los datos es igual al orden del instrumento aumentado en una dimensión, la cual hace referencia al número de muestras.

Tabla 3.1: Clasificación de los métodos analíticos en función de los datos disponibles por muestra

Datos obtenidos	Orden de los instrumentos	Ejemplos de instrumentación	Información obtenida	Técnicas de análisis matemático más comunes
Univariabes	0	pH-metro, Absorbancia a una λ , cromatografía	Escalar	Mínimos cuadrados
Multivariables	1	DAD-UV-Vis NIR, MIR	Vector Tensor de orden 1	PCA, PCR, PLS, MLR, ILS,
	2	DAD-UV-Vis midiendo cada t tiempo	Matriz Tensor de orden 2	MCR-ALS, U-PCA, U-PLS
	superior	Fluorímetro multidimensional midiendo cada t tiempo	Cubo Tensor de orden n (normalmente 3)	3W-PCA, 3W-PLS PARAFAC PARAFAC2 TUCKER 3

Esta memoria está dirigida al estudio y comparación de técnicas quimiométricas multidimensionales que se pueden aplicar a la resolución de sistemas evolutivos registrados de manera multivariable (datos de orden 3 y superior, cuadros en azul de la tabla 3.1). Se hará especial énfasis en el estudio de las técnicas de análisis tridimensional como 3W-PCA y 3W-PLS, 3W-MCR-ALS, HS-MCR-ALS y PARAFAC. Si bien, se ha de señalar que se hará uso puntual de metodologías de tratamiento de datos de orden 2 como PCA, PLS y MCR-ALS.

3.3. ANÁLISIS QUIMIOMÉTRICO MULTIDIMENSIONAL DE DATOS EVOLUTIVOS MULTIVARIABLE

3.3.1. DATOS EVOLUTIVOS. DATOS DE SEGUNDO Y TERCER ORDEN

Cuando se monitoriza una reacción con un instrumento de segundo orden (por ejemplo, un espectrofotómetro UV-Vis. equipado con un detector de diodos DAD), lo que generalmente se obtiene es un espectro (en un determinado rango continuo de longitudes de onda) por cada unidad de tiempo de medida. Traduciéndolo a lenguaje

matemático, lo que se está obteniendo es una matriz bi-dimensional \mathbf{M} ($J \times K$), donde la dirección de las filas representa los J espectros obtenidos para cada uno de los J tiempos de medida y las columnas representan el rango de las K longitudes de onda de medida (figura 3.1). Si repetimos este proceso para I muestras, lo que obtendremos es una matriz tri-dimensional $\underline{\mathbf{M}}$ ($I \times J \times K$), donde la tercera dimensión hará referencia al número de muestras monitorizadas.

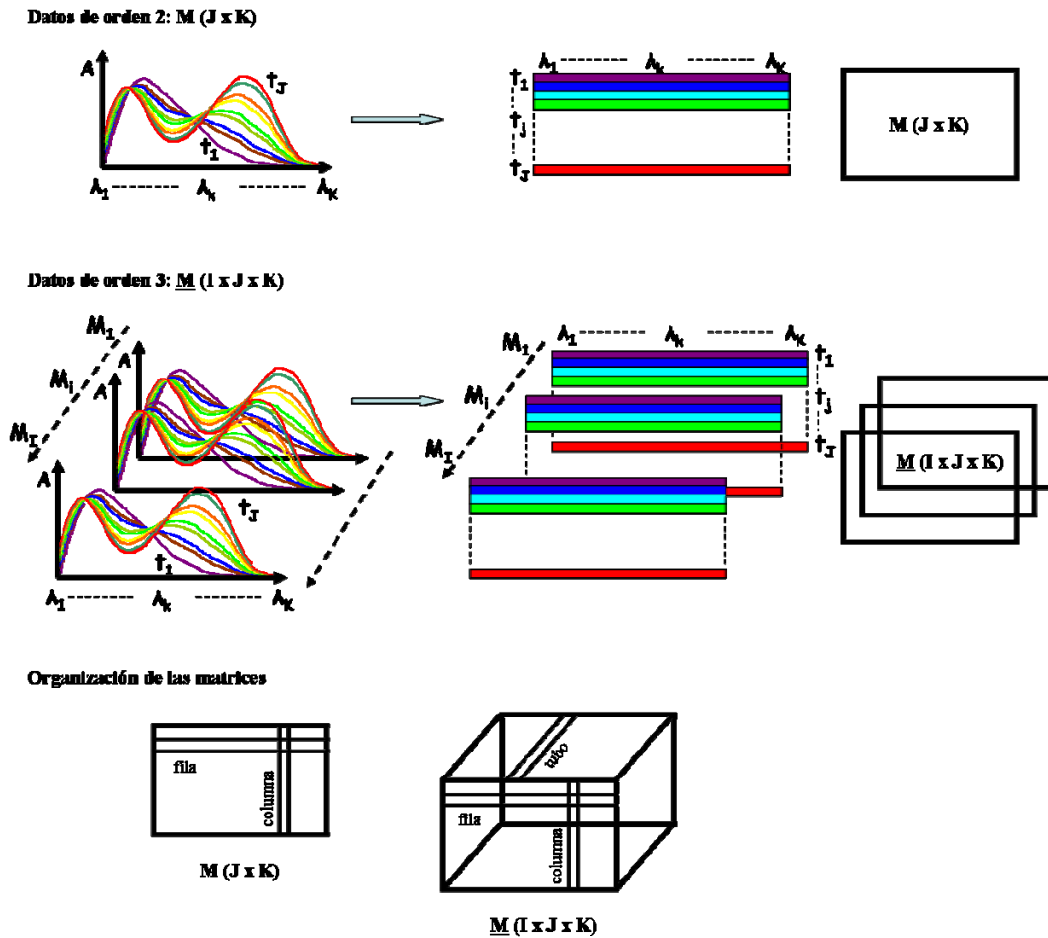


Figura 3.1: Organización de datos de segundo y tercer orden.

Existen métodos de análisis de matrices bidimensionales, tipo \mathbf{M} ($J \times K$) obtenidas para una muestra, como son la Resolución Multivariante de Curvas (MCR-ALS), la cual se estudiará con profundidad más adelante (ver sección 3.7). Pero existe un claro interés en estudiar un sistema enzimático monitorizando varias reacciones en las que las condiciones iniciales varíen, como por ejemplo variación en la concentración inicial de

sustratos, la temperatura o el pH. De esta manera podremos obtener más información de nuestro sistema si se crea una matriz tridimensional \mathbf{M} ($I \times J \times K$) donde cada muestra represente una condición diferente.

Antes de empezar con la descripción de cada uno de los métodos para el tratamiento de este tipo de datos es importante aclarar diversos conceptos de carácter matemático que serán aplicados a lo largo de este capítulo y que son de vital importancia a la hora de entender explicaciones posteriores.

3.3.2. CONCEPTOS DE LINEALIDAD Y RANGO DE UNA MATRIZ MULTIDIMENSIONAL.

3.3.2.1. Linealidad, bilinealidad y trilinealidad

Linealidad

Los conceptos “lineal” y “linealidad de un sistema” son ampliamente conocidos, si bien, existe mucha confusión sobre cómo utilizarlos adecuadamente. Se dice que en un sistema lineal se cumple una relación matemática entre las variables (y) y los parámetros (x) del tipo

$$y_i = b_1 x_i + b_0 + e_i; \quad i = 1, \dots, I \quad (3.1)$$

Siendo y_i los valores predichos y x_i los predictores para cada una de I medidas realizadas. Esta es la manera habitual de representar un modelo de regresión clásico (mínimos cuadrados) con una ordenada en el origen b_0 , una pendiente b_1 y un término de error e_i [smilde 2004]. En este caso se dice que el modelo es lineal tanto en los parámetros como en las variables. El caso más conocido es la ley de Beer-Lambert, donde la absorbancia sigue un modelo lineal con la concentración de analito dentro de un rango determinado de concentraciones. Un ejemplo de modelo lineal en los parámetros, pero no lineal en x es:

Capítulo 3

$$y_i = b_2x_i^2 + b_1x_i + b_0 + e_i; \quad i = 1, \dots, I \quad (3.2)$$

Es muy útil diferenciar entre estos dos tipos de linealidad. Las cinéticas son casos bien conocidos en los que normalmente no se cumple la condición de linealidad en los parámetros. En particular, la mayoría de cinéticas enzimáticas se caracterizan por seguir el modelo cinético enzimático de Michaelis-Menten, el cual no es un modelo lineal en los parámetros.

Bilinealidad

El concepto de bilinealidad es fácilmente extrapolable del concepto de linealidad. La única diferencia es que ahora, la relación lineal se debe cumplir en las dos direcciones de la matriz bidimensional (parámetros y variables). Un ejemplo claro es un modelo PCA de dos componentes principales para una matriz \mathbf{M} ($I \times J$) (para más detalles de modelos PCA y de lo que entendemos por componente principal ver sección 3.4). Cada elemento m_{ij} de esta matriz puede ser descompuesto de la siguiente manera:

$$m_{ij} = t_{i1}p_{j1} + t_{i2}p_{j2} + e_{ij}; \quad i=1, \dots, I; j=1, \dots, J \quad (3.3)$$

La ecuación 3.3 será considerada un modelo bilineal de \mathbf{M} cuando t_{i1} y t_{i2} sean lineales (manteniendo constantes p_{j1} y p_{j2}) y también lo sean p_{j1} y p_{j2} (manteniendo constantes t_{i1} y t_{i2}). Esta consideración se mantiene cuando el modelo PCA contiene más de dos componentes principales [**Kruskal 1984**].

Trilinealidad

De la misma manera podemos extender el concepto de bilinealidad a trilinealidad. En este caso, descomponemos cada elemento m_{ijk} de la matriz \mathbf{M} ($I \times J \times K$) de la siguiente manera:

$$m_{ijk} = a_{i1}b_{j1}c_{k1} + a_{i2}b_{j2}c_{k2} + \dots + e_{ijk}; \quad i=1, \dots, I; j=1, \dots, J; k=1, \dots, K \quad (3.4)$$

Si se cumplen las mismas premisas de linealidad que en los casos anteriores, podremos decir que m_{ijk} es trilineal en los parámetros a , b y c . Un ejemplo de estos modelos es el Análisis Paralelo de Factores (PARAFAC) que se estudiará más adelante en este capítulo (ver sección 3.6). La noción de trilinealidad se puede extender a cuatrilinealidad, etc. La idea básica es la misma.

3.3.2.2. Rango matemático–rango químico. Pseudo-rango. Fuentes químicas de variación.

Rango de matrices bidimensionales

El concepto “rango químico de un sistema” puede definirse como el número de fuentes que hagan variar la señal química y que sean distinguibles de la señal del ruido. Esta idea coincide con la definición matemática de pseudo-rango. Las fuentes de variación no siempre están relacionadas con variación química del sistema. Muchas veces la señal química se ve afectada por efectos físicos, por lo que el rango químico será superior al número de especies.

Aplicar el concepto matemático de rango al modelado de datos químicos a menudo no es muy conveniente [Smilde 2004]. Supongamos, por ejemplo, que tenemos una matriz \mathbf{X} ($I \times J$) formada por números aleatorios, cuyas filas y columnas sean linealmente independientes. Si los números son realmente aleatorios, una manera sencilla de definir el rango matemático de \mathbf{X} es decir que el rango es la mínima dimensión de la matriz en filas, o en columnas.

Ahora supongamos que esa matriz \mathbf{X} está compuesta por 10 espectros UV-Vis. correspondientes a diferentes concentraciones de la misma especie. Cada espectro se compone de 50 longitudes de onda. Si se cumple la ley de Beer-Lambert, la matriz

Capítulo 3

resultante \mathbf{X} (10 x 50) tendrá, básicamente, rango 1. Debido al error experimental, sin embargo, el rango matemático de esta matriz es 10. Podemos escribir \mathbf{X} como:

$$\mathbf{X} = \mathbf{c}\mathbf{s}' + \mathbf{E} \quad (3.5)$$

En la ecuación 3.5 el vector \mathbf{c} (10 x 1) contiene las concentraciones de la especie en las 10 muestras y \mathbf{s} (50 x 1) contiene el espectro puro (absortividad molar x camino óptico) de la especie absorbente. \mathbf{E} (10 x 50) es la matriz asociada al ruido. La ecuación 3.5 se puede escribir como:

$$\mathbf{X} = \hat{\mathbf{X}} + \mathbf{E} \quad (3.6)$$

$\hat{\mathbf{X}}$ tiene rango matemático 1 al representar la variación sistemática de \mathbf{X} . Diremos que el pseudo-rango de \mathbf{X} viene definido por el rango de $\hat{\mathbf{X}}$. Este concepto de pseudo-rango es lo que los químicos entendemos por rango químico de un sistema.

En otros casos, por el contrario, tendremos sistemas donde la absorbancia de una especie se mantenga constante. Esto provocará que el rango químico sea inferior al número de especies químicas en nuestro sistema. Estaremos en un caso de deficiencia de rango químico. Esta deficiencia de rango genera problemas a la hora de resolver el sistema mediante determinadas técnicas de resolución de curvas. Se ha de señalar que cada vez que se hable de rango en esta Memoria se referirá al pseudo-rango matemático o al rango químico.

Rango de matrices tridimensionales

Definir y estudiar el rango matemático de una matriz tridimensional es mucho más complicado que el de una matriz bidimensional. Se han publicado muchos trabajos al respecto [Bro 1998], pero quizás, para explicarlo de una manera sencilla nos quedaremos con la definición dada por Smilde [Smilde 2004]: El rango de una matriz tridimensional $\underline{\mathbf{X}}$ es el mínimo número de componentes trilineales que se necesitan para definir

perfectamente \underline{X} . Esta definición coincide con la definición de pseudo-rango o rango químico para matrices bidimensionales.

De la misma manera que en las matrices bidimensionales, a la hora de calcular el rango químico de una matriz tridimensional podemos tener el mismo problema de deficiencia de rango. El cálculo del rango de matrices tridimensionales se estudiará cuando hablemos del Análisis Paralelo de Factores, PARAFAC (ver sección 3.6).

3.4. TRANSFORMACIÓN DE UNA MATRIZ TRIDIMENSIONAL EN BIDIMENSIONAL. TÉCNICAS *UNFOLDING* DE FACTORIZACIÓN

Dos de las metodologías más conocidas para el análisis multivariable de datos de segundo orden son el Análisis en Componentes Principales, PCA (*Principal Components Analysis*), para visualización y clasificación previa de los datos; y la Regresión Parcial por Mínimos Cuadrados, PLS-R (*Partial Least Square Regression*) para cuantificación [De Jong 1993, Geladi 1988, Helland 1988, Höskuldsson 1988, Phatak 1997, Jonson 2002, Kruskal 1978, Stanimirova 2004, Wold 1987b].

Hasta hace unos años, estas metodologías no eran aplicables a datos tridimensionales, por lo que los datos se tenían que adaptar a una estructura bidimensional compatible. Para hacer compatible la estructura tridimensional a los algoritmos PCA y PLS surgieron las denominadas técnicas de desdoblamiento o *unfolding* [Wold 1987]. Pondremos un ejemplo para visualizar el objetivo de las técnicas *unfolding*. Monitorizando una reacción enzimática, para una muestra \mathbf{M} ($J \times K$), desdoblado la matriz se obtiene un tensor m ($1 \times JK$), si el desdoblamiento se realiza por filas; o un tensor m ($JK \times 1$), si el desdoblamiento se realiza por columnas (figura 3.2).

Por lo que de esta manera, si tenemos I muestras, podremos analizar esta nueva matriz \mathbf{M} ($I \times JK$) con las metodologías clásicas de análisis PCA y PLS (figura 3.3).

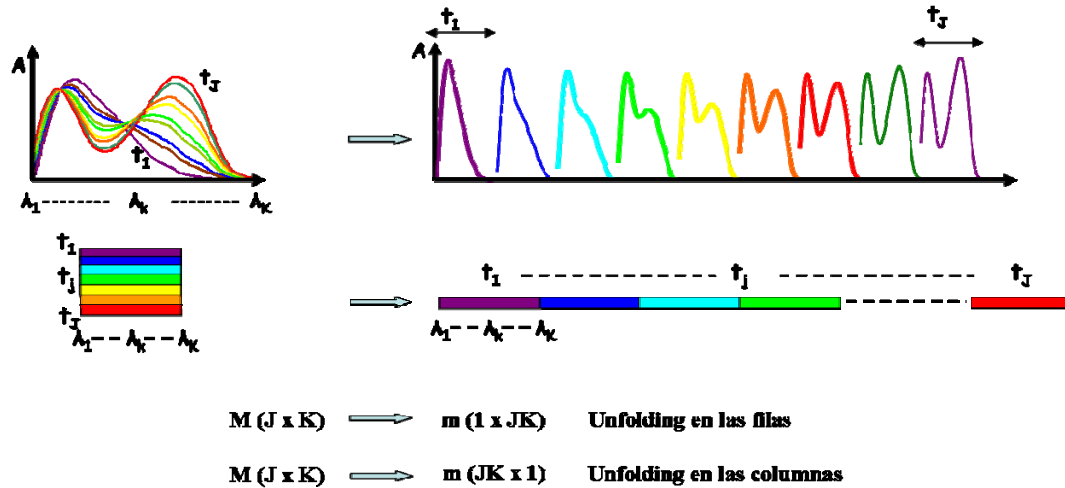


Figura 3.2: Desdoblamiento de una matriz bidimensional en un tensor de primer orden, o vector.

Si tenemos I muestras, obtendremos una matriz bidimensional como la que muestra la figura 3.3.

