



UNIVERSITAT DE BARCELONA



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

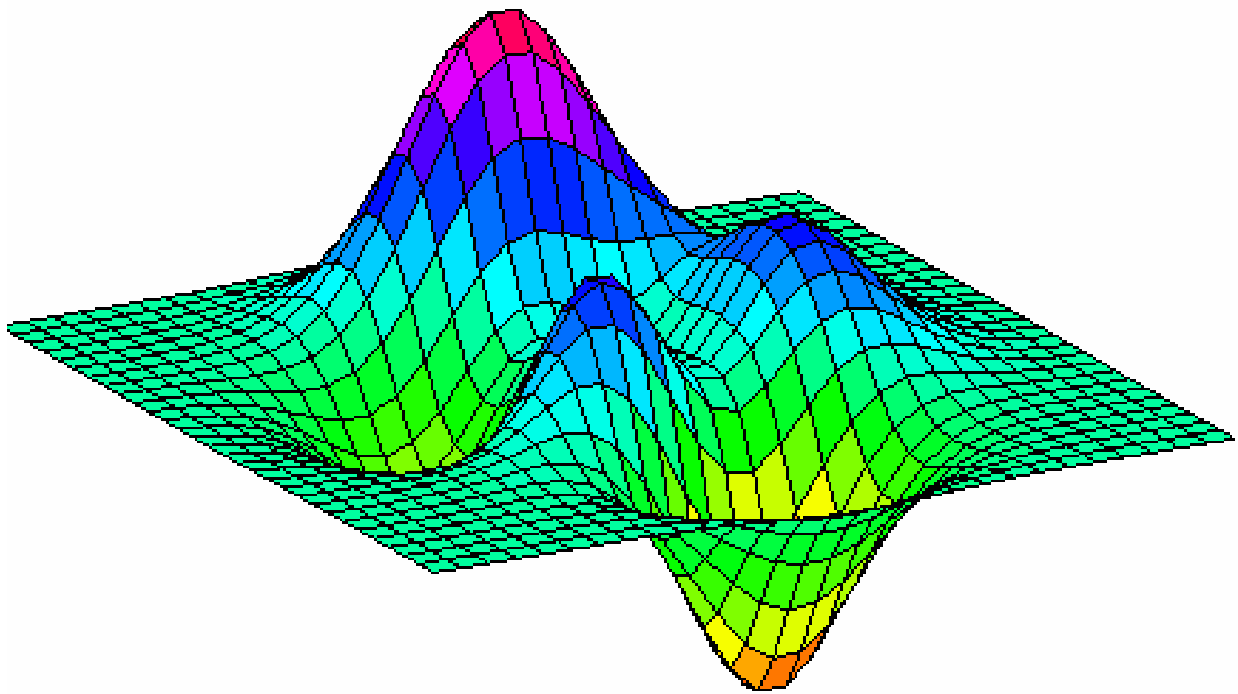
Joaquim Jaumot Soler

**Tesi Doctoral
20 de juny de 2006**

Bloc

II

**Resultats i
Discussió**



Capítol 4

**Nous desenvolupaments i
validació de mètodes de
resolució multivariant**

En aquest apartat es presenta la contribució d'aquesta Tesi Doctoral a la millora del mètode de Resolució Multivariant de Corbes per Mínims Quadrats Alternats (MCR-ALS), a l'avaluació de la fiabilitat dels resultats obtinguts pel mètode MCR-ALS i dels obtinguts en l'anàlisi de dades cinètiques mitjançant mètodes de modelització rígida.

En els darrers anys, el mètode MCR-ALS ha assolit una notable popularitat en l'anàlisi i la resolució de processos i mesclades multicomponent. Aquesta acceptació ha estat deguda a la capacitat d'adaptació d'aquest mètode a sistemes experimentals diversos i a la flexibilitat en la implementació de diferents restriccions durant el procés d'optimització iterativa. Malgrat aquests avantatges, els algoritmes existents presentaven la dificultat de la destresa que, en alguns casos, era necessària per dur a terme la selecció i aplicació de restriccions adequades, sobretot quan es treballa amb un nombre elevat de matrius. Per tal de superar aquests inconvenients, es va dirigir el treball de desenvolupament del mètode MCR-ALS cap a la creació d'una interfície gràfica entre l'usuari i l'algoritme d'optimització que permetés la selecció de les restriccions d'una forma molt més intuïtiva.

En segon lloc, es presenten els treballs que s'han dut a terme en aquesta Tesi Doctoral per tal d'avaluar la fiabilitat dels resultats obtinguts mitjançant els mètodes de resolució multivariant. Els treballs resultants s'han agrupat en dos blocs en els quals es diferencien els estudis realitzats amb el mètode de modelització flexible MCR-ALS i els realitzats amb un mètode de modelització rígida.

En el primer d'aquest dos blocs s'estudia com avaluar l'efecte de l'error experimental en les solucions MCR-ALS i de les ambigüitats intrínseques del mètode.

