



UNIVERSITAT DE BARCELONA



**Nous desenvolupaments,
aplicacions bioanalítiques i
validació de mètodes de
resolució multivariant**

Joaquim Jaumot Soler

**Tesi Doctoral
20 de juny de 2006**



UNIVERSITAT DE BARCELONA

U

B

NOUS DESENVOLUPAMENTS, APLICACIONS BIOANALÍTiques I VALIDACIÓ DE MÈTODES DE RESOLUCIÓ MULTIVARIANT

Joaquim Jaumot Soler

**Departament de Química Analítica
Barcelona, 2006**



UNIVERSITAT DE BARCELONA



Departament de Química Analítica

**NOUS DESENVOLUPAMENTS,
APLICACIONS BIOANALÍTiques I
VALIDACIÓ DE MÈTODES DE
RESOLUCIÓ MULTIVARIANT**

Tesi doctoral presentada per
Joaquim Jaumot Soler
per optar al grau de
Doctor en Química per la Universitat de Barcelona

Direcció
Dr. Romà Tauler Ferré
Dr. Raimundo Gargallo Gómez

Programa de doctorat
QUÍMICA ANALÍTICA DEL MEDI AMBIENT I DE LA POL·LUCIÓ (Bienni 2001-2003)

NOUS DESENVOLUPAMENTS, APLICACIONS BIOANALÍTiques I VALIDACIÓ DE MÈTODES DE RESOLUCIÓ MULTIVARIANT

Memòria presentada per Joaquim Jaumot Soler
per optar al grau de **Doctor en Química**

El Dr. Romà Tauler Ferré i el Dr. Raimundo Gargallo Gómez,

CERTIFIQUEM

que la present Tesi Doctoral presentada per Joaquim Jaumot Soler ha estat realitzada sota la nostra direcció al Departament de Química Analítica de la Universitat de Barcelona.

Barcelona, març de 2006

Dr. Romà Tauler Ferré

Professor d'Investigació del Departament de Química Ambiental de l'Institut d'investigacions Químiques i Ambientals de Barcelona adscrit al Consejo Superior de Investigaciones Científicas

Dr. Raimundo Gargallo Gómez

Investigador "Ramón y Cajal" adscrit al Departament de Química Analítica de la Universitat de Barcelona

*The laws of Nature are written in the language of mathematics,
the symbols are triangles, circles and other geometrical figures,
without whose help it is impossible to comprehend a single word.
Measure what is measurable, and make measurable what is not so.*