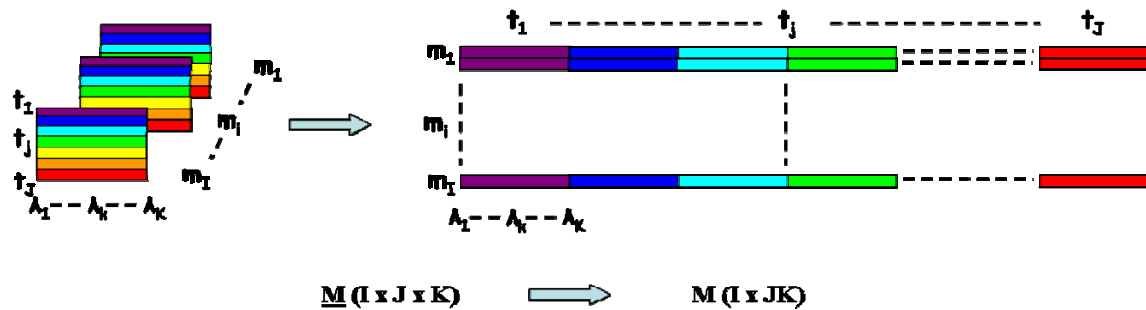


Figura 3.3: Obtención de una matriz bidimensional desdoblada a partir de una matriz tridimensional.

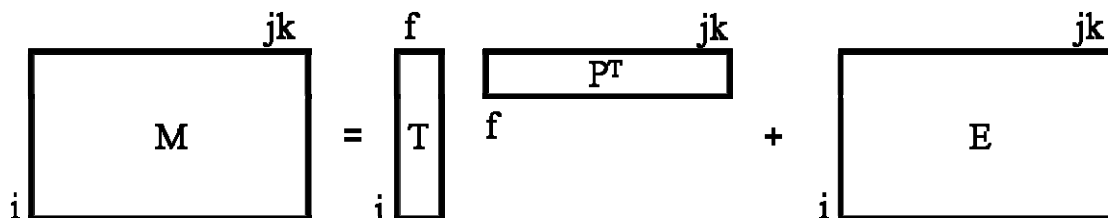
3.4.1. UNFOLDING-PCA Y UNFOLDING-PLS

La forma en la que el PCA y el PLS descomponen la matriz M es la misma que si se tratase de un sistema bidimensional convencional. El análisis en componentes principales (*Principal Components Analysis*) PCA, es una técnica de reducción de variables que condensa el gran número de variables, correlacionadas en mayor o menor medida, en un número pequeño de parámetros representativos ortogonales (no correlacionados) denominados componentes principales, factores o variables latentes.

Los componentes principales recogen la máxima variabilidad existente entre las diferentes muestras y sus variables (espectros desdoblados). Cada uno de ellos se relaciona con diferentes variabilidades fisico-químicas del sistema, por lo que el PCA es una técnica muy útil a la hora de visualizar el comportamiento cinético de nuestro sistema así como la existencia de cualquier interferencia. Matemáticamente hablando, la matriz \mathbf{M} ($I \times JK$) se descompone en el producto de dos matrices (ecuación 3.7)

$$\mathbf{M} = \mathbf{TP}^T + \mathbf{E} \tag{3.7}$$

La primera matriz, \mathbf{T} ($I \times F$), es la matriz de *scores*. Esta matriz está relacionada con la variabilidad en las muestras. La segunda matriz, \mathbf{P} ($JK \times F$), es la matriz de *loadings* y está relacionada con las variables. La matriz \mathbf{E} ($I \times JK$) es la matriz de residuales. El superíndice “T” indica que la matriz es transpuesta. La figura 3.4 representa la descomposición de cada elemento m_{ijk} , donde f representa el número de Componentes Principales.



$$m_{ijk} = \sum_{f=1}^F t_{if} p_{jkf}^T + e_{ijk}$$

Figura 3.4: Representación gráfica de expresión matricial de una descomposición PCA

Existen diferentes algoritmos de cálculo para obtener \mathbf{T} y \mathbf{P} , siendo uno de los más conocidos el NIPALS (*Nonlinear Iterative Partial Least Squares*) [WoldH 1966].

Análogamente, podríamos hablar de la regresión por mínimos cuadrados parciales clásica aplicada a la matriz desdoblada \mathbf{M} , *unfolding*-PLS. En este caso, lo que se busca

es la creación de modelos de regresión robustos para poder predecir nuevas muestras. La diferencia con PCA es que ahora se crea un modelo de regresión buscando la máxima varianza explicada entre la matriz desdoblada \mathbf{M} (como en la figura 3.3) y la correspondiente matriz de la propiedad que se quiere cuantificar \mathbf{Y} (por ejemplo, concentración inicial de sustrato).

La utilización del PCA y PLS con matrices desdobladas conlleva ciertas desventajas, ya que se pierde la estructura tridimensional de los datos y los datos son tratados de forma bidimensional. Los principales efectos son:

- Al desdoblar la matriz \mathbf{M} se inserta una correlación inexistente entre el final de un espectro y el inicio del siguiente.
- Los modelos de calibración *unfolding*-PLS son menos robustos.
- La interpretación visual de los resultados es complicada.

3.5. 3W-PCA Y 3W-PLS

Con el fin de eliminar los problemas derivados de trabajar con matrices desdobladas y aplicar estas técnicas directamente a la matriz tridimensional se han generado diversos algoritmos, tales como el *three-way* PCA (3W-PCA) y el *three-way* PLS (3W-PLS), que son capaces de aplicar los mismos fundamentos teóricos que sus respectivos bidimensionales [Menrion 1994, Bro 1996, Smilde 1997, De Jong 1998].

3.5.1. 3W-PCA

Existen diversas maneras de aplicar 3W-PCA a la matriz $\underline{\mathbf{M}}$ [Wold 1987, Bro 1996]. En esta tesis estudiaremos la manera más sencilla. La matriz $\underline{\mathbf{M}}$ se descompone de forma análoga a la realización de un desdoblamiento de la matriz. Pero ahora, el desdoblamiento se realiza en cada una de las direcciones de $\underline{\mathbf{M}}$ (figura 3.5).

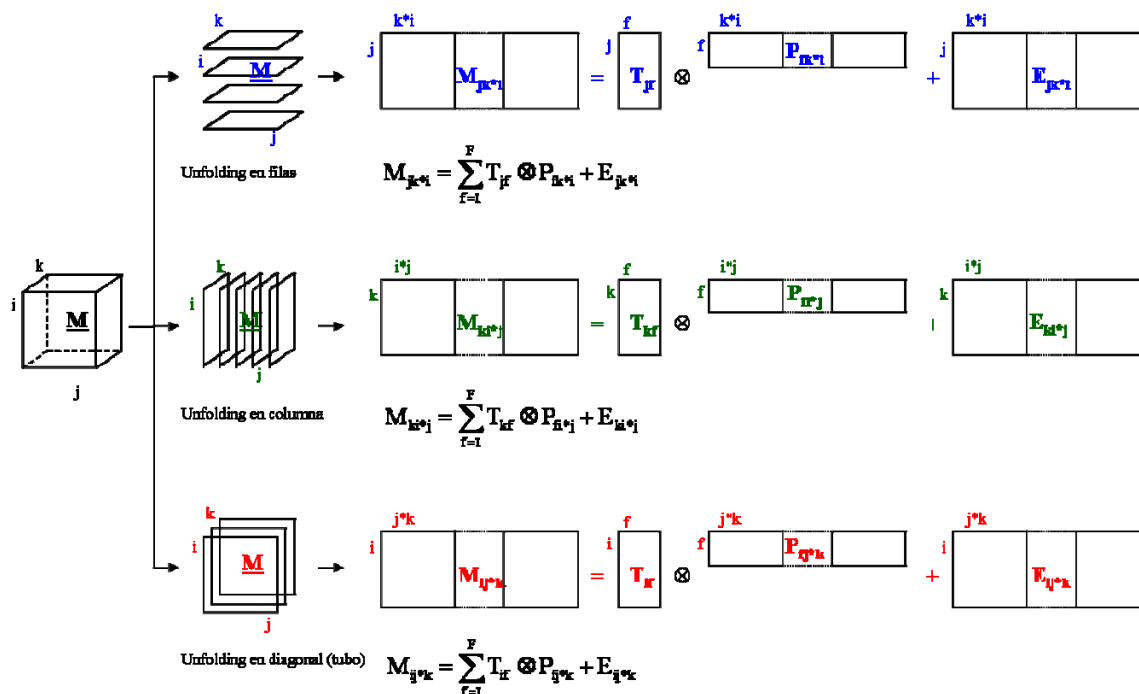


Figura 3.5: Aplicación del 3W-PCA a la matriz $\underline{\mathbf{M}}$. \otimes representa el producto de Kronecker [Smilde 2004].

El algoritmo descompone la matriz tridimensional como la suma de la multiplicación de las matrices de *scores* \mathbf{T} obtenidas para cada un de los modos i , j y k por las matrices de *loadings* \mathbf{P} , teniendo en cuenta el número apropiado de factores \mathbf{F} . De esta manera se consigue obtener una matriz de *scores* para cada una de las direcciones del desdoblamiento. Las matrices de *scores* obtenidas (\mathbf{T}_{if} , \mathbf{T}_{jf} , \mathbf{T}_{kf}) nos dan una importante información cualitativa sobre la variabilidad en la evolución de las concentraciones, en los espectros y en las muestras, respectivamente.

Esta manera de aplicar 3W-PCA a una matriz tridimensional no es más que una extensión del *unfold*-PCA para los tres modos. La principal ventaja del análisis 3W-PCA radica, precisamente, en la interpretabilidad de estas matrices de *scores*, aportando mucha más información que un análisis de PCA bilineal.

3.5.2. 3W-PLS

En 1996 Rasmus Bro [Bro 1996] desarrolló un algoritmo, N-PLS (regresión parcial por mínimos cuadrados de n dimensiones), que ampliaba la aplicación del PLS a sistemas multidimensionales sin necesidad de convertir la matriz en bidimensional con técnicas de desdoblamiento. En la literatura existe cierta confusión acerca de este algoritmo, ya que N-PLS se ha referido también a menudo a PLS no lineal (*non linear PLS*). Por lo que en esta Memoria se hará referencia al 3W-PLS.

La descomposición de la matriz $\underline{\mathbf{M}}$ mediante 3W-PLS es análoga a PARAFAC (ver sección 3.6) [Bro 1996, Bro 1997]. La matriz $\underline{\mathbf{M}}$ se descompone en un conjunto de cubos de manera óptima y que, a diferencia de PARAFAC, el rango (número de componentes) puede ser superior que el número de especies absorbentes (figura 3.6).

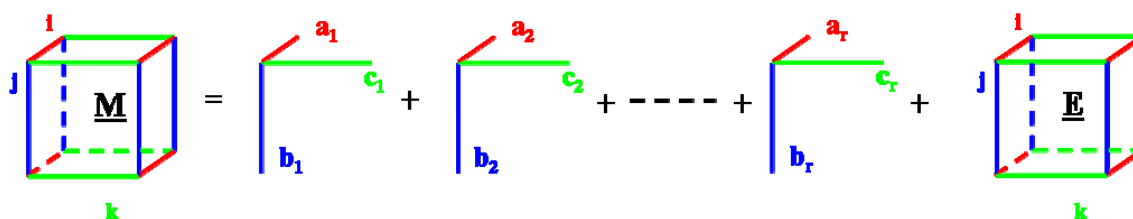


Figura 3.6: Descomposición 3W-PLS de una matriz $\underline{\mathbf{M}}$ ($I \times J \times K$) vista como la suma de cada uno de los componentes r en los que se ha descompuesto la matriz.

La filosofía es la misma que la del PLS bilineal. Se descompone la matriz de tal manera que los vectores de *scores* y *loadings* producidos describen la máxima covarianza de las variables dependientes e independientes (figura 3.6). Matemáticamente hablando, no existen diferencias entre los *scores* y los *loadings* porque ambos son calculados de la misma manera. Por eso, $\underline{\mathbf{M}}$ se define con las matrices \mathbf{A} ($I \times R$), \mathbf{B} ($J \times R$) y \mathbf{C} ($K \times R$), cuyos elementos son a_{ir} , b_{jr} y c_{kr} respectivamente. R es el número de componentes. El algoritmo usado en esta memoria viene descrito adecuadamente en las referencias [Bro 1996, Smilde 1997, De Jong 1998, Bro 1998].

Una ventaja importante del 3W-PLS, al igual que el 3W-PCA, es la fácil interpretación de los *scores* y los *loadings* cuando hay un gran número de variables [Crouch 2000]. El algoritmo es rápido debido a la poca cantidad de parámetros a calcular y al hecho de que el problema se reduce a una descomposición en valores propios [Bro 1996]. Otra ventaja de este método de calibración es la posibilidad de estabilización de la solución en caso de baja contribución neta del analito, debido a la búsqueda de un modelo que explique la máxima covarianza entre las variables dependientes e independientes. Pero, como era de esperar, hay ciertas desventajas, la principal de las cuales es la pérdida de ajuste comparada con una metodología bilineal.

El algoritmo 3W-PLS se puede extender a 3W-PLS2 de la misma manera que su análogo bilineal PLS2. Los algoritmos PLS2 (bi y tri dimensional) se caracterizan porque se modelan varias variables dependientes al mismo tiempo. Ahora, **M** se descompone produciendo vectores de *scores* y *loadings* que tienen la máxima covarianza con la parte no explicada de todas las variables dependientes [Amigo 2005]. Este hecho hace que se tengan ciertas ventajas en la modelización de posibles correlaciones o co-linealidades entre las variables dependientes.

3.6. ANÁLISIS PARALELO DE FACTORES. PARALLEL FACTOR ANALYSIS. PARAFAC

3.6.1. INTRODUCCIÓN

El Análisis Paralelo de Factores, PARAFAC (*PARAllel FACtor Analysis*) es un método de descomposición de matrices tridimensionales que puede ser comparado al PCA bilineal [Bro 1997, Bro 1998]. Tiene su origen en la psicometría y fue propuesto independientemente por Harshman y Carrol & Chang en 1970 [Harshman 1970, Carrol 1970]. Éstos últimos lo denominaron CANDECOMP, descomposición canónica (*CANonical DECOMPosition*).

Básicamente, PARAFAC es un método de resolución de curvas que descompone la matriz tridimensional $\underline{\mathbf{M}}$ ($I \times J \times K$) en tres sub-matrices bidimensionales \mathbf{A} ($I \times R$), \mathbf{B} ($J \times R$), \mathbf{C} ($K \times R$), tal y como se muestra en la figura 3.7 [Andersen 2003, Bro 1997b]. En esta figura cada color representa a una de las dimensiones de la matriz.

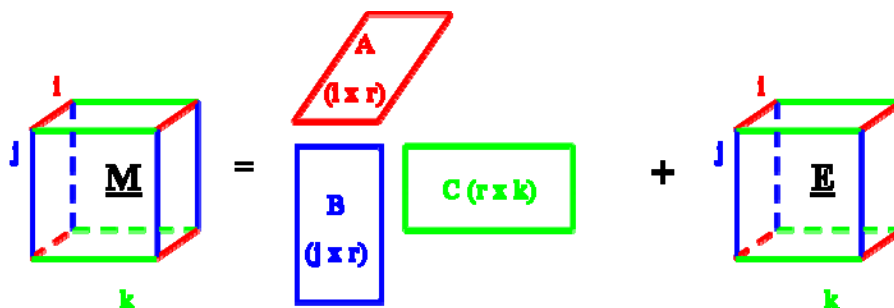


Figura 3.7: Descomposición PARAFAC de una matriz $\underline{\mathbf{M}}$ ($I \times J \times K$) vista como el producto de las tres matrices de *loadings* obtenidas.

En la descomposición PARAFAC nos referimos al concepto de componentes R y no a componentes principales F . Es importante destacar esta diferencia por lo que se ha descrito con detalle en la sección 3.7. Cada componente R viene definido por tres vectores de *loadings*, los cuales representan o indican el comportamiento de los componentes \mathbf{R} en cada una de las tres dimensiones (modo de concentración, *loading* de concentración; modo cinético, *loading* cinético; modo espectral, *loading* espectral) (figura 3.7).

Matemáticamente podemos decir que la descomposición PARAFAC busca el mejor modelo trilineal que minimice la suma de cuadrados de los residuales, e_{ijk} en el modelo (ecuación 3.8):

$$m_{ijk} = \sum_{f=1}^F a_{ir} b_{jr} c_{kr} + e_{ijk} \quad (3.8)$$

Este método ha ganado interés en la resolución de sistemas dinámicos y, en especial en sistemas de fermentación monitorizados con fluorescencia multidimensional

por su simplicidad y la información cualitativa que aporta. Tiene ciertas ventajas que lo hacen preferible a un modelo 3W-PCA:

- Propiedad de unicidad de la solución (*uniqueness*). La solución es única. El cálculo de los *loadings* de cada componente \mathbf{R} en PARAFAC se realiza sin imposición de la restricción de ortogonalidad, tal y como se realiza en MCR-ALS, por lo que los *loadings* pueden ser asociados a las fuentes que provocan la variación en cada modo.
- Poder de visualización: Los *loadings* obtenidos con PARAFAC aportan una mayor información visual que los *scores* y *loadings* obtenidos mediante 3W-PCA.
- Posibilidad de trabajar con más de tres dimensiones.

Pero también tiene ciertas desventajas. Los componentes \mathbf{R} no están escalonados como en 3W-PCA. Debido a la libertad rotacional de la que goza PARAFAC, los componentes deben ser identificados después de la construcción del modelo. Por esta razón, la correcta determinación del número de componentes es un parámetro crítico y crucial. Pero quizás la desventaja más importante de los modelos PARAFAC es que la estructura de los datos debe ser OBLIGATORIA Y ABSOLUTAMENTE trilineal. Este aspecto es un importante handicap que limita su correcta utilización. Por ejemplo, la utilización de los modelos PARAFAC para cinéticas enzimáticas registradas con un espectrofotómetro DAD-UV-Vis nos puede llevar a resultados y conclusiones erróneas en nuestro sistema, ya que el modo cinético no sigue un comportamiento lineal. En cambio, donde PARAFAC demuestra todo su potencial es en la monitorización de procesos de fermentación a través de la fluorimetría multidimensional, siendo uno de los campos donde su implantación se está llevando con mayor éxito (ver sección 3.8).