- El primer treball estudia la qualitat dels resultats obtinguts per MCR-ALS quan s'analitza una única matriu de dades. Per tal de dur a terme aquest estudi es van proposar i comparar diferents mètodes que han permès avaluar la propagació de l'error experimental a les solucions obtingudes. Aquests estudis es van dur a terme amb dades simulades i experimentals corresponents als equilibris àcid-base del polinucleòtid poly(I)-poly(C) seguits per absorció molecular a l'ultraviolat.

- El segon treball estudia la qualitat dels resultats obtinguts per MCR-ALS en el cas d'analitzar més d'una matriu de dades simultàniament. A partir dels resultats obtinguts al treball anterior es va seleccionar el mètode més apropiat per dur a terme l'estudi de la propagació de l'error experimental a les solucions obtingudes. A més, s'estudia l'efecte de les ambigüitats rotacionals en les solucions obtingudes i com l'estratègia d'analitzar simultàniament diversos experiments contribueix a minimitzar aquests efectes. Aquests estudis es van dur a terme amb dades experimentals corresponents a mesclades de diferents tipus de benzines seguits mitjançant espectroscòpia d'infraroig proper obtingudes, durant una estada de recerca en el grup del Prof. Dr. José C. Menezes de l'Instituto Superior Técnico de Lisboa.

En el segon bloc, es presenta el treball realitzat sobre l'ambigüitat d'intercanvi de les constants que apareix quan s'intenta modelitzar determinats mecanismes cinètics utilitzant mètodes de modelització rígida. Així, es va observar que quan el mecanisme consta de dos o més reaccions es produeix una ambigüitat en la resolució deguda a l'intercanvi de les constants de velocitat associades a aquestes reaccions. Es proposa l'aplicació de la restricció de no-negativitat a l'interior del cicle d'optimització iterativa per tal d'eliminar aquesta ambigüitat. Aquest estudi s'ha dut a terme analitzant conjunts de dades simulats i experimentals en els quals es produïa aquest tipus d'ambigüitat. Aquest treball es va dur a terme en una estada de recerca en el grup del Prof. Dr. Paul J. Gemperline a la East Carolina University de Greenville (Carolina del Nord, Estats Units).

4.1. “An user friendly interface for MCR-ALS : a new tool for Multivariate Curve Resolution in MATLAB”

*Chemometrics and Intelligent
Laboratory Systems,
(2005),
76, 101-110*

*Joaquim Jaumot
Raimundo Gargallo
Anna de Juan
Romà Tauler*

4.2. “*Estimation of error propagation and prediction intervals in MCR-ALS using resampling methods*”

Journal of Chemometrics,
(2004),
18, 327-340

Joaquim Jaumot
Raimundo Gargallo
Romà Tauler

4.3. “Quality assesment of the results obtained by Multivariate Curve Resolution analysis of multiple processes of gasoline blending processes ”

(enviat per a la seva publicació)

*Joaquim Jaumot
José C. Menezes
Romà Tauler*

4.4. “Non-negativity constraints for elimination of multiple solutions in fitting of multivariate kinetic models to spectroscopic data ”

Journal of Chemometrics,
(2005)
19, 97-106

Joaquim Jaumot
Paul J. Gemperline
Alexandra Stang

4.5. Discussió dels resultats

En aquest capítol es presenta una discussió global dels treballs presentats en aquest capítol com a complement de la que es pot trobar als articles presentats anteriorment. La discussió s'ha dividit en dos subapartats segons l'aspecte tractat als treballs. En primer lloc es tracta sobre la interfície gràfica desenvolupada pel mètode de resolució multivariant de corbes MCR-ALS. En segon lloc es tracta sobre la validació dels resultats obtinguts tant pels mètodes de modelització flexible com rígida, tenint en compte les ambigüitats intrínseques a aquests mètodes i els possibles efectes del soroll experimental.

4.5.1. Nous desenvolupaments del mètode MCR-ALS

El mètode de resolució multivariant de corbes mitjançant mínims quadrats alternats (MCR-ALS) ha esdevingut un mètode quimiomètric força utilitzat per a l'anàlisi multicomponent de mescles complexes. Aquest fet es pot comprovar en fer una recerca bibliogràfica del concepte "Multivariate Curve Resolution", el qual retorna aproximadament 300 entrades al portal "ISI Web of Knowledge" [1] a data de febrer de 2006. Una part d'aquestes entrades corresponen a treball realitzats en el sí del Grup de Quimiometria (com, per exemple, els treballs realitzats en aquesta Tesi Doctoral), però també hi ha treballs de grups de recerca tant catalans [2-4] com europeus [5-7] o americans [8-10]. A més, en dur a terme aquesta recerca s'observa també una gran varietat d'àmbits d'aplicació diferents en els quals s'ha aplicat amb èxit aquest mètode com, per exemple, el seguiment de processos industrials [11, 12] i biofísics [13, 14], l'anàlisi d'imatges [15, 16], l'anàlisi de dades ambientals [17, 18] o l'anàlisi de dades de micromatrius d'ADN [19, 20]. Malgrat aquesta acceptació per part de la comunitat científica, temes com la selecció de les restriccions i de les opcions d'optimització en l'algoritme ALS resulten complexes sobretot en el cas de l'anàlisi de diferents experiments simultàniament.