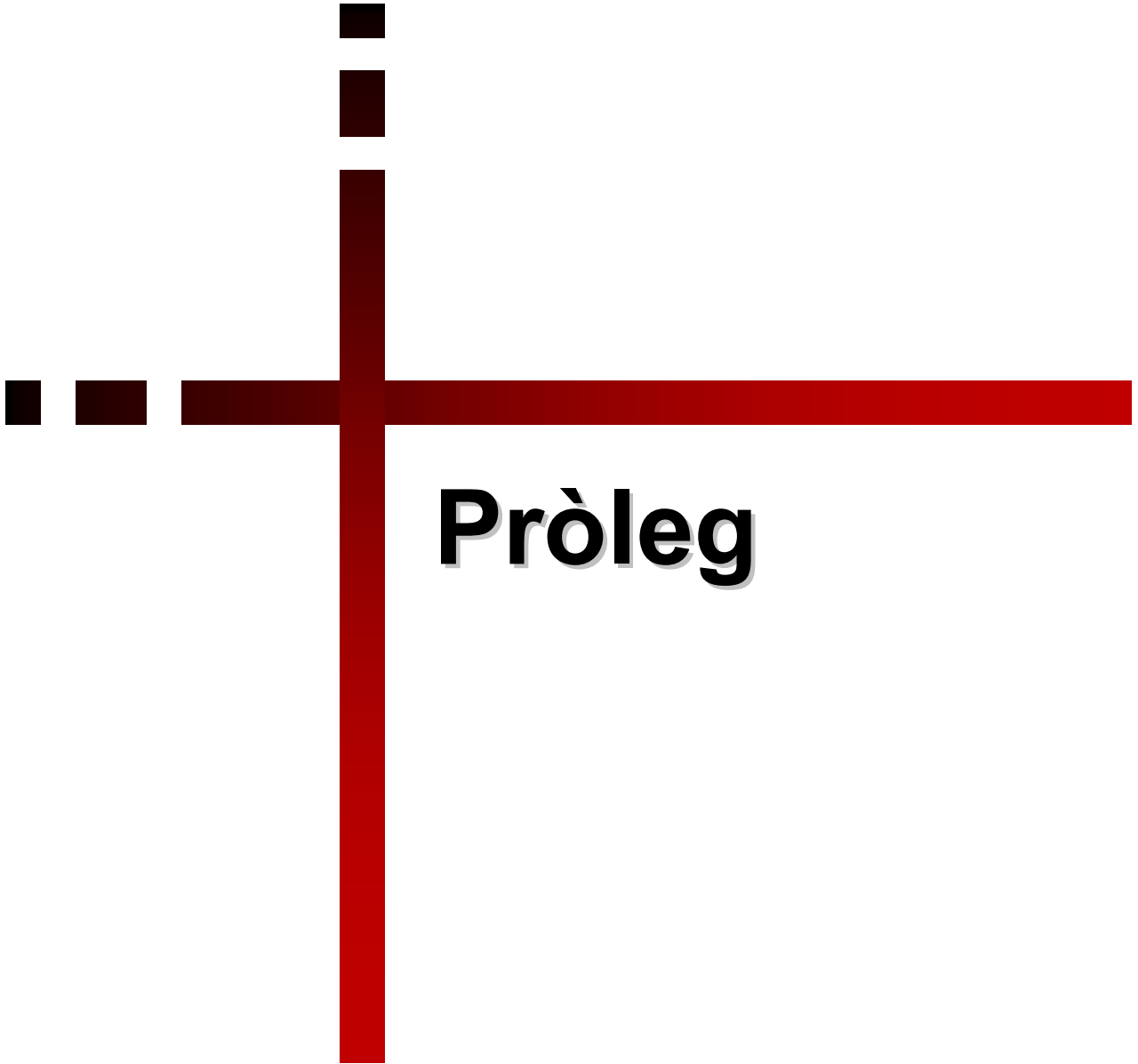
(Galileo Galilei)

A la Susana i a la meva família,

<u>PRÒLEG</u>	1
Estructura de la Tesi Doctoral	3
Antecedents	7
Objectius	9
<u>I. INTRODUCCIÓ</u>	13
1. Mètodes quimiomètrics	15
1.1. Mètodes multivariants d'anàlisi de dades	20
1.2. Modelització flexible	24
1.2.1. Consideracions bàsiques: nombre de components i estimacions inicials	24
1.2.2. Optimització ALS	32
1.2.3. Avaluació de la qualitat dels resultats	42
1.3. Modelització rígida	49
1.3.1. Consideracions bàsiques	49
1.3.2. Aplicació de mètodes de modelització rígida a l'estudi de sistemes en equilibri químic	56
1.3.3. Aplicació de mètodes de modelització rígida a l'estudi de sistemes químics cinètics	61
1.3.4. Limitacions dels mètodes de modelització rígida	66
1.4. Bibliografia	68
2. Química dels àcids nucleics	77
2.1. Estructura dels àcids nucleics	83
2.1.1. Conceptes bàsics i estructures habituals	83
2.1.2. Estructures secundàries dels ADN poc habituals	88
2.2. Variables experimentals que afecten a les conformacions dels àcids nucleics	96
2.3. Descripció dels sistemes experimentals estudiats	103
2.4. Micromatrius d'ADN	105
2.4.1. Conceptes bàsics	105
2.4.2. Tractament de dades	110
2.4.2.1. Mètodes de classificació	112
2.5. Bibliografia	118