3.6.2. APLICACIÓN DE PARAFAC

Los pasos para la aplicación de PARAFAC son los mismos que para MCR-ALS, siendo puntos fundamentales la determinación de los componentes del sistema y la aplicación de restricciones a la convergencia mediante el algoritmo ALS.

3.6.2.1. Número de componentes en PARAFAC

Decidir el número de componentes de un modelo PARAFAC es mucho más fácil que para problemas bidimensionales, debido a la propiedad de solución única que posee el algoritmo. El análisis de residuales, varianza explicada por el modelo y el análisis de la llamada “consistencia del núcleo”, *core consistency*, son las metodologías más utilizadas.

La “consistencia del núcleo” [Bro 1998] mide la robustez de la descomposición matemática comparando los resultados obtenidos con la matriz original. Es un parámetro muy útil a la hora de discernir si se ha obtenido una solución degenerada o un modelo inestable y poco fiable.

Cuando se decide el número de componentes para un modelo PARAFAC es recomendable el estudio conjunto de la evolución de la varianza explicada por el modelo y la consistencia del núcleo respecto al número de componentes **R**.

3.6.2.2. Iniciando PARAFAC. Estimaciones iniciales y restricciones.

Al igual que en MCR-ALS, contar con buenas estimaciones iniciales ayuda a que el algoritmo encuentre el mínimo absoluto de convergencia. A este respecto se han propuesto diversas maneras de inicialización. Harshman y Lundy [Harshman 1984 y 1994] propusieron el uso de valores aleatorios y varios puntos de inicio del algoritmo simultáneamente. Otra manera de obtener estimaciones iniciales es usar los vectores propios de un análisis SVD previo (ver sección 3.7). Diversos autores [Bro 1998] proponen el uso de valores iniciales obtenidos por diversas metodologías como una Descomposición Trilinear, TLD (*Trilinear Decomposition*) o Métodos Generalizados de Anulación de Rango, GRAM (*generalized rank annihilation method*).

También son de gran utilidad las restricciones naturales con las que se puede ayudar a la convergencia al algoritmo. El fundamento de estas restricciones se puede ver

en la sección 3.7, siendo la nonegatividad y la unimodalidad las dos restricciones más utilizadas.

3.6.2.3. Algoritmo ALS en PARAFAC

La solución al modelo PARAFAC puede ser encontrada por el algoritmo *Alternating Least Squares* ALS [Bro 1998] utilizando iteraciones sucesivas en un proceso análogo al MCR-ALS. El proceso general puede ser descrito como:

- 1) Inicialización de **B** y **C**.
- 2) Estimación de **A** a partir de **M**, **B** y **C** por mínimos cuadrados.
- 3) Estimación de **B** y **C** a partir de la **A** estimada y **M** de la misma manera.
- 4) Cálculo de la nueva matriz **M** a partir de las nuevas matrices **A**, **B** y **C**.
- 5) Cálculo de los residuales y ver criterio de convergencia.
- 6) Se prosigue con el paso 2 hasta la convergencia.

La convergencia se consigue con un cambio pequeño en el ajuste o en los *loadings* entre dos iteraciones sucesivas. A veces sucede que pequeños cambios en el ajuste pueden estar relacionados con grandes diferencias en los *loadings* [Harshman 1970]. Para evitar esto, un criterio de parada bastante común es detener las iteraciones cuando el cambio relativo en el ajuste entre iteraciones es inferior a un cierto valor (por ejemplo 10^{-6}). En algunas ocasiones se utiliza como criterio de parada que el cambio en los *loadings* para dos iteraciones sucesivas sea muy pequeño [Tu 1992].

3.7. RESOLUCIÓN MULTIVARIABLE DE CURVAS. *MULTIVARIATE CURVE RESOLUTION (MCR-ALS)*

3.7.1. INTRODUCCIÓN.

La resolución Multivariable de Curvas-Mínimos Cuadrados Alternos, MCR-ALS (*Multivariate Curve Resolution-Alternating Least Square*) quizás sea la metodología de resolución de curvas que mayor aplicabilidad ha demostrado a la hora de resolver problemas en los más diversos sistemas evolutivos [De Juan 2000, De Juan 2000b, De Juan 2003, Tauler 1995, Tauler 1995b]. De una manera sencilla permite obtener una visión cualitativa o, si se quiere, semi-cuantitativa de nuestro sistema en estudio sin necesidad de conocer las leyes enzimáticas por las que se rigen. Por esta razón se les denominan métodos de resolución de curvas de modelado blando (*Soft Modelling*) [De Juan 2000]. En nuestro caso, tratándose de medidas espectrofotométricas, si se asume que la absorbancia a cada longitud de onda es la suma de cada componente de la mezcla (sistema bilineal), la matriz M puede ser descompuesta tal y como indica la figura 3.8 [Rodríguez-Rodríguez 2007].

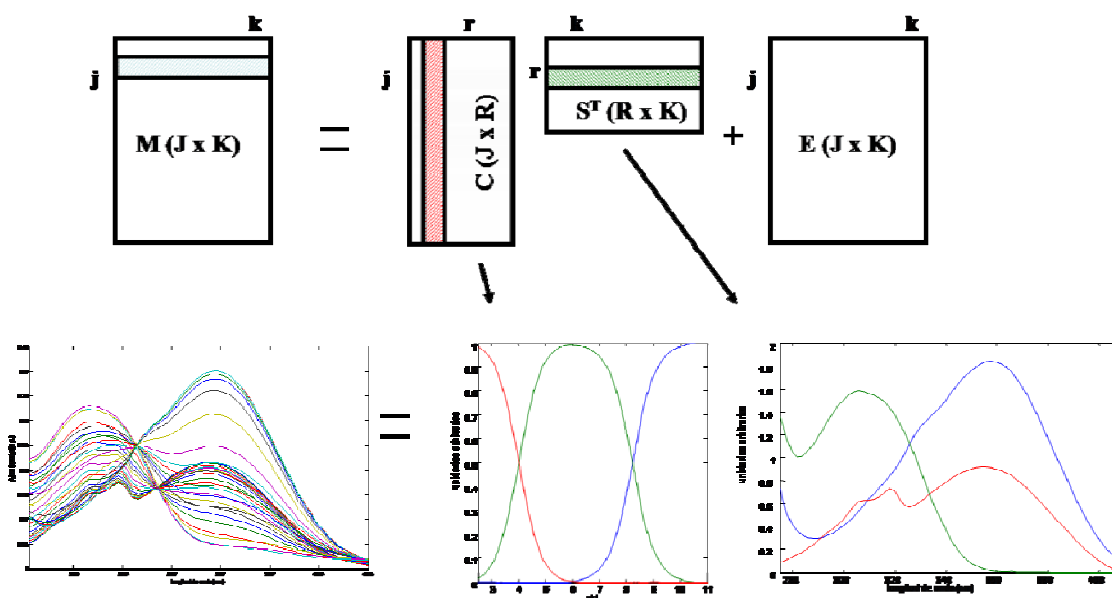


Figura 3.8: Resolución Multivariante de Curvas MCR-ALS aplicada a la descomposición de la matriz M . Evolución del espectro del ácido 8-hidroxiquinolina-5-sulfónico con el pH [Rodríguez-Rodríguez 2007].

Matemáticamente, cada elemento de la matriz \mathbf{M} se descompone en dos sub-elementos (ecuación 3.9)

$$m_{jk} = \sum_{r=1}^R c_{jr} s_{kr}^T + e_{jk} \quad (3.9)$$

De forma global, podremos decir que la matriz \mathbf{M} queda descompuesta en dos sub-matrices (ecuación 3.10)

$$\mathbf{M} = \mathbf{C}\mathbf{S}^T + \mathbf{E} \quad (3.10)$$

\mathbf{C} ($J \times R$) y \mathbf{S} ($K \times R$) son las sub-matrices que corresponden a los perfiles de concentración y los perfiles espectrales de cada uno de los \mathbf{R} componentes, respectivamente. La matriz \mathbf{E} ($J \times K$) es la matriz de residuales.

MCR resuelve de forma iterativa la ecuación 3.10 mediante un algoritmo de mínimos cuadrados alternos, ALS (*Alternating Least Squares*), siguiendo los pasos que se pueden observar en la figura 3.9:

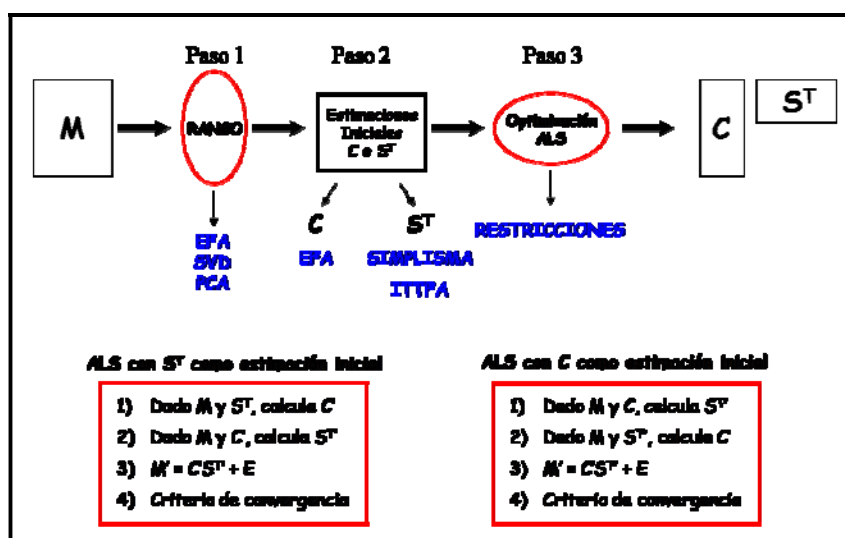


Figura 3.9: Pasos en una descomposición MCR-ALS de la matriz \mathbf{M} ($J \times K$)

Esta descomposición se realiza para el número de componentes previamente propuesto y usando estimaciones iniciales de C o S^T .

A la hora de afrontar un análisis MCR-ALS se deben tener en cuenta ciertos problemas que pueden surgir y que pueden hacer que los resultados sean matemáticamente perfectos pero químicamente incoherentes. No se debe de olvidar nunca que se aplican metodologías matemáticas a sistemas que evolucionan debido a fenómenos físico-químicos. Siempre tiene que haber un consenso entre los resultados matemáticos obtenidos y los efectos físico-químicos que se producen en el sistema. En este sentido pueden surgir dos tipos de problemas que hay que tener presentes: problemas del tipo químico y del tipo matemático. Para la resolución de estos problemas y llevar a cabo una buena optimización MCR-ALS se debe pasar necesariamente por los tres pasos indicados en la figura 3.9.

3.7.2. PASO 1: DETERMINACIÓN DEL NÚMERO DE COMPONENTES. SISTEMAS DEFICIENTES DE RANGO.

Uno de los puntos fundamentales en un análisis de resolución de curvas es la determinación del número de componentes que conforman el sistema. Es importante destacar la diferencia conceptual que existe entre Componente Principal \mathbf{F} y componente \mathbf{R} . El concepto de Componente Principal es más matemático que químico. Como ya hemos visto con anterioridad, los Componentes Principales \mathbf{F} están restringidos a ser ortogonales entre sí, por lo que no ofrecen una descripción real de la evolución del sistema, sino una idea intuitiva. Los componentes \mathbf{R} no están obligados a ser ortogonales entre sí. Están directamente ligados directamente a una entidad físico-química del sistema que hacen variar la señal de forma significativamente diferente del ruido. Por lo que el sentido físico-químico del componente toma mayor importancia. Se espera que cada componente \mathbf{R} explique una respuesta real de nuestra cinética. Esto incluye ejemplos tan

diversos como el espectro de los componentes químicos absorbentes de la reacción o posibles fuentes de contaminación que hacen variar la señal de forma significativa.

Para determinar el número de componentes \mathbf{R} existen diversas metodologías [Otto 1999, Tauler 1997, Sharaf 1986]. Tres de las más empleadas son el PCA, la descomposición en valores singulares (*Singular Value Decomposition*) SVD, y el análisis de factores evolutivos. Con el fin de ilustrar la potenciabilidad de cada uno de los métodos propuestos, se usará un ejemplo sencillo. En la figura 3.10 se muestra los equilibrios ácido-base del ácido 8-hidroxiquinolina-5-sulfónico. Se monitoriza la evolución de su espectro UV-Vis con la variación del pH. El ejemplo se describe con detalle en capítulos sucesivos de esta Memoria [Rodríguez- Rodríguez 2007]. Se trata de un sistema diprótico en la que las tres especies involucradas en los equilibrios ácido-base absorben en el rango de longitudes de onda seleccionado.

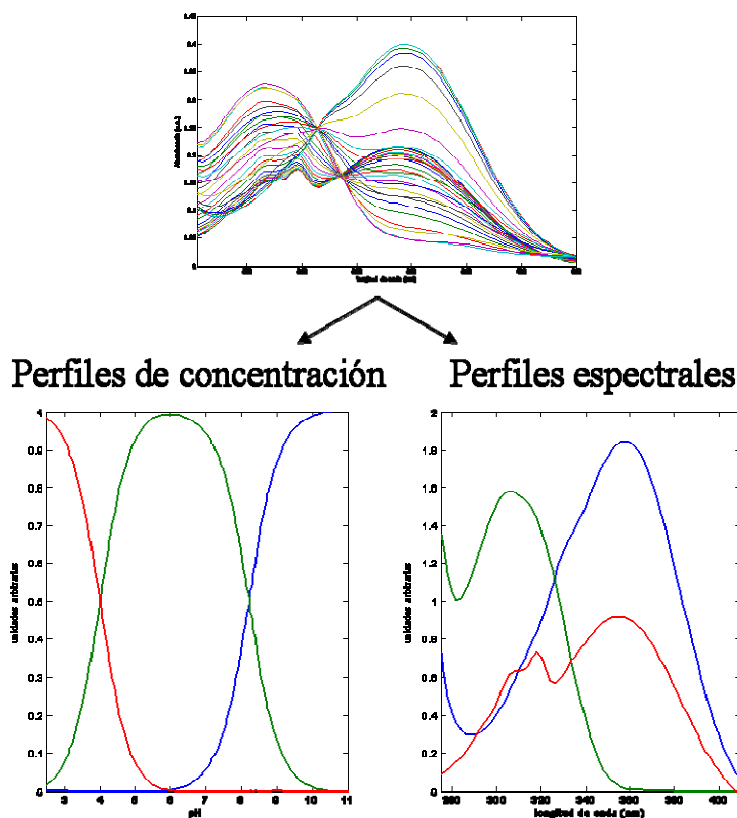


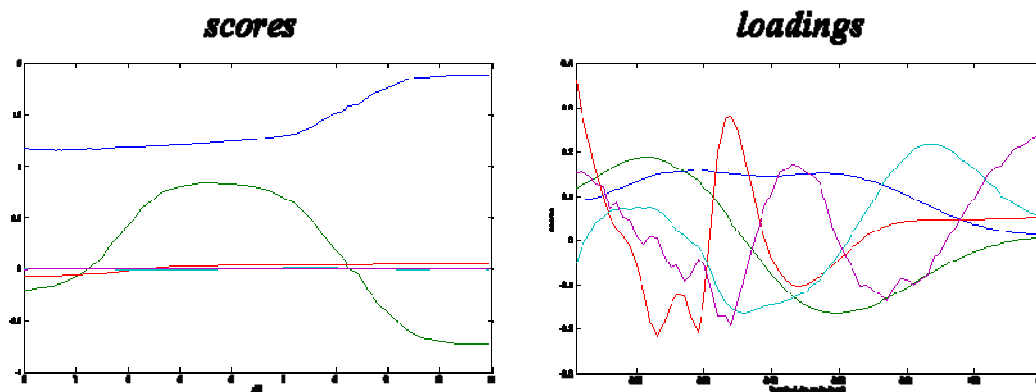
Figura 3.10: Evolución del espectro ácido-base en un sistema diprótico [Rodríguez-Rodríguez 2007].

Observando los espectros, se ve cómo los perfiles del primer compuesto (rojo) y del último (azul) tienen una alta correlación. Sin partir del conocimiento previo del sistema, el distinguir estos dos compuestos puede presentar dificultades, a pesar de contar con un perfil de concentración totalmente diferente.

3.7.2.1. Análisis en Componentes Principales. *Principal Components Analysis*. PCA.

Anteriormente se ha estudiado el análisis PCA como una herramienta cualitativa de visualización de la evolución de un sistema. De esta manera, las matrices de *scores* \mathbf{T} y *loadings* \mathbf{P} obtenidas pueden ser asociadas a la evolución de la concentración y al perfil espectral de cada uno de los \mathbf{F} factores escogidos. La elección del número de factores es complicada, ya que se están obteniendo componentes principales \mathbf{F} y no componentes \mathbf{R} . Un parámetro muy útil es estudiar cómo varía la varianza explicada con el número de componentes principales.

Realizando un PCA en el ejemplo propuesto para 5 componentes principales se obtienen los *scores* y *loadings* representados en la figura 3.11. Los perfiles de los *scores* y *loadings* de los dos primeros factores (azul y verde) tienen una forma definida y continuada. Además, su porcentaje de varianza explicada es elevado. El perfil del *score* en el tercer factor (rojo) no presenta una forma muy diferenciada de los demás *scores*. La varianza explicada por este tercer factor es muy pequeña comparada con los dos primeros. La cuestión es discernir si este factor es significativo o no. En este caso, se sabe que existen tres componentes químicos, pero el PCA no clarifica esta cuestión.