3. Procediments experimentals i instrumentació	125
3.1. Realització d'experiments	127
3.2. Tècniques instrumentals	130
3.3. Bibliografia	158
<u>II. RESULTATS I DISCUSSIÓ</u>	161
4. Nous desenvolupaments i validació de mètodes de resolució multivariant	163
4.1. <i>"An user friendly interface for MCR-ALS : a new tool for Multivariate Curve Resolution in MATLAB"</i>	167
4.2. <i>"Estimation of error propagation and prediction intervals in MCR-ALS using resampling methods"</i>	179
4.3. <i>"Quality assesment of the results obtained by Multivariate Curve Resolution analysis of multiple processes of gasoline blending processes"</i>	195
4.4. <i>"Non-negativity constraints for elimination of multiple solutions in fitting of multivariate kinetic models to spectroscopic data"</i>	211
4.5. Discussió dels resultats	223
4.5.1. Nous Desenvolupaments del mètode MCR-ALS	223
4.5.2. Validació de mètodes de resolució multivariant	228
4.6. Bibliografia	237
5. Aplicacions bioanalítiques	241
5.1. <i>"Application of multivariate resolution methods to the study of biochemical and biophysical processes"</i>	247
5.2. <i>"Multivariate curve resolution: a powerful tool for the analysis of conformational transitions in nucleic acids"</i>	263
5.3. <i>"Resolution of parallel and antiparallel oligonucleotide triple helices formation and melting processes by multivariate curve resolution"</i>	275
5.4. <i>"Resolution of a structural competition involving dimeric G-quadruplex and its C-rich complementary strand"</i>	289

5.5. <i>“Multivariate resolution of NMR labile signal by means of hard- and soft-modelling methods”</i>	303
5.6. <i>“Multivariate curve resolution applied to the analysis and resolution of two-dimensional [¹H, ¹⁵N] NMR reaction spectra”</i>	317
5.7. <i>“Exploratory data analysis of DNA microarrays by Multivariate Curve Resolution”</i>	327
5.8. Discussió dels resultats	341
5.8.1. Discussió dels resultats des d'un punt de vista quimiomètric	341
5.8.2. Discussió dels resultats des d'un punt de vista biofísic	350
5.9. Bibliografia	359
<u>III. CONCLUSIONS</u>	361
6. Conclusions	363
<u>IV. EPÍLEG</u>	369



Pròleg

Estructura de la Tesi Doctoral

El treball portat a terme durant la realització d'aquesta Tesi Doctoral s'estructura en tres grans blocs.

En un primer bloc d'introducció es descriuen les bases dels mètodes quimiomètrics i experimentals emprats així com dels sistemes biològics estudiats en aquesta Tesi. Aquest bloc es divideix en tres capítols.

En el capítol primer es fa una descripció dels mètodes quimiomètrics utilitzats, tant de modelització flexible com de modelització rígida.

En el capítol segon s'introdueix la química dels àcids nucleics i es descriuen els factors experimentals que afecten els seus equilibris conformacionals. En aquest capítol també es fa una breu introducció dels fonaments experimentals de la tecnologia de micromatrius d'ADN i dels mètodes estadístics emprats per analitzar aquest tipus de dades.

Finalment, el capítol tercer descriu les tècniques instrumentals i els procediments experimentals que s'han emprat en aquesta Tesi.

En un segon bloc de resultats es descriu detalladament el treball que s'ha realitzat en aquesta Tesi. Aquest bloc es divideix en dos capítols.

El capítol quart mostra els nous desenvolupaments proposats pel mètode de resolució multivariant, MCR-ALS, i la validació dels resultats obtinguts mitjançant aquest mètode. Es mostra, en primer lloc, el desenvolupament d'una nova interfície gràfica pel mètode de Resolució Multivariant de Corbes per Mínims Quadrats Alternats, MCR-ALS, [1]. A continuació es mostren dos treballs sobre la validació del mètode MCR-ALS, tant pel cas d'analitzar un únic experiment [2], com quan s'analitzen diferents experiments simultàniament [3]. Finalment, en el darrer treball d'aquest capítol, s'analitzen les ambigüitats presents en els mètodes de modelització rígida de dades cinètiques i

s'investiga com la imposició de la restricció de no-negativitat durant el procés d'optimització permet obtenir resultats més fiables [4].

El capítol cinquè presenta les aplicacions bioanalítiques investigades mitjançant els mètodes de resolució multivariant. En primer lloc es va dur a terme un estudi bibliogràfic sobre les aplicacions recents d'aquests mètodes a problemes bioanalítics [5]. A continuació es descriuen tres aplicacions d'aquests mètodes quimiomètrics a problemes concrets on s'estudien els equilibris conformacionals de diferents oligonucleòtids. Així, es presenta un treball on s'estudien els equilibris de dimerització d'oligonucleòtids cíclics [6], un treball on s'estudia la formació d'estructures triples paral·leles i antiparal·leles [7], i, finalment, un treball on s'estudia la formació d'estructures quàdruples [8]. D'altra banda, es presenten dos treballs relacionats amb l'aplicació dels mètodes de resolució multivariant per a l'anàlisi de dades de Ressonància Magnètica Nuclear monodimensional [9] i multidimensional [10]. Finalment, es mostra un treball que tracta sobre l'aplicació del mètode MCR-ALS a l'anàlisi de dades de micromatrius d'ADN [11].

Finalment, el bloc final de conclusions presenta de forma resumida els aspectes més significatius dels treballs realitzats tant en el desenvolupament com en la validació dels mètodes de resolució multivariant. També es resumeix la importància que tenen les aplicacions realitzades d'aquests mètodes quimiomètrics en l'anàlisi de dades que provenen de diferents tècniques experimentals i en l'estudi de nous sistemes i problemes biològics, de complexitat creixent.

Llista de treballs originals relacionats amb aquesta Tesi Doctoral

- (1) Jaumot, J.; Gargallo, R.; de Juan, A.; Tauler, R. (2005) A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB. *Chemometrics and Intelligent Laboratory Systems*, **76**, 101-110.
- (2) Jaumot, J.; Gargallo, R.; Tauler, R. (2004) Noise propagation and error estimations in multivariate curve resolution alternating least squares using resampling methods. *Journal of Chemometrics*, **18**, 327-340.
- (3) Jaumot, J.; Menezes, J. C.; Tauler, R. (Enviat per a la seva publicació) Quality assessment of the results obtained by Multivariate Curve Resolution analysis of multiple runs of gasoline blending processes.
- (4) Jaumot, J.; Gemperline, P. J.; Stang, A. (2005) Non-negativity constraints for elimination of multiple solutions in fitting of multivariate kinetic models to spectroscopic data. *Journal of Chemometrics*, **19**, 97-106.
- (5) Jaumot, J.; Vives, M.; Gargallo, R. (2004) Application of multivariate resolution methods to the study of biochemical and biophysical processes. *Analytical Biochemistry*, **327**, 1-13.
- (6) Jaumot, J.; Escaja, N.; Gargallo, R.; Gonzalez, C.; Pedroso, E.; Tauler, R. (2002) Multivariate curve resolution: a powerful tool for the analysis of conformational transitions in nucleic acids. *Nucleic Acids Research*, **30**, e92.
- (7) Jaumot, J.; Aviño, A.; Eritja, R.; Tauler, R.; Gargallo, R. (2003) Resolution of parallel and antiparallel oligonucleotide triple helices formation and melting processes by multivariate curve resolution. *Journal of Biomolecular Structure & Dynamics*, **21**, 267-278.
- (8) Jaumot, J.; Eritja, R.; Tauler, R.; Gargallo, R. (2006) Resolution of a structural competition involving dimeric G-quadruplex and its C-rich complementary strand. *Nucleic Acids Research*, **34**, 206-216.
- (9) Jaumot, J.; Vives, M.; Gargallo, R.; Tauler, R. (2003) Multivariate resolution of NMR labile signals by means of hard- and soft-modelling methods. *Analytica Chimica Acta*, **490**, 253-264.
- (10) Jaumot, J.; Marchan, V.; Gargallo, R.; Grandas, A.; Tauler, R. (2004) Multivariate curve resolution applied to the analysis and resolution of two-dimensional [¹H,¹⁵N] NMR reaction spectra. *Analytical Chemistry*, **76**, 7094-7101.
- (11) Jaumot, J.; Tauler, R.; Gargallo, R. (Enviat per a la seva publicació) Exploratory data analysis of DNA microarrays by Multivariate Curve Resolution.