Factor	% Varianza explicada por factor	% Varianza explicada acumulada
1	83.93	83.93
2	15.90	99.82
3	0.18	100
4	0.00	100
5	0.00	100

Figura 3.11: Perfiles de *scores* y *loadings* y porcentaje de varianza explicada por factor en un análisis PCA del ejemplo propuesto.

3.7.2.2. Descomposición en valores singulares. *Singular Value Decomposition. SVD.*

La Descomposición en Valores Singulares (también llamados valores propios), SVD, es una de las metodologías más empleadas en la determinación del número de componentes. La definición matemática de lo que son los valores propios de una matriz puede ser extrapolada a su utilidad en quimiometría. Por lo que se puede decir que un sistema está compuesto de tantos valores propios como analitos linealmente independientes tenga, hablando tanto cinética como espectralmente. Cada valor propio está asociado a un vector propio, vectores propios, en cada uno de los modos. En un análisis SVD, la matriz \mathbf{M} ($J \times K$) se descompone en tres matrices \mathbf{U} ($J \times J$), \mathbf{S} ($J \times K$), \mathbf{V} ($J \times J$) de la forma que indica la ecuación 3.11:

$$X=USV^T \tag{3.11}$$

Siendo las matrices **U** y **V** las matrices de vectores propios “*left*” y “*right*”, matrices ortogonales entre sí (figura 3.12). Estas matrices contienen información similar a las matrices de *scores* y *loadings* en PCA, respectivamente, pero normalizados a longitud uno. La matriz **S** es una matriz diagonal que contiene los valores propios.

A la hora de seleccionar el número de componentes a partir de un análisis SVD se pueden aplicar diversos criterios. Uno de los que aporta mejores resultados es comparar los valores de los valores propios sucesivos para ver cuáles son diferentemente significativos y compararlos con sus correspondientes vectores propios (figura 3.12).

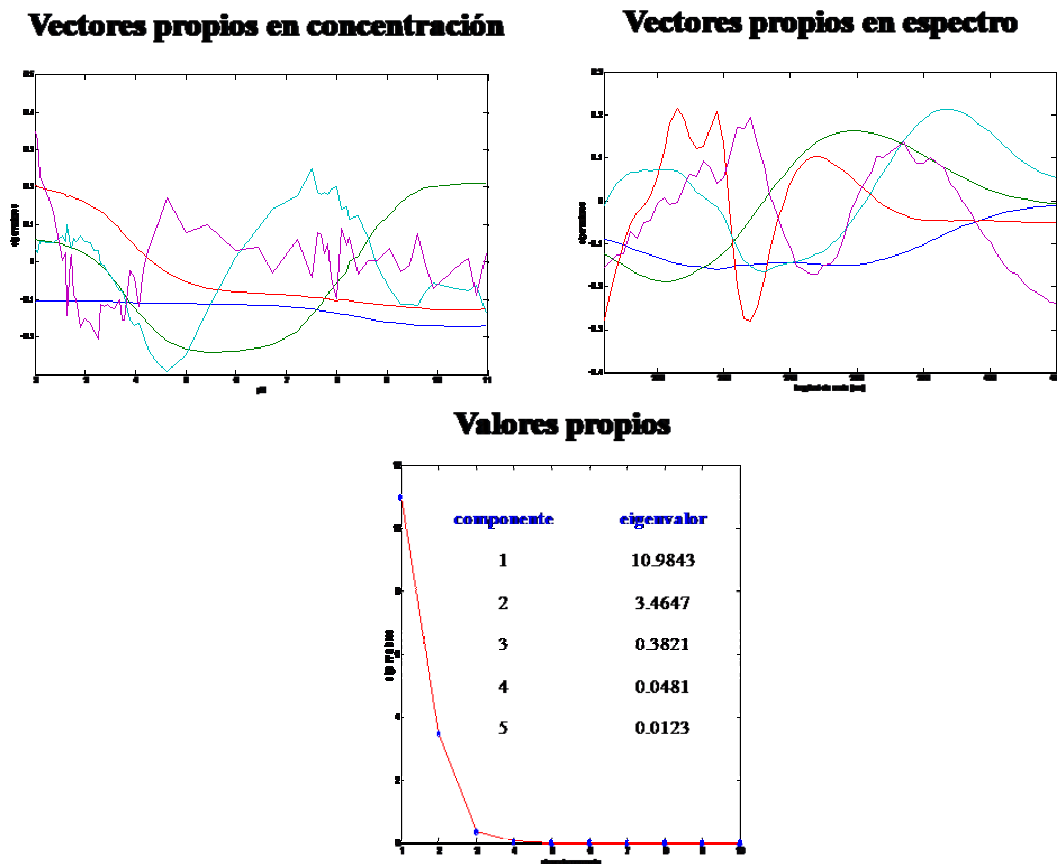


Figura 3.12: Análisis SVD del ejemplo propuesto. En las figuras se pueden observar los cinco primeros valores y vectores propios.

En la figura 3.12 se observa cómo los tres primeros valores propios son significativamente diferentes del resto. Además, los tres primeros vectores propios en concentración (azul, verde y rojo) presentan un perfil definido. Atendiendo a los vectores propios en el espectro se puede observar que el cuarto vector propio (cian) también presenta un perfil definido sin excesivo ruido. Pero su valor propio es pequeño en comparación con los tres primeros. Este análisis confirmaría la existencia de tres componentes absorbentes en nuestro sistema.

3.7.2.3. Análisis de Factores Evolutivos. *Evolving Factor Analysis*, EFA

El método de Análisis de Factores Evolutivos, *Evolving Factor Analysis*, EFA [Gemperline 1989, Maeder 1987], establece y diferencia las regiones de existencia de cada uno de los analitos que hacen variar la señal significativamente del ruido. Esta técnica se basa en la realización de diferentes análisis PCA partiendo del primer espectro y añadiendo sucesivamente un nuevo espectro para cada PCA. El análisis se realiza de principio a fin del proceso (*forward*) y viceversa (*backward*). El resultado final es la unión de la evolución de los valores propios más significativos en cada PCA realizado (figura 3.13).

Una de las propiedades del análisis EFA es que señala las zonas de existencia de los componentes a lo largo de la reacción. En el ejemplo propuesto se puede observar cómo los dos primeros componentes contienen la mayor variabilidad en el sistema, mientras que el tercero (rojo) podría ser considerado como despreciable.

A pesar de que el ejemplo es un sistema sencillo, la gran similitud entre el espectro de la primera y tercera especie (figura 3.12) hace que el análisis del número de componentes sea complicado y no un asunto trivial.

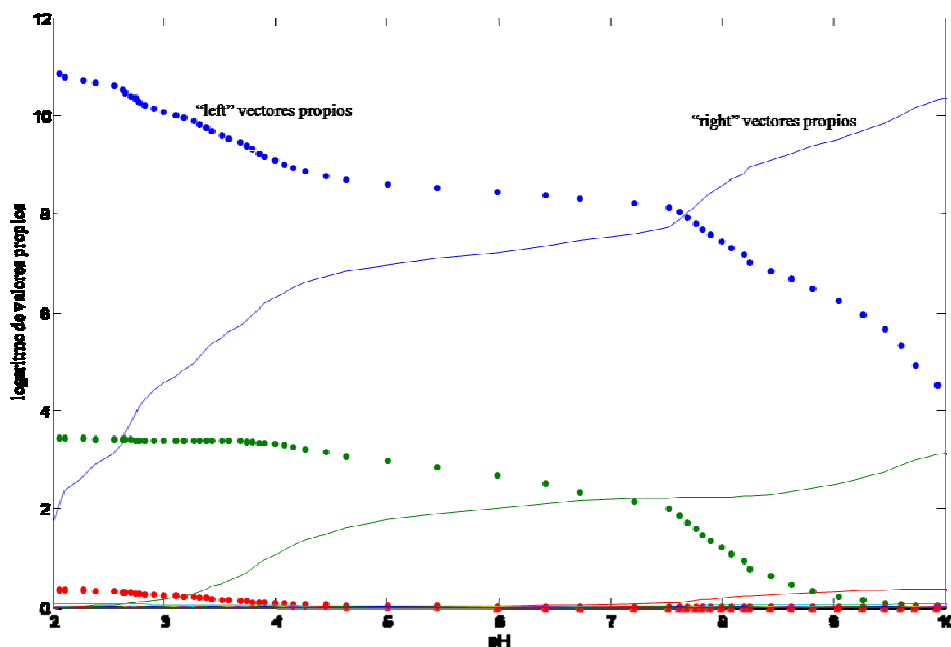


Figura 3.13: Análisis EFA del ejemplo propuesto. Primer componente, azul. Segundo componente, verde. Tercer componente, rojo. Punteado, *Left* vectores propios. Continuo, *right* vectores propios.

3.7.2.4. Sistemas deficientes de rango.

Uno de los problemas más importantes de estas metodologías aplicadas a sistemas enzimáticos es la existencia de interferencias espectrales cuya absorbancia permanece constante a lo largo de la reacción. Por ejemplo, en la figura 3.14 la matriz D1 muestra una reacción enzimática de hipoxantina oxidándose a ácido úrico a través de la catálisis con el enzima xantina oxidasa [Amigo 2006, Amigo 2006b]. La reacción se lleva a cabo en una matriz no absorbente. En esta reacción se conoce que existen tres especies absorbentes (hipoxantina, xantina, ácido úrico). Si esta reacción se lleva a cabo en una matriz absorbente en el mismo rango de longitudes de onda como la orina, D12, el sistema se compondrá de cuatro especies absorbentes. Sin embargo, realizando un estudio del número de componentes, obtendremos que son 3 los componentes significativos en las dos reacciones (figura 3.14).

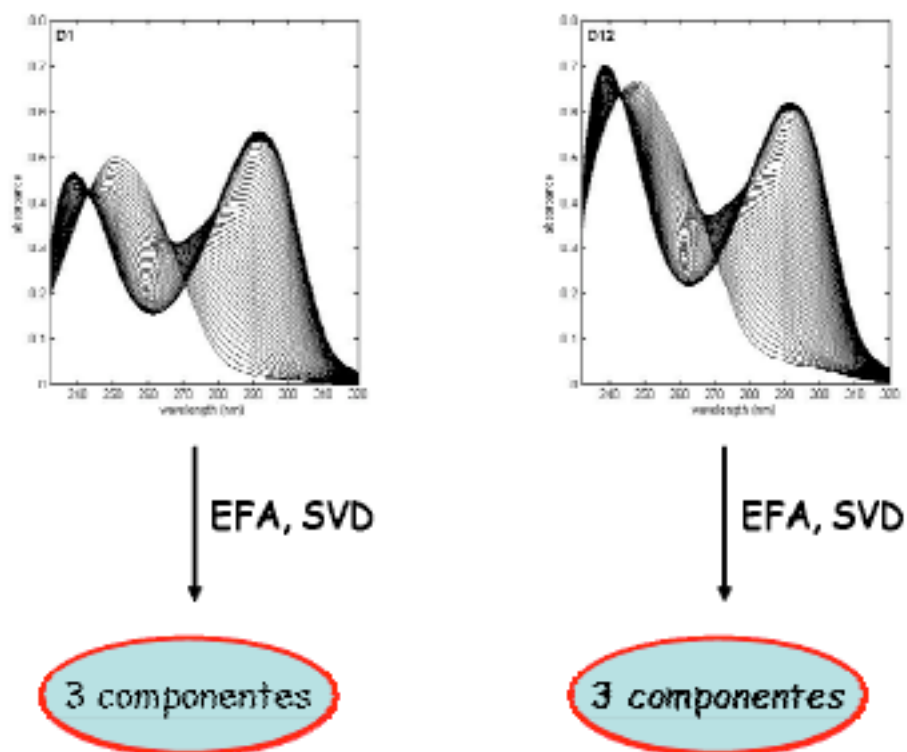


Figura 3.14: Sistema deficiente de rango.

Averiguar si un sistema es deficiente de rango es un problema difícil de resolver si no se tiene cierto conocimiento previo del sistema que se está tratando. Para resolver este problema, una de las metodologías más aplicadas es la concatenación de matrices deficientes de rango con matrices de rango completo. Esta cuestión se estudiará más adelante en este capítulo.

3.7.3. PASO 2: NECESIDAD DE ESTIMACIONES INICIALES.

Todas las metodologías de resolución de curvas necesitan de estimaciones iniciales de los perfiles de concentración o espectrales para empezar la optimización del

algoritmo. Tener buenas estimaciones iniciales ayuda al algoritmo a la convergencia hacia la solución real. Se pueden usar estimaciones iniciales tanto de los perfiles de concentración como de los perfiles espectrales. Es importante que estas estimaciones iniciales no sean muy diferentes a los perfiles reales.

Lo más común en sistemas enzimáticos es usar estimaciones iniciales de los perfiles espectrales. Se pueden usar los espectros inicial y final, y alguno de los intermedios, obtenidos en la monitorización de la reacción. También se pueden usar espectros puros de los analitos involucrados en la reacción si se conocen. Aún así, existen diversas metodologías matemáticas para la obtención de estas estimaciones, tanto en los perfiles espectrales como en los de concentración (SIMPLISMA, *Needle Search*, ITTFA) [Cuesta 1994, Gemperline 1986, Vandeginste 1985, Winding 1988].

3.7.4. PASO 3: OPTIMIZACIÓN ALS. AMBIGÜEDADES.

Aún habiendo hecho un buen análisis del número de componentes y teniendo buenas estimaciones iniciales, la factorización de la matriz \mathbf{M} no está exenta de problemas de tipo matemático. Los más importantes son las ambigüedades a las que está sometida la descomposición matemática.

Las ambigüedades matemáticas implican que la resolución de la matriz \mathbf{M} puede conseguirse, sin perder calidad de ajuste, con infinitos pares de matrices \mathbf{C} y \mathbf{S}^T diferentes.

La ecuación 3.12 refleja la **ambigüedad rotacional**. Para una determinada descomposición, siempre existe un infinito número de matrices \mathbf{X} invertibles que hacen que la solución no sea única. A esta ambigüedad también se la denomina “de forma”

porque en función de cómo sea \mathbf{X} los perfiles \mathbf{C} y \mathbf{C}^* , al igual que \mathbf{S} y \mathbf{S}^{T*} pueden ser totalmente diferentes.

$$\mathbf{A} = \mathbf{C} \mathbf{S}^T = \mathbf{C} \mathbf{X} \mathbf{X}^{-1} \mathbf{S}^T = (\mathbf{C}\mathbf{X}) (\mathbf{X}^{-1}\mathbf{S}^T) = \mathbf{C}^* \mathbf{S}^{T*} \quad (3.12)$$

La ecuación 3.13 refleja las **ambigüedades de intensidad o de magnitud**. Dado cualquier esquema de factorización, siempre se puede extraer un escalar genérico, α , de la matriz \mathbf{C} y, su inverso $1/\alpha$, de la matriz \mathbf{S}^T . De esta manera, \mathbf{C}^* es α veces \mathbf{C} y \mathbf{S}^{T*} es $1/\alpha$ veces \mathbf{S}^T , aunque \mathbf{C}^* y \mathbf{S}^{T*} tienen la misma forma que \mathbf{C} y \mathbf{S}^T , respectivamente.

$$\mathbf{A} = \mathbf{C} \mathbf{S}^T = (\alpha \mathbf{C}) (1/\alpha \mathbf{S}^T) = (\alpha \times 1/\alpha) (\mathbf{C} \mathbf{S}^T) = \mathbf{C}^* \mathbf{S}^{T*} \quad (3.13)$$

3.7.4.1. Minimización de las ambigüedades. Uso de restricciones.

Las restricciones ayudan a optimizar la convergencia del algoritmo ALS minimizando el efecto de las ambigüedades. Las restricciones pueden ser aplicadas tanto a los perfiles de concentración como a los perfiles espectrales utilizándose el conocimiento previo que se tenga del sistema enzimático. En la bibliografía se distinguen dos tipos de restricciones: Las denominadas naturales (no negatividad, unimodalidad, sistema cerrado, selectividad, igualdad) y las denominadas restricciones de modelo.

Restricciones naturales

Este tipo de restricciones obedecen a propiedades fisico-químicas de los perfiles de concentración y espectrales del sistema (figura 3.15).

Capítulo 3

No negatividad

Esta es la restricción más general en MCR-ALS. Se obliga a que los perfiles de concentración o los perfiles espectrales han de ser siempre positivos [**Winding 1992, Maeder 1990**]. Esta restricción es fácilmente extrapolable a sistemas espectrofotométricos, donde se supone que los perfiles espectrales estarán compuestos siempre de valores positivos de absorbancia.

Unimodalidad

Esta restricción se puede aplicar en aquellos perfiles en los que se presupone que sólo poseen un máximo. Esta situación es bastante corriente para perfiles de concentración y no tan corriente para los perfiles espectrales. Existen varios algoritmos para la aplicación de unimodalidad [**Bro 1998b**], aunque la forma más sencilla e intuitiva de aplicar esta restricción es encontrar el máximo del perfil, los valores en el entorno de este máximo han de disminuir de forma monótona, siempre dependiendo del nivel de ruido esperado por la técnica analítica.

Sistema cerrado

Esta restricción se puede emplear en sistemas donde la cantidad total de materia es constante. Tiene efectos importantes en la minimización de la ambigüedad de intensidad y en el escalado del sistema, ya que se tiende a la normalización de los perfiles de concentración.

Selectividad

Otra restricción importante es la de selectividad. Esta restricción se puede usar en áreas donde se sabe con seguridad que el cambio en la absorbancia es debido únicamente a un componente específico.