Antecedents

Aquest treball s'integra en una de les línies de recerca del Grup de Quimiometria del Departament de Química Analítica de la Universitat de Barcelona. Aquesta línia de recerca es centra en el desenvolupament de mètodes quimiomètrics d'anàlisi multivariant de dades i en la seva aplicació a l'estudi analític dels canvis de conformació i/o de les interaccions entre biomolècules.

Els antecedents d'aquesta línia de recerca són els estudis sobre determinacions de constants d'estabilitat a partir de mesures potenciomètriques i espectrofotomètriques dels anys vuitanta, i en la utilització de programes de càlcul basats en optimitzacions minimoquadràtiques basades en models fisicoquímics [1, 2]. Des d'aleshores s'ha produït una revolució tecnològica que ha afectat tant a l'adquisició de les dades experimentals com al tractament d'aquestes. D'aquesta forma s'ha passat de recollir mesures univariants (és a dir, on es mesura un únic canal com, per exemple, l'absorbància a una sola longitud d'ona a cada temps de reacció d'una cinètica) a mesures multivariants que permeten obtenir tot un espectre o adquirir les intensitats de fluorescència de milers de canals en uns segons. Com és lògic, aquest increment en la capacitat d'adquisició de dades ha provocat també un increment de la complexitat d'aquestes dades i, per tant, la necessitat de l'aparició de mètodes numèrics que siguin capaços de tractar-les.

Durant aquests anys, l'activitat investigadora del Grup de Quimiometria ha tingut una evolució paral·lela. Així, a partir dels treballs de determinació potenciomètrica i espectrofotomètrica de constants dels anys vuitanta [1, 2] mitjançant mètodes de modelització rígida, es va passar al desenvolupament, a principi de la dècada dels noranta, de mètodes de modelització flexible. Així, en primer lloc es va desenvolupar el programa SPFAC [3], implementat en FORTRAN, en el qual es realitzava una optimització iterativa per mínims quadrats alternats (ALS) basada en l'anàlisi per components principals i en l'anàlisi de factors evolutius de la matriu de dades experimental. Més endavant es van dur a terme una sèrie de millores en aquest mètode consistentes en la possibilitat d'introduir restriccions durant el procés d'optimització ALS i en l'anàlisi simultània de diferents experiments que van donar lloc a l'aparició del mètode conegut com a Resolució Multivariant de Corbes en l'entorn de

treball de MATLAB [4, 5]. Més endavant, aquest mètode es va anomenar MCR-ALS (a partir de les inicials en anglès de *Multivariate Curve Resolution by Alternating Least Squares*). A partir de la proposta inicial d'aquest mètode, el treball en el sí del grup de recerca s'ha diversificat en diferents línies.

Una primera línia de treball es basa en el desenvolupament i millora del mètode MCR-ALS, que s'ha convertit en un mètode quimiomètric molt flexible i versàtil [6]. Així, per una banda, s'ha augmentat i millorat el nombre de restriccions que es poden aplicar durant l'optimització de forma que augmenti el sentit químic de les solucions obtingudes [7, 8]. Entre aquestes restriccions cal destacar la referent a la imposició d'un model químic (ja sigui cinètic o d'equilibri basat en la llei d'acció de masses) que possibilita la resolució d'un sistema fent servir alhora mètodes de modelització flexible i rígida [9, 10]. Per altra banda, una altra gran avantatge del mètode MCR-ALS és la possibilitat d'analitzar simultàniament diverses matrius de dades correlacionades [7, 11] de manera que es poden estudiar tant sistemes formats per un únic experiment seguit per una única espectroscòpia, com estudiar un nombre x d'experiments seguits per un nombre y d'espectroscòpies, o com estudiar dades experimentals procedents d'espectroscòpies multidimensionals o d'imatges [12, 13].

Una altra línia de treball del grup ha permès avaluar la qualitat dels resultats obtinguts pel mètode MCR-ALS. Així, s'han estudiat en detall els problemes que poden afectar la fiabilitat dels resultats obtinguts degut a problemes de deficiència de rang [14] o l'efecte de les ambigüitats rotacionals i d'escala associades a aquests mètodes de resolució [7, 15]. En molts casos, s'ha aconseguit proposar restriccions i condicions de resolució (per exemple, a partir de l'anàlisi simultània de diversos experiments) que permeten la minimització d'aquests problemes i, en molts casos, aconseguir la seva eliminació i obtenir solucions úniques.