Igualdad

Si se posee algún tipo de información, tanto de los perfiles de concentración como espectrales la restricción de igualdad puede ser utilizada actuando con funciones penalizadoras, que hacen que las estimaciones obtenidas en cada uno de los ciclos del algoritmo ALS se adapten en mayor o menor grado a la información suministrada como restricción de igualdad [Bro 1998, Gemperline 2003].

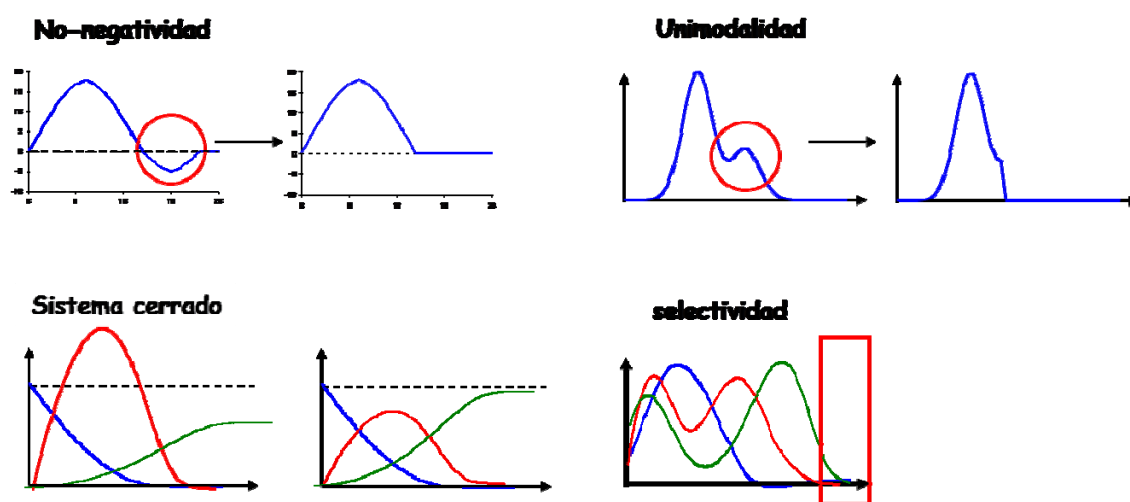


Figura 3.15: Restricciones naturales.

3.7.4.2. Restricciones de modelo. *Hard-Soft* MCR-ALS, HS-MCR-ALS

En el capítulo 2 de esta memoria se hacía mención de la utilización de las metodologías de ajuste duro (*Hard Modelling*) de los datos obtenidos a modelos empíricos de cinética enzimática. Esto es, ajustar los perfiles de concentración (con su correspondiente corrección de la absorbancia) a un modelo enzimático propuesto a través de ajustes no lineales. Esta metodología de trabajo ofrece excelentes resultados si el modelo propuesto es el apropiado y si toda la variación relacionada con la respuesta espectrofotométrica está unida a los componentes que forman el sistema. Los resultados

empeoran cuando surgen variaciones incontroladas durante el proceso o cuando existen interferencias absorbentes no reflejadas en el modelo propuesto.

Un planteamiento novedoso para resolver estos problemas es el usar los modelos enzimáticos (ajuste duro, *Hard*) de manera combinada a las metodologías de resolución de curvas (ajuste blando, *Soft*) para aprovechar de forma simultánea las ventajas de ambos métodos [Amigo 2006, Amigo 2006b, De Juan 2000, Diework 2003, Haario 1998].

De esta manera, los perfiles de concentración obtenidos en cada iteración de ALS son ajustados al modelo enzimático propuesto por medio de algoritmos de ajuste no-lineal (figura 3.16).

La principal ventaja del uso simultáneo de ambos métodos es la minimización de la ambigüedad de rotación inherente a los métodos de resolución, lo cual permite la monitorización del sistema en estudio y la obtención de una estimación de los diferentes parámetros del modelo enzimático propuesto.

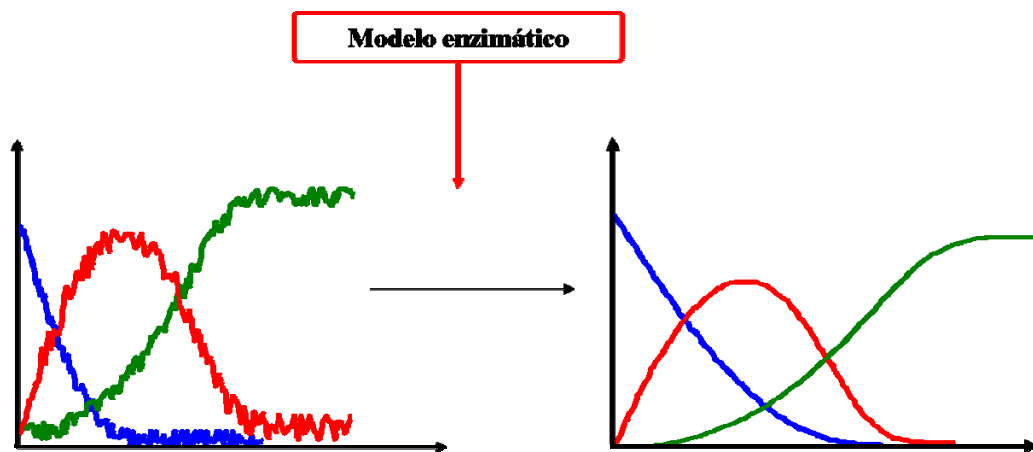


Figura 3.16: Aplicación del modelo enzimático a los perfiles de concentración obtenidos en un análisis MCR-ALS.

Además, también permite el ajuste del proceso enzimático en presencia de interferentes absorbentes que no intervienen en el modelo enzimático, ya que la versatilidad de la metodología de trabajo permite seleccionar las especies a las que se quiere ajustar el modelo enzimático, dejando las demás especies libres para ser modeladas de forma *soft-modelling*.

Su utilización no es trivial, por esta razón se desarrollará con más detenimiento en el capítulo de resultados, donde se mostrará su aplicación a los sistemas enzimáticos estudiados en esta memoria (ver capítulo 5).

El MCR-ALS debe ser considerado una metodología cualitativa debido al gran número de ambigüedades existentes en la descomposición matemática. Sin embargo, con el adecuado uso de las restricciones se puede convertir en una técnica cuantitativa.

El uso de espectros de concentración conocida o de perfiles de concentraciones reales como estimaciones iniciales ayuda a minimizar las ambigüedades [De Juan 2000]. Aún así las ambigüedades siguen existiendo. Por ejemplo, la aparición de una interferencia inesperada en el sistema anularía la capacidad como técnica cuantitativa a pesar de las estimaciones iniciales.

Es este aspecto, la restricción de modelo es la que más puede ayudar a minimizar las ambigüedades y convertirla en una técnica cuantitativa, ya que se está utilizando un ajuste no-lineal a un modelo para los perfiles de concentración que se crea conveniente, dejando fuera de ese modelo las interferencias inesperadas (las cuales son modeladas por *soft-modelling*).

3.7.5. CRITERIO DE CONVERGENCIA Y PARÁMETROS DE CALIDAD DEL AJUSTE EN MCR-ALS

La convergencia del algoritmo ALS se consigue cuando en dos ciclos consecutivos la diferencia de los residuales entre los datos experimentales y los perfiles

Capítulo 3

calculados por ALS es menor que un valor predeterminado. Típicamente este valor es de 0.1%, siendo un valor variable dependiendo de los objetivos del análisis y del tipo de datos que se están tratando.

Los parámetros de calidad del proceso de optimización son el porcentaje de falta de ajuste, el porcentaje de varianza explicada y la desviación estándar de los residuales respecto a los datos experimentales.

El porcentaje de falta de ajuste se define como la diferencia entre los datos de la matriz \mathbf{M} y la matriz reproducida a partir del producto \mathbf{CS}^T obtenido por MCR-ALS. Este valor se calcula a partir de la expresión [Jaumot 2005]:

$$\% \text{ de falta de ajuste} = 100 \times \sqrt{\frac{\sum_{jk} e_{jk}^2}{\sum_{jk} m_{jk}^2}} \quad (3.14)$$

Donde m_{jk} es cada elemento de la matriz \mathbf{M} y e_{jk} es cada elemento de la matriz de residuales \mathbf{E} .

El porcentaje de varianza explicada y la desviación estándar de los residuales son calculados de acuerdo a las ecuaciones 3.15 y 3.16, respectivamente:

$$R^2 = \frac{\sum_{jk} m_{jk}^2 - \sum_{jk} e_{jk}^2}{\sum_{jk} m_{jk}^2} \quad (3.15)$$

$$\sigma = \sqrt{\frac{\sum_{jk} e_{jk}^2}{J_{\text{filas}} \cdot K_{\text{columnas}}}} \quad (3.16)$$

3.7.6. MCR-ALS APLICADO A SISTEMAS TRIDIMENSIONALES. 3W-MCR-ALS

Hasta ahora, hemos hablado de MCR-ALS como una técnica de análisis de sistemas evolutivos aplicada a una sola matriz bidimensional. Esta matriz representa la evolución enzimática de un sistema en unas determinadas condiciones. Pero el MCR-ALS también se puede aplicar a varias matrices bidimensionales simultáneamente. Esto es, un sistema tridimensional, donde se recogen las evoluciones de un sistema enzimático bajo diferentes condiciones iniciales de catálisis puede ser tratado con MCR-ALS de forma simultánea. De esta forma se analiza la matriz tridimensional $\underline{\mathbf{M}}$ ($I \times J \times K$) de manera simultánea, siendo I el número de muestras o procesos utilizados para el análisis. Esta forma de trabajar con MCR-ALS se conoce como 3W-MCR-ALS. Las I matrices individuales \mathbf{M} son encadenadas dependiendo de las características del sistema enzimático y de los objetivos que se persigan [Amrhein 1996, Saurina 1998, Tauler 1995, Tauler 1995b, Tauler 2001].

Al desdoblar la matriz $\underline{\mathbf{M}}$ se pierde la trilinealidad del sistema, analizándose de forma bilineal. De esta manera, pueden surgir dos tipos de encadenamiento, principalmente: encadenamiento por columnas y encadenamiento por filas.

3.7.6.1. Encadenamiento por columnas

El encadenamiento por columnas se aplica cuando el sistema de registro es el mismo para todas las cinéticas, es decir, los perfiles espectrales esperados son los mismos para todas las I matrices y lo que varía de matriz a matriz son los perfiles de concentración (figura 3.17). Éste es nuestro caso, en el que se monitoriza la misma reacción enzimática con la misma técnica instrumental, pero variando la concentración inicial de sustrato. En estas condiciones, los perfiles espectrales serán siempre los debidos a los mismos analitos, pero los perfiles de concentración dependerán de la concentración inicial de sustrato. La ecuación 3.17 muestra el modelo para I experimentos, dando como

resultado i matrices C_{jr} de perfiles de concentración y una matriz común S_{rk} de perfiles espectrales.

$$\begin{bmatrix} M_{11} \\ \vdots \\ M_{jk} \end{bmatrix} = \begin{bmatrix} C_{11} \\ \vdots \\ C_{jr} \end{bmatrix} S^T + \begin{bmatrix} E_{11} \\ \vdots \\ E_{jk} \end{bmatrix} \quad (3.17)$$

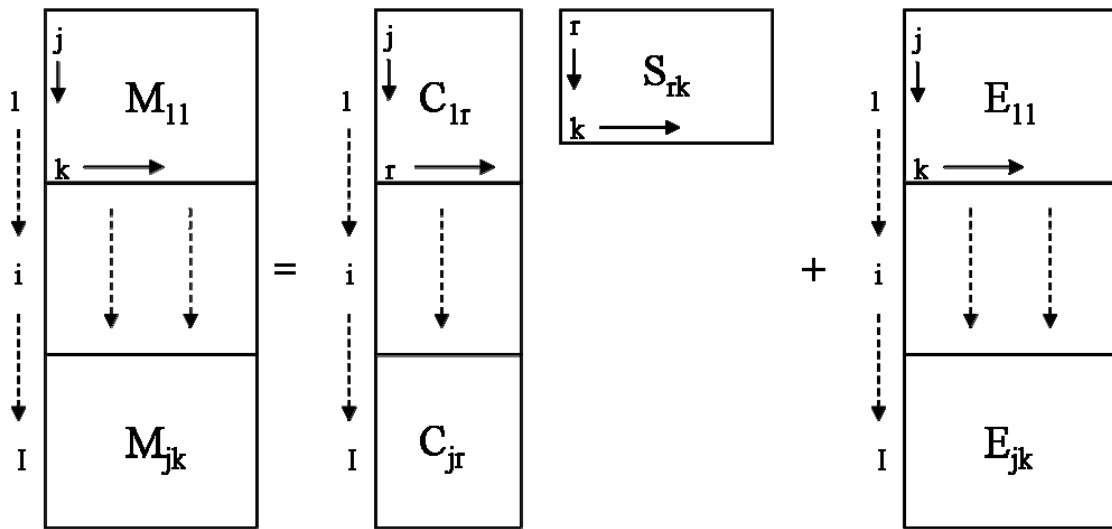


Figura 3.17: Encadenamiento 3W-MCR-ALS por columna.

3.7.6.2. Encadenamiento por filas

Cuando un mismo sistema enzimático es monitorizado en las mismas condiciones iniciales pero usando diferentes técnicas analíticas de análisis (por ejemplo, espectrofotometría UV-Vis y fluorescencia) las diferentes i matrices M se deben encadenar por filas. Ahora sólo se obtendrá una matriz de perfiles de concentración y una matriz de perfiles espectrales para cada una de las i matrices (figura 3.18)

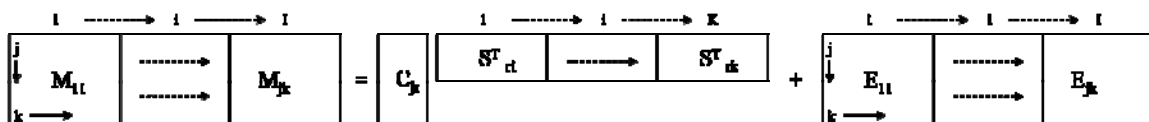


Figura 3.18: Encadenamiento 3W-MCR-ALS por fila

En esta ocasión la ecuación 3.18 muestra el modelo para i experimentos, dando como resultado una matriz común C_{jr} de perfiles de concentración y una matriz S_{rk} de perfiles espectrales para cada una de las i matrices.

$$\begin{bmatrix} M_{11} & \dots & M_{jk} \end{bmatrix} = C_{jr} \begin{bmatrix} S_{r1}^T & \dots & S_{rk}^T \end{bmatrix} + \begin{bmatrix} E_{11} & \dots & E_{jk} \end{bmatrix} \quad (3.18)$$

Finalmente, existe otro tipo de encadenamiento de matrices en sistemas en los que varían tanto los perfiles de concentración como los perfiles espectrales, dando lugar a un encadenamiento conjunto por filas y por columnas.

3.7.7. METODOLOGÍA DE TRABAJO EN 3W-MCR-ALS

La metodología de trabajo para la resolución de sistemas encadenados puede ser extrapolada de la metodología de trabajo de MCR-ALS para matrices individuales. Los pasos a realizar y los parámetros de ajuste son exactamente los mismos. La gran ventaja de trabajar con matrices encadenadas es la resolución de sistemas deficientes de rango. Encadenando una matriz deficiente de rango con otra matriz de rango completo, 3W-MCR-ALS utiliza la información que obtiene de la matriz de rango completo para eliminar la deficiencia de rango en la otra matriz.

Otra gran ventaja de trabajar con sistemas encadenados se observa cuando se implementa la restricción de modelo. La gran versatilidad de esta restricción hace que podamos aplicarla a componentes individuales de una de las matrices de manera específica, aprovechando la información obtenida para las demás matrices de forma simultánea.

3.8. UTILIZACIÓN DE PARAFAC *ON-LINE* PARA EL CONTROL A TIEMPO REAL DE PROCESOS

3.8.1. INTRODUCCIÓN

En el capítulo anterior se ha destacado la importancia que está adquiriendo los espectrofotómetros de fluorescencia multicanal para la monitorización de bioprocesos. [Lindemann, 1998; Hisinger, 2005]. Aplicados a cualquier proceso, estos instrumentos producen una señal tridimensional (tiempo x longitud de onda de excitación x longitud de onda de emisión). Esta señal compleja requiere de un método como PARAFAC para extraer la información estrictamente relevante del proceso. Esta información es la que proporcionan las entidades químicas (fluoróforos) que hacen variar la señal de forma significativamente diferente al ruido durante el bioproceso [Bro 1996, Bro 1997, Bro 1997b, Anderssen 2003].

La Tecnología Analítica de Proceso, PAT (*Process Analytical Technology*), se está convirtiendo en una herramienta importante para entender y controlar los bioprocesos. El uso de la Estadística Multivariable del Control de Procesos, MSPC (*Multivariate Statistical Process Control*) ofrece un efectivo y eficiente medio de adquirir información para facilitar el conocimiento del bioproceso, genera estrategias de minimización de riesgos y facilita mejoras continuas en el mismo [Wise 1996, Louwerse 2000]. Por lo que la combinación de las dos metodologías (modelado PARAFAC de los datos de fluorescencia multicanal más MSPC) ofrecerá mejoras substanciales en la tecnología de bioprocesos.

De esta manera, el objetivo de esta unión es mostrar como el uso combinado de PARAFAC y la estadística multivariable de control de procesos, MSPC aplicada a señales espectrofluorimétricas multidimensionales ofrece una forma fácil y visualmente directa de monitorización y control a tiempo real de bioprocesos. PARAFAC permitirá obtener toda la información necesaria acerca del proceso a través de la creación de un modelo de calibración con procesos monitorizados en condiciones normales de trabajo,

NOC (*normal operating conditions*). El modelo se compondrá de las tres matrices de *loadings*, dos de ellas relacionadas con los espectros de fluorescencia de excitación y emisión de los componentes del sistema, mientras que la tercera matriz de *loadings* contendrá la información necesaria de la evolución de la concentración relativa de los diferentes componentes a lo largo de la reacción. Los *loadings* de concentración servirán para predecir los perfiles de concentración en nuevos procesos.

3.8.2. CONSTRUCCIÓN DEL MODELO PARAFAC DE CALIBRACIÓN

El modelo de calibración se debe construir a partir de uno o varios procesos completos los cuales se han monitorizado en condiciones normales de operación, NOC (*normal operating conditions*). Esto significa que estos procesos deben recoger el comportamiento real del sistema bioquímico. Los procesos se monitorizaron a lo largo del tiempo y se registró la evolución de sus espectros de excitación y emisión (λ_{em} , λ_{ex}) para los n tiempos. De esta manera se creó la matriz tridimensional $\underline{\mathbf{M}}$ ($n \times \lambda_{ex} \times \lambda_{em}$) (figura 3.19a).

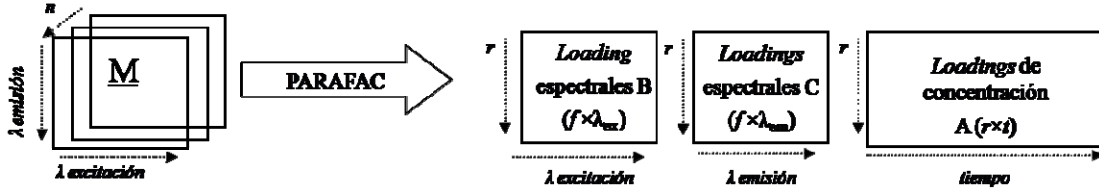
Tal y como se expone en la sección 3.6, PARAFAC es capaz de descomponer esta matriz en tres sub-matrices bidimensionales, las cuales son optimizadas de tal manera que expliquen la máxima varianza de los datos (figura 3.19a). Las primeras dos matrices, denominadas matrices de *loadings* espectrales, están relacionadas con los perfiles espectrales de excitación (\mathbf{B} , $r \times \lambda_{ex}$) y emisión (\mathbf{C} , $r \times \lambda_{em}$). La tercera matriz, denominada *loadings* de concentración \mathbf{A} ($r \times n$), está relacionada con la variación de la señal de los r componentes durante el bioproceso. Matemáticamente, cada elemento m_{ijk} de $\underline{\mathbf{M}}$ ($n \times \lambda_{ex} \times \lambda_{em}$) se puede expresar tal y como se expone en la ecuación 3.8.

Otra manera de expresar esta descomposición es la siguiente [Louwerse 2000, Smilde 2004]:

$$M = A(C \circ B)^T + E \quad (3.19)$$

Donde \circ representa el producto matricial de Khatri-Rao [Rao 1971].

a) Creación del modelo PARAFAC con el primer bioproceso



b) Monitorización y control a tiempo real de un nuevo bioproceso

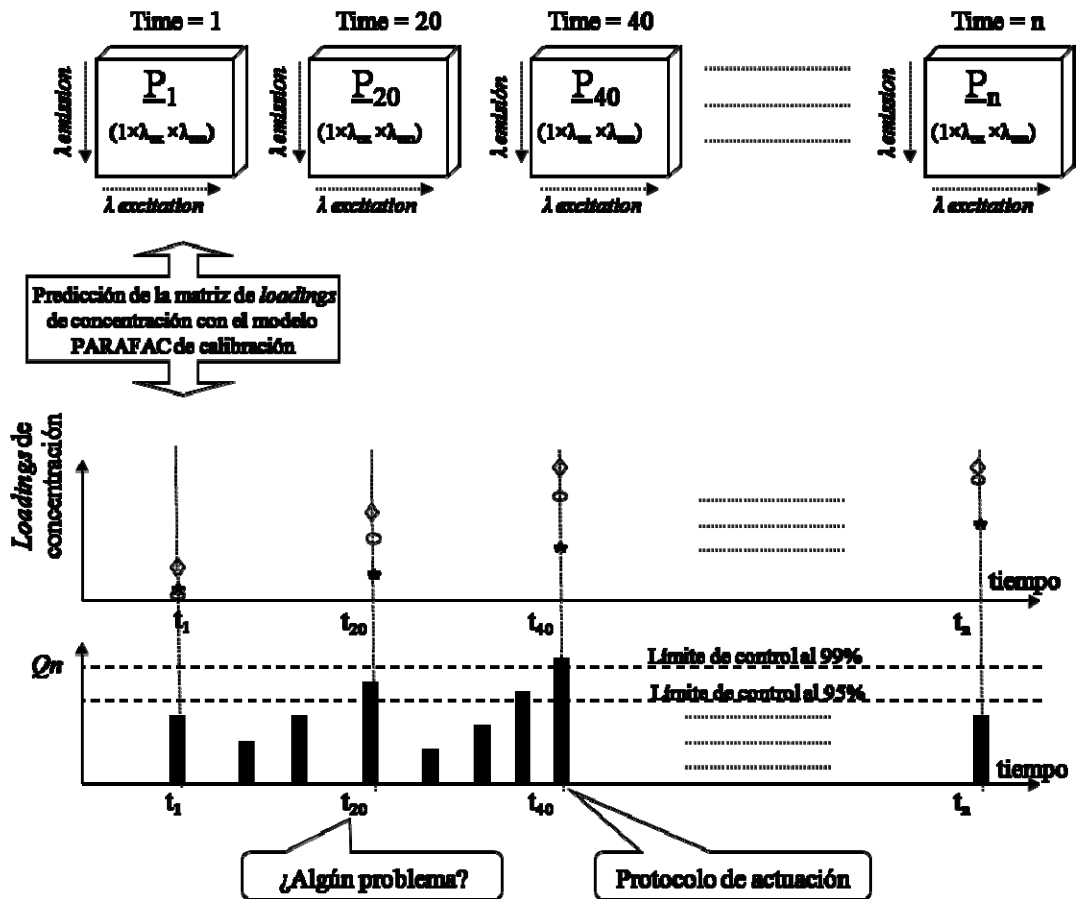


Figura 3.19: a) Creación del modelo PARAFAC de calibración con un bioproceso monitorizado en condiciones normales de operación (NOC). B) Control a tiempo real de un nuevo bioproceso.

3.8.3. MONITORIZACIÓN DEL BIOPROCESO EN TIEMPO REAL. PROYECCIÓN DE UNA MUESTRA EN EL MODELO PARAFAC DE CALIBRACIÓN.

Una vez creado el modelo PARAFAC, el siguiente paso es comenzar los bioprocesos que se quieren monitorizar y controlar a tiempo real. Para un nuevo bioproceso, $\underline{\mathbf{P}}$, los espectros se deben registrar usando el mismo intervalo de longitudes de onda que el utilizado para la creación del modelo. Obviamente, se espera que la variación de la señal sea causada por la evolución de los mismos fluoróforos, y, consecuentemente, los perfiles espectrales que se van a obtener en las matrices de *loadings* espectrales \mathbf{B} y \mathbf{C} permanecerán constantes.

Para un tiempo específico i del nuevo bioproceso, se obtiene una matriz bidimensional P_i , de dimensiones $\lambda_{ex} \times \lambda_{em}$. Sin embargo, esta matriz se puede visualizar como una matriz tridimensional, donde el tiempo tiene “dimensión 1”. Consecuentemente, las dimensiones de la nueva matriz \underline{P}_i serán $1 \times \lambda_{ex} \times \lambda_{em}$ (Fig. 3.19b).

La nueva matriz de *loadings* de concentración \mathbf{A}_i ($r \times 1$) para \underline{P}_i se puede predecir usando los *loadings* espectrales obtenidos en el modelo PARAFAC de calibración. Louwerse definió esta nueva proyección matemáticamente de la siguiente manera [Louwerse 2000]:

$$a_i = [(C \circ B)^T (C \circ B)]^{-1} (C \circ B)^T p_i \quad (3.20)$$

Donde p_i es el valor experimental correspondiente a cada elemento de la matriz \underline{P}_i . Los valores de a_i obtenidos ofrecen la información del comportamiento de cada componente a un tiempo determinado i . Los valores residuales de cada muestra e_i se calculan de la siguiente manera:

$$e_i = p_i - (C \circ B)a_i \quad (3.21)$$

Este cálculo se repite para cada tiempo, por lo que a efectos prácticos, se está obteniendo una monitorización a tiempo real del comportamiento y la evolución de los fluoróforos a lo largo del bioproceso.

3.8.4. ESTADÍSTICO RESIDUAL, Q_N , COMO PARÁMETRO DE CONTROL

Los dos parámetros más utilizados en la MSPC para el control de procesos son el estadístico T^2 de Hotelling (también llamado D-chart) y el estadístico residual Q_n .

El estadístico residual Q_n indica la varianza residual no capturada por los r componentes de las muestras del nuevo proceso respecto al modelo de “calibración” [Wise 1996]. Es el parámetro indicado cuando el control se realiza muestra por muestra. Este parámetro se define como la suma de los residuales cuadráticos obtenidos a tiempo i (ecuación 3.22, figura 3.20):

$$Q_n = \sum_{j=1}^J \sum_{k=1}^K e_i^2 \quad (3.22)$$

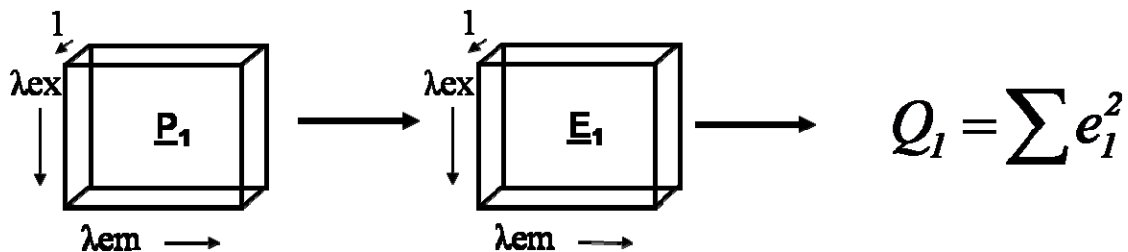


Figura 3.20: Estadístico residual para una muestra \underline{P} monitorizada a un tiempo determinado.

Donde e_i es el valor del residual definido en la ecuación 3.22. Este valor viene controlado por dos niveles de confianza [Louwerse 2000], los cuales son impuestos a partir de la varianza total explicada del modelo PARAFAC de calibración [Bro 1997b]. Asumiendo una distribución normal de los valores de los residuales del modelo de

calibración, los dos límites de confianza se calculan como el 95 y el 99% del intervalo de confianza (figura 3.19b). Estos valores son similares a los valores de la media de los valores residuales más 2 y 3 veces la desviación estándar, respectivamente. Si el valor de Q_n en un determinado tiempo i se encuentra por debajo del primer límite de confianza, se puede asumir que el proceso se encuentra en condiciones controladas. Si este valor se encuentra entre los dos límites, puede que se esté dando alguna anomalía en el proceso respecto a NOC. Y si el valor se encuentra por encima del límite al 99%, el proceso se encuentra claramente fuera de control. Una de las ventajas de esta nueva metodología es que las muestras se tratan de forma individual, por lo que el tiempo de registro puede ser variado de acuerdo a los requerimientos del sistema y la evolución de los valores de Q_n .

Esta nueva metodología de trabajo se puede aplicar a lo largo de todo el proceso de fermentación. Todos los pasos a seguir para aplicar esta metodología se encuentran en el diagrama de actuación de la figura 3.21. La ventaja es que este diagrama de actuación se puede adaptar a los requerimientos específicos de cada bioproceso de acuerdo al propósito final del mismo.

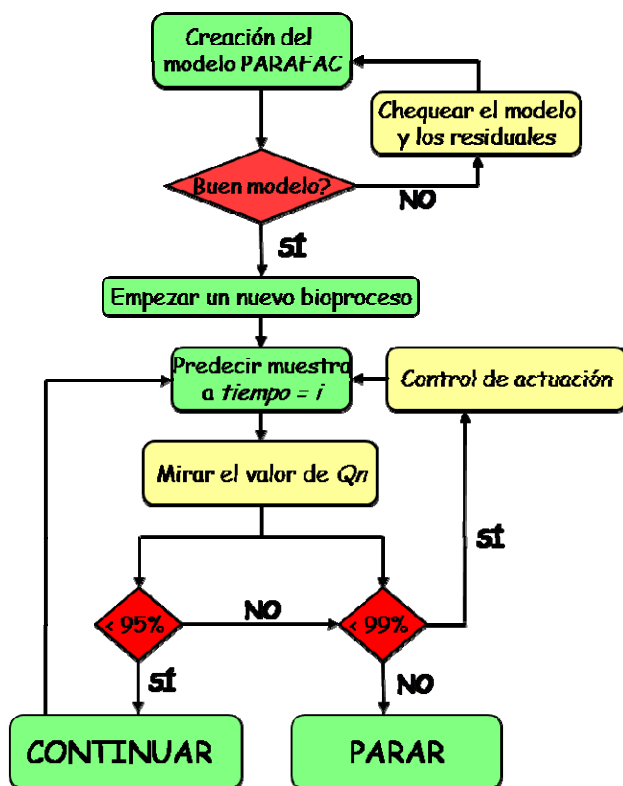


Figura 3.21: Diagrama de control de un nuevo bioproceso.

3.9. PROGRAMAS DISPONIBLES EN EL ANÁLISIS MULTIVARIABLE MULTIDIMENSIONAL DE DATOS EVOLUTIVOS

A continuación se hará un breve resumen de los algoritmos utilizados en esta memoria y su procedencia, si bien se tiene que tener en cuenta que no son los únicos existentes para realizar esta tarea, sí son los más conocidos.

La gran aplicabilidad de los métodos quimiométricos a sistemas dinámicos hace que cada día se generen nuevos algoritmos más rápidos y más fáciles de implementar a los sistemas en estudio. Los algoritmos tienden cada vez a ser más rápidos y a estar implementados dentro de interfaces de usuario fácilmente entendibles y muy intuitivas.

Se tiende a generar plataformas “*plug and play*” que hagan que apretando un botón se aplique el tratamiento escogido sin necesidad de escribir código. Este hecho facilita mucho al usuario, muchas veces no experto, el utilizar la quimiometría de una manera sencilla y didáctica.

La mayoría de estos algoritmos se encuentran escritos en código MatLab® [Mat]. Por lo que se hace necesario cierto dominio de este programa. Otros son programas independientes. En la tabla 3.2 se puede ver una comparativa entre los diferentes algoritmos usados en esta memoria respecto a su aplicabilidad. En la tabla 3.3 se hace un resumen de los principales programas disponibles y sus características.

Tabla 3.2: Comparación de los diferentes algoritmos en función de su aplicabilidad

ALGORITMO	APLICACIÓN	CAPACIDAD DE MODELAR SISTEMAS TRIDIMENSIONALES NO TRILINEALES	NECESIDAD DE ESTIMACIONES INICIALES	CÁLCULO DE FACTORES-COMPONENTES
Unfold PCA	CUALITATIVA	Es capaz de modelar sistemas que no se desvían demasiado de la trilinealidad.	NO	Cálculo de factores. Los <i>scores</i> y los <i>loadings</i> obtenidos son de interpretabilidad limitada
Unfold PLS	CUANTITATIVA. Pero necesita de una gran cantidad de muestras para realizar un modelo de calibración robusto			
3W-PCA	CUALITATIVA			Cálculo de componentes al basarse en una descomposición PARAFAC. Los <i>loadings</i> obtenidos son de interpretabilidad limitada
3W-PLS	CUANTITATIVA. Pero necesita de una gran cantidad de muestras para realizar un modelo de calibración robusto			
MCR-ALS	CUALITATIVA. Aunque con buenas estimaciones iniciales se puede usar como técnica cuantitativa	Es una técnica bilinear para matrices bidimensionales	SÍ. Deben ser buenas estimaciones iniciales	Cálculo de componentes. Los perfiles obtenidos corresponden a perfiles de concentración y espectros de los analitos absorbentes en nuestro sistema
3W-MCR-ALS		SÍ		
IIS-MCR-ALS	CUANTITATIVA. La restricción de modelo es lo suficientemente fuerte como para poder cuantificar. No necesita una gran cantidad de muestras.	SÍ		
PARAFAC	CUALITATIVA. Es una técnica análoga al PCA bilinear	NO. Cualquier desviación de la trilinealidad hará que el modelo no sea representativo del sistema en estudio	No externas. Calculadas a partir de M	

3.9.1. UNFOLD PCA Y PLS

Al ser el PCA y el PLS las dos metodologías más conocidas, existen numerosos programas de creación propia. Sin embargo, hay que destacar dos como los más representativos y útiles.

Unscrambler [UNS]

Sin duda, el programa Unscrambler, desarrollado por Camo, es uno de los más sencillos de utilizar y uno de los que ofrece una mayor facilidad a la hora de trabajar con los datos experimentales aplicando PCA y PLS.

Capítulo 3

Su uso es muy sencillo y no se requieren conocimientos de programación para utilizarlo. Tiene la gran ventaja de que es muy fácil la elección de parámetros de los modelos (número de factores, por ejemplo), la eliminación de muestras anómalas, etc. Uno de los aspectos más destacables es el entorno gráfico. Hacer gráficos de *scores* y *loadings*, residuales, varianza por muestra, etc. es muy sencillo y rápido.

Los requerimientos de memoria dependen mucho de la cantidad de información que se quiere utilizar y del ordenador personal del que se dispone. Aún así, Unscrambler necesita una cantidad de memoria RAM media-alta para su correcto funcionamiento. Este problema ha quedado resuelto con la generación de ordenadores personales cada vez más potentes (hasta 4 Gbytes de memoria RAM)

La gran desventaja a la hora de aplicar un análisis *Unfold* es que las matrices se han de introducir ya desdobladas. Por lo que se hace necesario el uso de algún otro programa de tratamiento de datos. Por ejemplo, es muy común el uso combinado de MatLab y Unscrambler debido a la facilidad de MatLab para el tratamiento de matrices. Unscrambler puede exportar e importar archivos con formato MatLab. No así MatLab, que no puede exportar archivos Unscrambler.