Finalment, una altra de les línies d'investigació es basa en l'aplicació dels mètodes quimiomètrics desenvolupats per nosaltres a diversos tipus de sistemes experimentals. D'aquesta forma s'aplica aquest mètode a l'estudi de la interacció d'àcids nucleics amb metalls [16, 17], als equilibris conformacionals de proteïnes [18, 19] i àcids nucleics [20, 21] o a la determinació de contaminants en mostres ambientals [22, 23].

Tenint en compte aquests antecedents històrics i les línies de recerca en el grup, es poden definir els objectius de la Tesi Doctoral que es presenta en aquesta memòria.

Objectius

Els tres objectius principals que s'han intentat assolir en aquesta Tesi Doctoral queden definits en el mateix títol: "Nous desenvolupaments, aplicacions bioanalítiques i validació dels mètodes de resolució multivariant".

El primer objectiu d'aquesta Tesi ha estat el de contribuir al desenvolupament i extensió del mètode de Resolució Multivariant de Corbes (MCR-ALS) a partir de la creació d'una interfície gràfica. Aquesta interfície pretén facilitar la utilització del mètode MCR-ALS permetent una millor selecció de diferents opcions possibles a utilitzar durant l'optimització per mínims quadrats alternats i facilitar el seu ús d'una manera més intuïtiva.

El segon objectiu d'aquesta Tesi ha estat l'estudi de la fiabilitat dels resultats que s'obtenen mitjançant l'aplicació del mètode MCR-ALS. S'ha investigat tant l'efecte de les ambigüitats rotacionals i d'escala com els efectes de la propagació del soroll experimental a les solucions obtingudes pel mètode MCR-ALS.

Finalment, el tercer objectiu d'aquesta Tesi ha estat l'aplicació del mètode MCR-ALS a l'estudi dels equilibris conformacionals de diferents oligonucleòtids mitjançant diverses tècniques instrumentals, especialment espectroscòpiques.

Referències

- (1) Casassas, E.; Tauler, R.; Filella, M. (1986) A Critical Comparison of Computer-Programs for the Potentiometric Determination of Stability-Constants. *Analytica Chimica Acta*, **191**, 399-411.
- (2) Casassas, E.; Filella, M.; Tauler, R. (1986) Assessment of the Results Obtained from Different Computer-Programs Applied to Potentiometric Complexation Data. *Analytica Chimica Acta*, **191**, 413-423.
- (3) Tauler, R.; Casassas, E. (1989) Application of Principal Component Analysis to the Study of Multiple Equilibria Systems. Study of Copper(II)/Salicylate/Mono-, Di- and Triethanolamine Systems. *Analytica Chimica Acta*, **223**, 257-268.
- (4) Tauler, R.; Casassas, E.; Izquierdo-Ridorsa, A. (1991) Self-Modeling Curve Resolution in Studies of Spectrometric Titrations of Multi-Equilibria Systems by Factor-Analysis. *Analytica Chimica Acta*, **248**, 447-458.
- (5) Tauler, R.; Kowalski, B.; Fleming, S. (1993) Multivariate Curve Resolution Applied to Spectral Data from Multiple Runs of an Industrial-Process. *Analytical Chemistry*, **65**, 2040-2047.
- (6) de Juan, A.; Tauler, R. (2003) Chemometrics applied to unravel multicomponent processes and mixtures - Revisiting latest trends in multivariate resolution. *Analytica Chimica Acta*, **500**, 195-210.
- (7) Tauler, R.; Smilde, A.; Kowalski, B. (1995) Selectivity, Local Rank, 3-Way Data-Analysis and Ambiguity in Multivariate Curve Resolution. *Journal of Chemometrics*, **9**, 31-58.
- (8) de Juan, A.; Van der Heyden, Y.; Tauler, R.; Massart, D. L. (1997) Assessment of new constraints applied to the alternating least squares method. *Analytica Chimica Acta*, **346**, 307-318.
- (9) de Juan, A.; Maeder, M.; Martinez, M.; Tauler, R. (2000) Combining hard- and soft-modelling to solve kinetic problems. *Chemometrics and Intelligent Laboratory Systems*, **54**, 123-141.
- (10) Diewok, J.; de Juan, A.; Maeder, M.; Tauler, R.; Lendl, B. (2003) Application of a combination of hard and soft modeling for equilibrium systems to the quantitative analysis of pH-modulated mixture samples. *Analytical Chemistry*, **75**, 641-647.
- (11) Tauler, R. (1995) Multivariate curve resolution applied to second order data. *Chemometrics and Intelligent Laboratory Systems*, **30**, 133-146.
- (12) de Juan, A.; Maeder, M.; Hancewicz, T.; Tauler, R. (2005) Local rank analysis for exploratory spectroscopic image analysis. Fixed Size Image Window-Evolving Factor Analysis. *Chemometrics and Intelligent Laboratory Systems*, **77**, 64-74.