Hoy en día, el problema de importar – exportar archivos desde un espectrofotómetro hasta un programa de tratamiento de datos no resulta un problema. Unscrambler puede importar archivos desde una gran cantidad de formatos.

Tabla 3.3: Comparación de algoritmos para el tratamiento de datos multivariable multidimensionales evolutivos

ALGORITMOS	SOFTWARE	NECESIDAD DE MATLAB	ALGORITMO O PLATAFORMA DE LIBRE ACCESO	NIVEL DE MEMORIA RAM REQUERIDO	NIVEL DE DIFICULTAD DE MANEJO	USADO EN ESTA MEMORIA
Unfold PCA Unfold PLS	Unscrambler	NO	NO	medio	bajo	sí
	PLS-Toolbox	sí	NO	medio	medio	sí
3W-PCA 3W-PLS	Unscrambler	NO	NO	medio	bajo	NO
	PLS-Toolbox	sí	NO	alto	medio	sí
SVD, EFA	MatLab	sí	EFA disponible en [web 3.1]	bajo	bajo	sí
MCR-ALS Y 3W-MCR-ALS	als2004	sí	Disponible en [web 3.1]	medio	medio	sí
	GUIPRO	sí	Disponible en [web 3.2] No disponible para 3W-MCR-ALS	medio	medio	NO
HS-MCR-ALS	MatLab	sí	Disponible en [Amigo 2006]	alto	alto	sí
PARAFAC	MatLab	sí	Disponible en [web 3.3]	alto	alto	sí
	PLS-Toolbox	sí	NO	alto	medio	NO

PLS-Toolbox [PLSTOOL]

La PLS-Toolbox es una herramienta que recoge diversos algoritmos para tratamientos quimiométricos. Fue creada por Eigenvector® y trabaja bajo formato MatLab. Esto es un handicap a la hora de su manejo, ya que se hace necesario un mínimo conocimiento del tratamiento de matrices con MatLab.

Cuenta también con una interfaz de usuario gráfica fácil de manejar, pero que resulta poco intuitiva al principio. En lo que a PCA y PLS se refiere, es una herramienta menos potente que Unscrambler ya que cuenta con muchas menos opciones de graficado y resulta menos didáctica. Además, MatLab, software bajo el cual trabaja la PLS-Toolbox, requiere un nivel alto de memoria RAM.

Pero la gran ventaja es que al trabajar desde MatLab, no hace falta un programa de tratamiento de datos para el desdoblamiento. El problema se nos puede presentar a la hora de importar datos desde un espectrofotómetro. MatLab acepta muchos formatos de archivos, pero, a diferencia de Unscrambler, no está pensado exclusivamente para el tratamiento de datos recogidos por equipos espectrofotométricos.

3.9.2. 3W-PCA Y 3W-PLS

Unscrambler

Hasta hace poco esta opción de análisis no era posible en Unscrambler. Pero en las últimas versiones del programa (Unscrambler 9.0) ya se han incluido algoritmos de análisis de 3W-PCA y 3W-PLS. De esta manera, se tiene la misma capacidad de graficado y facilidad a la hora de obtener resultados que con PCA y PLS bilinear.

El problema viene a la hora de introducir los datos. No es tan evidente como cabría de esperar debido a la dificultad de crear una matriz tridimensional en una hoja de cálculo bidimensional.

PLS-Toolbox

La PLS-Toolbox sigue ofreciendo las mismas posibilidades que con PCA y PLS bilinear, pero la ventaja es que, una vez que se domina el uso de MatLab, el manejo de matrices tridimensionales es mucho más sencillo que con Unscrambler.

3.9.3. SVD, EFA

Los códigos usados para la realización de SVD y EFA están escritos en código MatLab. La utilización del SVD requiere conocimientos de MatLab, y su utilización no es intuitiva.

Hay que destacar la sub-rutina EFA generada por la Dra. Anna de Juan [**Web 3.1**]. Es de libre acceso vía Web y también funciona con MatLab. Su uso es más didáctico y aporta mucha más información que la rutina propia de MatLab, por lo que su uso facilita mucho la comprensión de los resultados del análisis

3.9.4. MCR-ALS Y 3W-MCR-ALS

Unscrambler

Al igual que 3W-PCA y 3W-PLS, el MCR-ALS ha sido incluido en las últimas versiones de Unscrambler (versión 9.2). Esta memoria está basada en MatLab, por lo que se desconoce la potenciabilidad de este programa en la aplicación de MCR-ALS.

als2004 [Web 3.1]

El als2004 es uno de los programas más potentes que existen de MCR-ALS. Fue generado por R. Tauler en la Universidad de Barcelona. Al principio era un programa que trabajaba a través de código desde MatLab, por lo que su utilización resultaba difícil. En 2005 J. Jaumot [**Jaumot 2005**] desarrolló una plataforma “*plug and play*” que facilitaba su utilización. Esta plataforma es muy didáctica y muy fácil de usar por una persona no experta en el campo MatLab. La elección de las restricciones y el uso de información adicional se realizan de manera muy sencilla.

Una gran ventaja del als2004 es que permite el trabajo con matrices encadenadas en ambas direcciones. Que conozcamos, es el único programa que permite esta opción.

Las principales desventajas de este programa es que no cuenta con opciones de pre-tratamiento de datos, selección de longitudes de onda ni tiempo, ni con opciones de análisis del número de componentes.

GUIPRO [Gemperline 2003]

El GUIPRO es un programa desarrollado por P.J. Gemperline de libre acceso vía Web [Web 3.2]. Trabaja desde MatLab, pero cuenta con una plataforma “*plug and play*” muy sencilla de usar. La principal diferencia con als2004 es que el GUIPRO trabaja con funciones de penalización.

Una gran ventaja del GUIPRO es que la plataforma “*plug and play*” te permite la selección de intervalos de tiempos y longitudes de onda, el pretratamiento de datos y la selección del número de componentes a través de análisis SVD y EFA. Además, la implementación de restricciones se realiza componente a componente, a diferencia del als2004.

Otra ventaja es que permite el ajuste no-lineal de los datos cinéticos obtenidos a modelos empíricos. La capacidad de esta tarea es limitada, pero es una herramienta que puede ser utilizada como paso previo a mejores ajustes con otros programas.

La gran desventaja de este programa es que no permite el trabajo con matrices encadenadas.

3.9.5. HS-MCR-ALS

Las primeras subrutinas de HS-MCR-ALS fueron generadas por A. de Juan y M. Maeder [De Juan 2000]. Estas sub-rutinas trabajan desde MatLab y no se encuentran en plataformas “*plug and play*”, por lo que su uso es complicado. Son de libre acceso y específicas a casos cinéticos concretos.

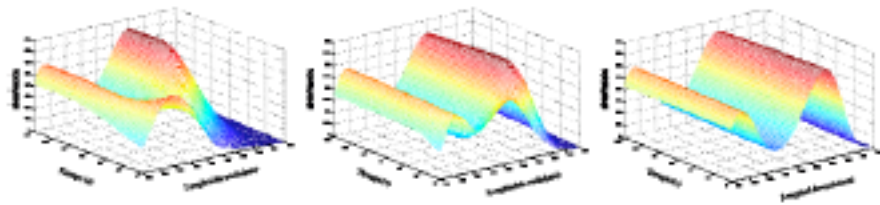
En esta memoria se ha desarrollado una nueva aplicación HS-MCR-ALS para el modelo enzimático de la catálisis de la hipoxantina, xantina y ácido úrico a través de la xantin-oxidasa. Esta aplicación se encuentra disponible vía los autores [Amigo 2006 y

2006b]. El completo funcionamiento de esta sub-rutina se explicará en el capítulo siguiente.

3.9.6. PARAFAC

Las rutinas PARAFAC para MatLab fueron desarrolladas por R. Bro [**Bro 1998**]. En la página Web [**Web 3.3**] se pueden encontrar varias de ellas. También se encuentra en la PLS-Toolbox y se puede utilizar desde la plataforma “*plug and play*”.

El principal problema de PARAFAC es que el algoritmo es lento y necesita una gran capacidad de memoria RAM. Además, al funcionar bajo MatLab, se necesitan conocimientos del manejo de matrices con MatLab.



Capítulo 4

METODOLOGÍA

PARTE I:

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

4.1. REACTIVOS

En los trabajos relacionados con el uso de los enzimas como elementos diferenciadores en una determinación simultánea de varios analitos en muestras biológicas, se utilizaron los siguientes reactivos y disoluciones:

Se prepararon disoluciones patrón de hipoxantina, xantina y ácido úrico de 5 mmol L⁻¹ de concentración (Merck p.a.). Se preparó una disolución patrón de 0.14 Unidades mL⁻¹ de enzima xantina oxidasa (obtenida de la mantequilla; 15% de contenido proteico; 0.4-1.0 Unidades / mg de proteína; Merck).

La orina liofilizada, Lyphocheck® (*Quantitative Urine Control from BIO-RAD*) se regeneró usando la disolución tampón tal y como aconseja el fabricante.

Todas las disoluciones se prepararon usando como disolvente un tampón de TRIS (*Tris(hydroxymethyl)-aminomethan*; Merck p.a.), ajustando el pH mediante una disolución de ácido clorhídrico 1 M (Panreac, p.a.). Este tampón se preparó usando agua

calidad Milli-Q. La concentración final del tampón fue de 50 mmol L⁻¹ y el pH final fue de 8.5.

4.2. CONDICIONES ÓPTIMAS DE CATÁLISIS ENZIMÁTICA

Tal y como se ha mencionado en el primer capítulo la actividad enzimática depende de las condiciones experimentales tales como el pH y la concentración del tampón, la temperatura y la concentración de enzima. Por lo que las condiciones óptimas de catálisis deben ser estudiadas y establecidas.

La temperatura y la concentración de tampón fue fijada en relación a la bibliografía [Escribano 1988]. El pH y la concentración de enzima se optimizaron para la hipoxantina y la xantina, estableciéndose como óptimos un pH de 8.5 y una concentración de enzima de 0.08 U / mL⁻¹.

4.3. APARATOS Y SOFTWARE

Los perfiles cinéticos fueron registrados usando un espectrofotómetro Ultravioleta-Visible equipado con un detector de diodos (HP-8453) (figura 4.2). A este espectrofotómetro se acopló un sistema de flujo interrumpido *stopped-flow* (Applied Photophysics RX 2000). Se usó una celda de cuarzo de 1.00 cm de camino óptico termostatzada a 25.0 ± 0.1 °C. Para mantener la temperatura constante se usó un módulo “Peltier” (Agilent 89090-A). La temperatura ambiente se mantuvo a 24 ± 1°C.

4.4. PROCEDIMIENTO EXPERIMENTAL PARA LA MONITORIZACIÓN DE LAS CINÉTICAS.

La orina liofilizada se regeneró con la disolución tampón y se filtró con filtros de nylon (0.45 µm). Esta solución se diluyó en una proporción 1:25 y fue dopada apropiadamente. Todas las disoluciones se prepararon justo antes de registrar sus cinéticas.

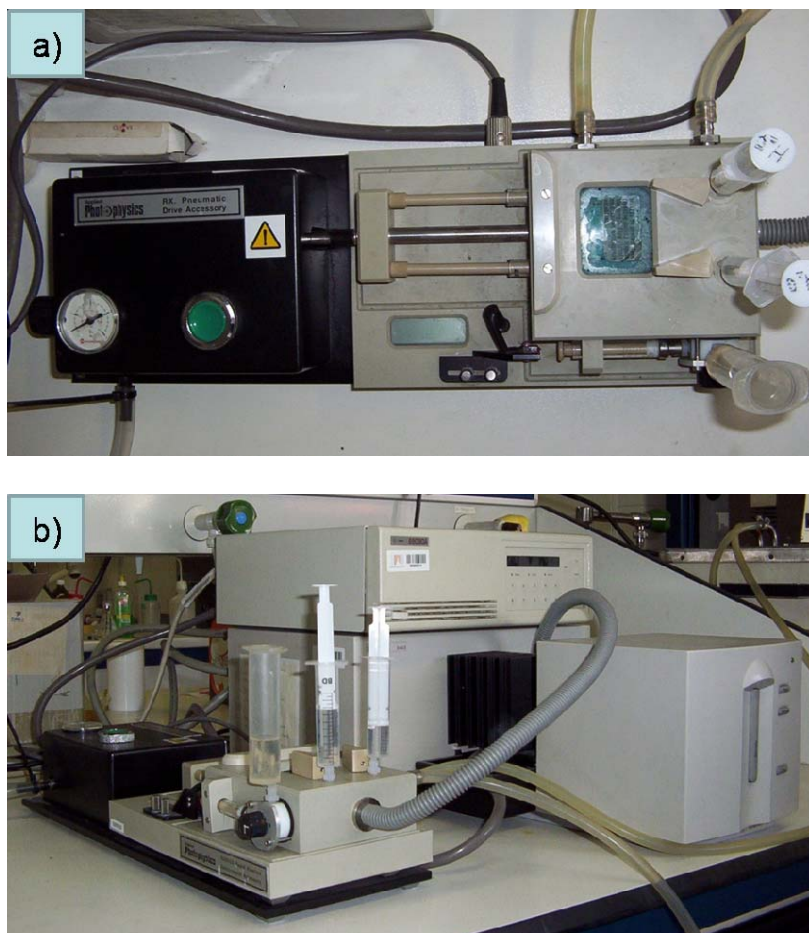


Figura 4.1:a) Sistema de flujo interrumpido (*stopped-flow*). b) Espectrofotómetro UV-Vis-DAD acoplado al módulo de flujo interrumpido y al módulo *Peltier*.

El sistema de flujo interrumpido consta de dos jeringuillas paralelas (figura 4.1). En la primera se introdujo la disolución tamponada de enzima xantina oxidasa. En la segunda jeringuilla se introdujo las diferentes disoluciones tamponadas de calibración, validación y orina. La reacción comenzó cuando una parte de cada una de las disoluciones incluidas en las jeringuillas (150 μl de cada una de ellas) se empuja hacia la celda de reacción. La concentración final de enzima en la celda de reacción fue 0.07 Unidades/ml. Se tomó como referencia espectrofotométrica la disolución de enzima mezclada con tampón.

El registro de los espectros UV-Vis. comenzó 0.3 segundos después de producirse la mezcla y se registraron espectros cada 1 segundo durante 160 segundos en un rango espectral de 225-320 nm en intervalos de 1 nm (figura 4.2). En este tiempo, se completaba todas las reacciones de la mezcla de mayor concentración, excepto la degradación del ácido úrico.

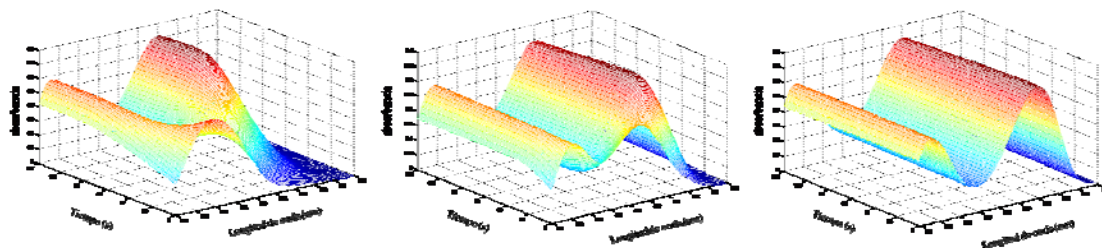


Figura 4.2: Evoluciones cinéticas obtenidas para tres de las muestras de calibración a) C2, cinética de hipoxantina; b) C3, cinética de xantina; c) C4, cinética del ácido úrico.

4.5. TRATAMIENTO DE LOS DATOS UV-VISIBLE

Una vez registrada la evolución de los espectros con el tiempo, los datos obtenidos se exportaron, en formato ASCII, desde el espectrofotómetro a diferentes programas. Se utilizaron los algoritmos 3W-PCA, 3W-PLS1 y 3W-PLS2 de la PLS-TOOLBOX 3.5 [PLSTOOL]. Estas subrutinas de análisis trabajan bajo MatLab® v. 7.0 [MAT]. El programa de HS-MCR-ALS también se generó en código MatLab [MAT] y se implementó el sistema enzimático propuesto. Se utilizó el programa Unscrambler v. 8.5 [UNS] para la realización de los modelos *unfold*-PLS1 y *unfold*-PLS2. El ajuste por mínimos cuadrados no lineal (*Hard-modelling*) se realizó con el programa Berkeley Madonna v. 8.0.1 [Web 4.1].

PARTE II:

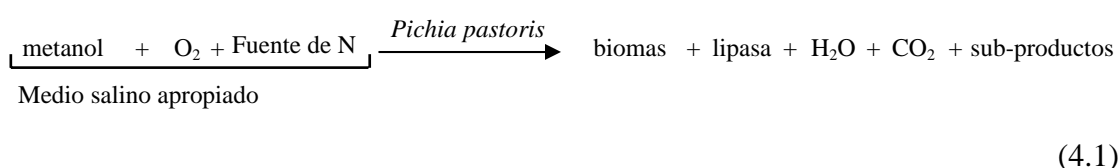
CONTROL EN TIEMPO REAL DE PROCESOS DE PRODUCCIÓN DE ENZIMAS. PRODUCCIÓN DE LIPASAS.

4.6. REACTIVOS

En la monitorización y control a tiempo real del proceso industrial de generación de lipasas, se utilizó el microorganismo *Pichia pastoris* tipo X-33, el cual contenía el vector pPIC α A_ROL para la expresión heteróloga de la *Rhizopus oryzae* lipasa (ROL). El control del microorganismo se hizo bajo el promotor *AOX1* [Minning 2001].

4.7. CONDICIONES DEL BIOPROCESO

Se monitorizaron bioprocesos utilizando dos tipos de sustratos como fuente de carbono, metanol en una concentración inicial de 10 gL^{-1} y glicerol en una concentración inicial de 10 gL^{-1} . El sustrato se metabolizó principalmente a biomasa y lipasa según el esquema (para el metanol):



Los diferentes bioprocesos se llevaron a cabo utilizando un medio sintético basal que contenía 0.02 M de H_3PO_4 , $6.8 \times 10^{-3} \text{ M}$ de CaSO_4 , 0.1 M de K_2SO_4 , 0.06 M de $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 0.073 M de KOH en un volumen final de 1 litro. Se usaron 0.1 mL de antiespumante Mazu DF 7960 (Mazer Chemicals, PPG Industries Inc., Guerne, Ill., USA) y 4.35 mL de una solución de PTM1. Todas las disoluciones se prepararon con agua destilada.

PTM1 es una mezcla de diferentes sales que contenía por litro de agua destilada: 6.0 g de $\text{CuSO}_4 \cdot 5 \text{ H}_2\text{O}$, 0.08 g de NaI , 3.0 g de $\text{MnSO}_4 \cdot \text{H}_2\text{O}$, 0.2 g de $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$, 0.02 g de H_3BO_3 , 0.5 g de CoCl_2 , 20.0 g de ZnCl_2 , 65.0 g de $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, 0.2 g de biotina, 5 ml de H_2SO_4 concentrado. La biotina y las sales se filtraron y se esterilizaron.

El medio usado como iniciador del bioproceso fue el medio YPD que contenía 1% (w/v) de extracto de levadura, 2% (w/v) de peptona y 2% (w/v) de dextrosa.

Todos los cultivos se monitorizaron en un bioreactor esterilizado de 2 litros, acoplado a una unidad termostaticadora que los mantenía a $30 \text{ }^\circ\text{C}$ (Biostat B, Sartorius Biotech, Alemania) (figura 4.3). El pH del medio se mantuvo a 5.5, controlándolo mediante la adición de una disolución amortiguadora del 30% (v/v) de NH_4OH . El

bioproceso se mantuvo bajo continua agitación y con una corriente de aire continua de 2.5 Lmin^{-1} .

4.8. MEDIDAS DE FLUORESCENCIA MULTICANAL ON-LINE

Las medidas de fluorescencia se realizaron con un sensor BioView[®] (*DELTA Light & Optics, Lyngby, Denmark*) [Lindemann 1998] conectado al bioreactor a través de una ventana de cuarzo de 25 mm. de diámetro (figura 4.3). El rango de longitudes de onda monitorizado fue desde 270 a 550 nm en intervalos de 20 nm para la excitación y de 310 a 590 nm, también en intervalos de 20 nm, para la emisión. Se programó el instrumento para realizar medidas cada 10 minutos.

4.9. TRATAMIENTO DE DATOS DE FLUORESCENCIA

Las matrices espectrales se exportaron en formato ASCII para tratamientos posteriores. Debido a las especificaciones de las medidas espectrales, existe un área que no contiene información sobre la intensidad de la señal. Para facilitar la convergencia del algoritmo, se introdujeron valores de NaN (*Not A Number*) en esta área. Pero un gran número de valores NaN también puede afectar a la convergencia (Surribas et al., 2006c; Thygesen et al., 2004), por lo que los valores de NaN se reemplazaron por valores 0 donde se puede asegurar que la intensidad de fluorescencia es, efectivamente, cero (Figura 4.4). El algoritmo PARAFAC usado se encuentra implementado en la PLS-Toolbox 3.5 [PLS], la cual trabaja bajo MatLab v. 7.0 [Mat].

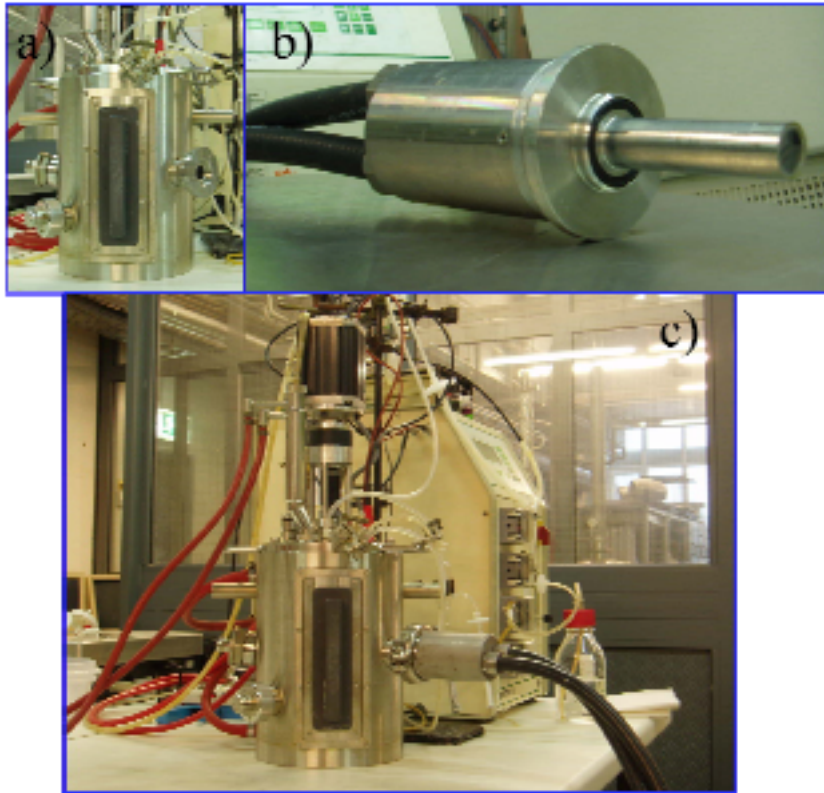


Figura 4.3. a) Reactor esterilizado donde se llevaron a cabo los bioprocesos. b) Detalle de la sonda de fluorescencia multicanal. c) Acoplamiento de la sonda multicanal con el bioreactor (fotos realizadas por Anna Surribas. Escola Tècnica Superior D'Enginyeria, ETSE. Universitat Autònoma de Barcelona)

4.10. MEDIDAS *OFF-LINE*

Para el control del proceso se utilizaron medidas *off-line* de biomasa, metanol, glicerol y lipasa para corroborar las conclusiones obtenidas con la metodología de trabajo *on-line* PARAFAC propuesta.

La biomasa se analizó filtrando 10 mL de muestra con filtros de microfibras de vidrio (Whatman GF/F; Maidstone, UK), lavada con agua destilada y secada a 105 °C hasta pesada constante. La desviación estándar relativa de estas medidas fue del 5%. El metanol y glicerol se determinó mediante HPLC (Merck LaChrom 7000) usando una columna *Inertsil*. La fase móvil usada fue una disolución de 0.1 mol L⁻¹ de NH₄H₂PO₄ tamponado a pH = 2.5. El volumen de inyección fue de 10 µL. Los cromatogramas se

Capítulo 4

analizaron con el programa D7000-HSM (VWR GmbH, Darmstadt, Germany). La desviación estándar relativa de los análisis fue del 4%.

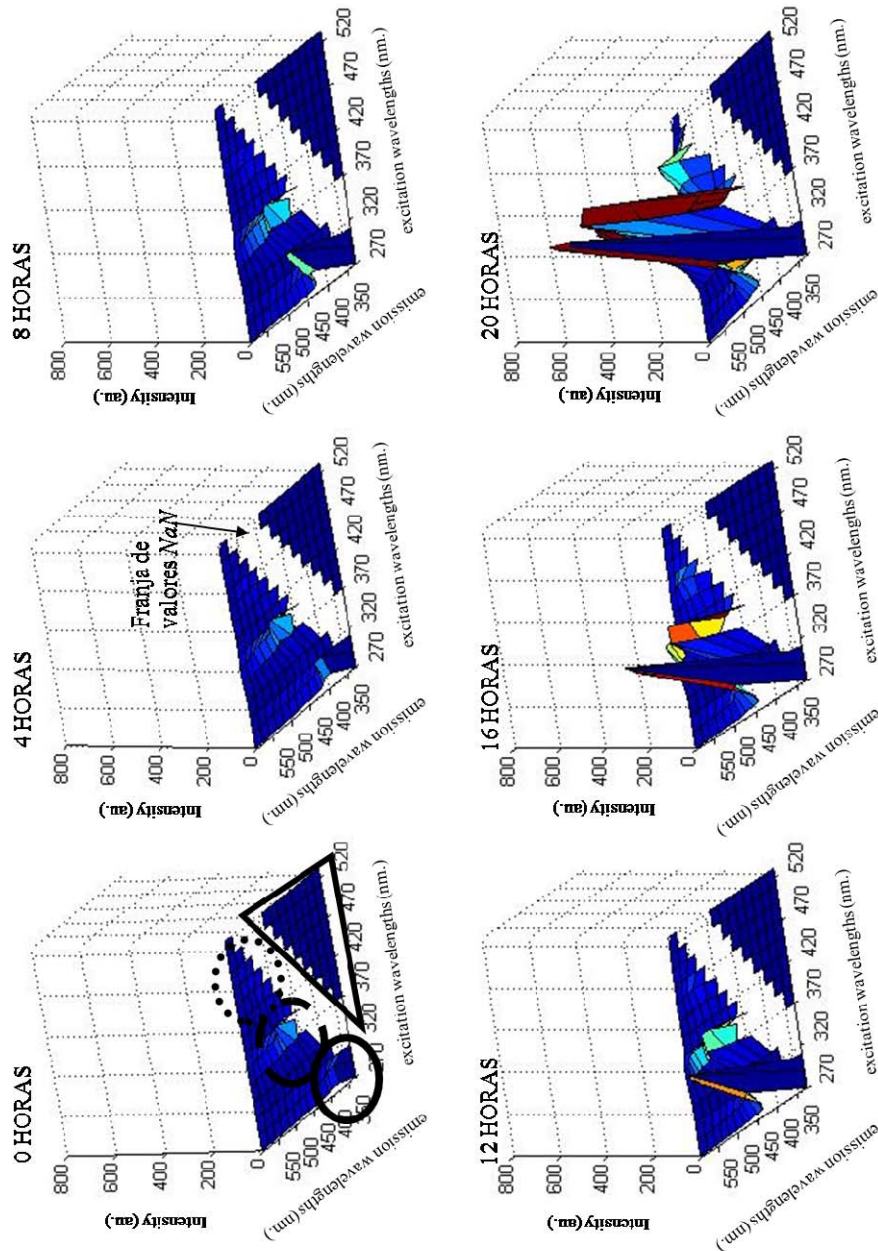


Figura 4.4. Evolución de la señal de fluorescencia con el tiempo para el primer bioproceso. En círculos se señala el área de excitación-emisión para los tres fluoróforos. Triptófano, línea continua; NADH, línea rayada; Rivoflabina, línea punteada. El triángulo indica el área de fluorescencia cero. La franja blanca indica los valores NaN (*Not a Number*).

PARTE III:

ESTUDIO ESPECTROFOTOMÉTRICO DE LOS EQUILIBRIOS ÁCIDO BASE DEL ÁCIDO 8-HIDROXIQUINOLINA-5-SULFÓNICO. INTRODUCCIÓN AL MCR-ALS.

4.11. REACTIVOS E INSTRUMENTAL

Para llevar a cabo la práctica se prepararon las siguientes disoluciones en agua Milli-Q. Todos los reactivos fueron de calidad p.a. (*Merck*):

- Disolución patrón de ácido 8-hidroxiquinolina-5-sulfónico (8HQS) 4.0×10^{-4} M.
- Ácido perclórico 0.023 M.
- Perclorato sódico 1.5 M.
- Disolución tampón de Tris-hidroximetil-aminometano 0.02 M. El pH fue de 7.4. El ajuste de pH se realizó con la disolución de ácido perclórico.
- Hidróxido sódico 0.5 M.

El instrumental necesario es el siguiente:

- Reactor de vidrio de doble pared.
- Baño termostático de recirculación *Thermo Haake DC10*
- Microbureta *Hamilton*
- Bomba peristáltica *Ismatec Reglo*
- Espectrofotómetro UV-Vis equipado con un detector de diodos HP-8452A
UV-Vis Diode Array
- Cubeta de cuarzo de 1.00 cm. de paso de luz.
- Electrodo combinado de pH *Crison* acoplado a un pH-metro *Knick*.
- Agitador magnético.
- Tubos de teflón y tubos FIA de *Tygon* de 0.5 mm de diámetro interno.

- Ordenador personal equipado con MatLab v6.0 [Mat].

4.12. MONTAJE DEL EXPERIMENTO.

El experimento se puede realizar en dos sesiones de laboratorio. En la primera sesión el estudiante realiza la valoración ácido-base del 8HQS y registra los espectros. El tiempo estimado de esta sesión puede ser de 3-4 horas. La segunda sesión de trabajo está dedicada al tratamiento quimiométrico de los datos. Si se tiene cierto conocimiento previo sobre MatLab, esta sesión puede durar 3-4 horas.

Este experimento se puede realizar solamente enfocándolo al tratamiento quimiométrico de los datos, ya que se puede disponer de los datos de la valoración ácido-base a través de los autores [Rodríguez-Rodríguez 2007]. En el material disponible vía página Web [Web 4.2] se encuentra todas las instrucciones necesarias para realizar la experimentación, los datos registrados de la valoración y una *demo* totalmente desarrollada para la aplicación del MCR-ALS usando el programa desarrollado por el profesor Romà Tauler y colaboradores [Web 3.1, Jaumot 2005].

4.12.1. PRIMERA SESIÓN: VALORACIÓN ÁCIDO-BASE DEL 8HQS

Se mantuvo una fuerza iónica constante de 0.5 M de perclorato sódico a lo largo de todo el experimento. Para evitar la precipitación de perclorato potásico en la membrana del electrodo de pH, la disolución interior de KCl se reemplazó por una disolución de NaClO₄ 3M saturada con AgCl. Previamente a la valoración espectrofotométrica, el electrodo fue calibrado siguiendo la metodología propuesta por Gran [Gran 1952, Boiani 1986].

El diagrama de la figura 4.5 muestra el montaje experimental. La valoración ácido-base se llevó a cabo en un reactor de vidrio de doble pared de 50 mL que se mantuvo a 25 ± 0.2 °C y bajo continua agitación. Se añadieron las siguientes disoluciones: 10 mL de la disolución de HClO₄, 10 mL de la disolución de NaClO₄, 10 mL de la disolución de Tris-hidroximetil-aminometano y 10 mL de la disolución de 8HQS. La función de la disolución tampón fue de evitar un salto de pH elevado en el

punto de equivalencia y permitir un cambio espectral más gradual. Se realizaron microadiciones a volumen controlado de la disolución de NaOH.

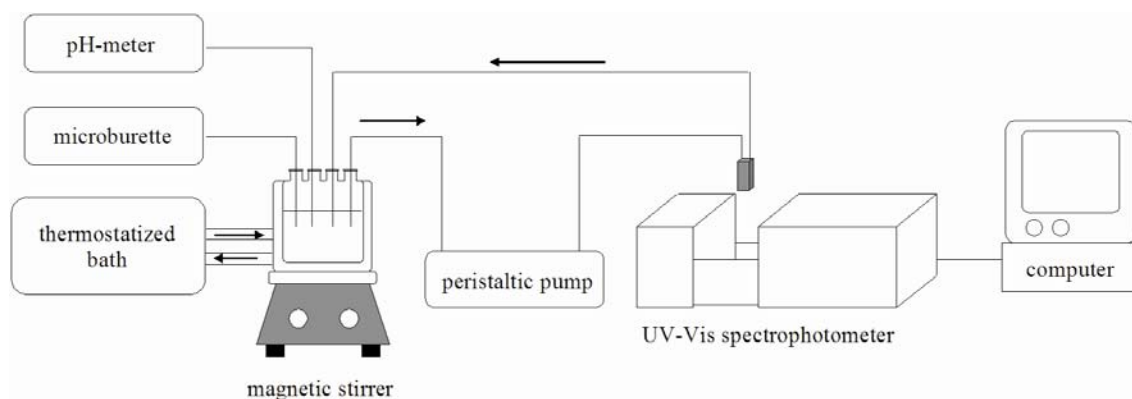


Figura 4.5: Diagrama del montaje de la valoración ácido-base.

El rango de pH fue de 2 a 11, medido con un electrodo de pH combinado. Se registraron los espectros UV-Vis en el intervalo de 250 a 500 nm, cada 2 nm. La disolución se recirculó continuamente a través de una cubeta de cuarzo manteniendo un flujo constante de 2 mL min^{-1} .

4.12.2. SEGUNDA SESIÓN: TRATAMIENTO QUIMIOMÉTRICO DE LOS ESPECTROS REGISTRADOS

Los espectros registrados para la valoración se convirtieron en formato ASCII y se agruparon en una matriz \mathbf{M} ($J \times K$) para su posterior análisis con el programa MCR-ALS [Web 3.1] funcionando bajo MatLab [MAT]. En esta matriz, los J espectros registrados a K longitudes de onda se colocaron en filas en función del pH (figura 3.8).

Se asumió que la absorbancia medida a cada longitud de onda era la suma de las absorbancias individuales de cada componente en el sistema. Como se asume que la absorbancia es el producto de dos variables desconocidas, se tuvieron que aplicar ciertas restricciones (nonegatividad, unimodalidad, sistema cerrado) para no obtener una solución ambigua. Además, se usaron 3 de los espectros registrados en determinados valores de pH como estimaciones iniciales. Más detalles sobre su aplicación se pueden encontrar en el material adicional de los anexos.