- (13) de Juan, A.; Tauler, R.; Dyson, R.; Marcolli, C.; Rault, M.; Maeder, M. (2004) Spectroscopic imaging and chemometrics: a powerful combination for global and local sample analysis. *Trac-Trends in Analytical Chemistry*, **23**, 70-79.
- (14) Saurina, J.; Hernandez-Cassou, S.; Tauler, R.; Izquierdo-Ridorsa, A. (1998) Multivariate resolution of rank-deficient spectrophotometric data from first-order kinetic decomposition reactions. *Journal of Chemometrics*, **12**, 183-203.
- (15) Tauler, R. (2001) Calculation of maximum and minimum band boundaries of feasible solutions for species profiles obtained by multivariate curve resolution. *Journal of Chemometrics*, **15**, 627-646.
- (16) Gargallo, R.; Tauler, R.; Izquierdo-Ridorsa, A. (1997) Acid-base and copper(II) complexation equilibria of poly(inosinic)-poly(cytidylic) acid. *Biopolymers*, **42**, 271-283.
- (17) Vives, M.; Tauler, R.; Moreno, V.; Gargallo, R. (2001) Study of the interaction of a cis-dichloroaminopyrrolidine Pt(II) complex and the polynucleotide poly(I)-poly(C) acid by means of H-1-NMR and multivariate curve resolution. *Analytica Chimica Acta*, **446**, 439-450.
- (18) Navea, S.; de Juan, A.; Tauler, R. (2002) Detection and resolution of intermediate species in protein folding processes using fluorescence and circular dichroism spectroscopies and multivariate curve resolution. *Analytical Chemistry*, **74**, 6031-6039.
- (19) Navea, S.; de Juan, A.; Tauler, R. (2003) Modeling temperature-dependent protein structural transitions by combined near-IR and mid-IR spectroscopies and multivariate curve resolution. *Analytical Chemistry*, **75**, 5592-5601.
- (20) Gargallo, R.; Tauler, R.; Izquierdo-Ridorsa, A. (1997) Application of a multivariate curve resolution procedure to the analysis of second-order melting data of synthetic and natural polynucleotides. *Analytical Chemistry*, **69**, 1785-1792.
- (21) Vives, M.; Eritja, R.; Tauler, R.; Marquez, V. E.; Gargallo, R. (2004) Synthesis, stability, and protonation studies of a self-complementary dodecamer containing the modified nucleoside 2'-deoxyzebularine. *Biopolymers*, **73**, 27-43.
- (22) Tauler, R.; Barcelo, D.; Thurman, E. M. (2000) Multivariate correlation between concentrations of selected herbicides and derivatives in outflows from selected US midwestern reservoirs. *Environmental Science & Technology*, **34**, 3307-3314.
- (23) Pere-Trepat, E.; Hildebrandt, A.; Barcelo, D.; Lacorte, S.; Tauler, R. (2004) Fast chromatography of complex biocide mixtures using diode array detection and multivariate curve resolution. *Chemometrics and Intelligent Laboratory Systems*, **74**, 293-303.

B

I



htrducci ó