Així, en aquesta Tesi s'ha pretès facilitar i millorar la interacció entre l'usuari i el programa d'optimització ALS. S'ha intentat posar aquest mètode quimiomètric a l'abast d'un usuari sense coneixements previs de programació MATLAB® [21], ni d'utilització prèvia del programa MCR-ALS. Aquest objectiu s'ha dut a terme mitjançant el desenvolupament d'una interfície gràfica dins de l'entorn MATLAB® (apartat 4.1.).

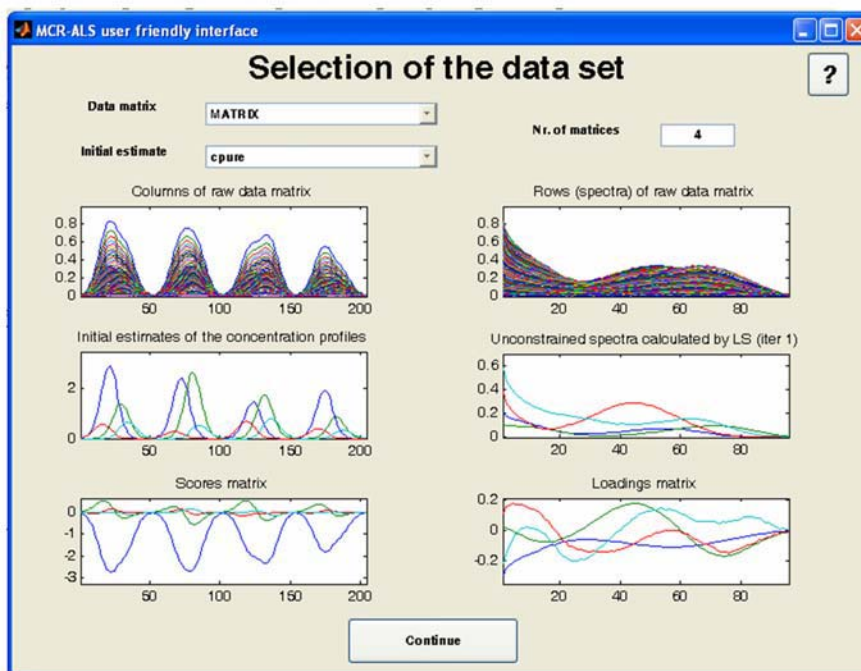
La nova interfície gràfica consta de les següents finestres:

- Finestra per a la introducció de la informació inicial (Figura 4.1.a). En primer lloc, es selecciona la matriu de dades a analitzar, les estimacions inicials dels perfils de concentració o espectrals que es faran servir en la resolució i el nombre de matrius a analitzar (cal tenir en compte les dues possibilitats d'augmentació de matrius per files i per columnes [11]). Així, després de seleccionar aquestes matrius es mostra una representació gràfica de les dades experimentals per files i per columnes. També es mostra una representació de les estimacions inicials dels perfils de concentració o espectrals seleccionats i dels perfils espectrals o de concentració, respectivament, calculats per mínims quadrats a partir de les dades experimentals i de les estimacions inicials. Finalment, es mostra a la part inferior una representació de les matrius de *scores* i de *loadings* obtingudes per l'anàlisi de components principals (PCA) de les dades experimentals.
- Finestres per a la selecció de les restriccions a aplicar durant l'optimització. En aquest cas, s'han de distingir dos casos diferents.

D'una banda, quan s'analitza una única matriu, s'obrirà la finestra de selecció de restriccions que es mostra a la Figura 4.1.b. En aquesta finestra únicament apareixen les restriccions que es poden aplicar quan s'analitza una única matriu. Així, es podran seleccionar les restriccions de no-negativitat, unimodalitat, sistema tancat (balanç de matèria) i selectivitat, tant pels perfils de concentració com pels espectrals. La selecció de les restriccions es porta a terme mitjançant clics del ratolí a caselles de selecció, botons de radio o menús emergents que s'activen progressivament quan l'usuari selecciona una determinada restricció de forma que el guien en el procés.

D'altra banda, si s'analitza una matriu augmentada (ja sigui en la direcció de les files o de les columnes) es tindrà la finestra de selecció de restriccions que es mostra a la Figura 4.1.c. Prèviament a l'aparició d'aquesta finestra, la interfície demanarà informació sobre la matriu de dades augmentada estudiada, és a dir, el nombre d'experiments i/o tècniques estudiades i les dimensions de cadascuna d'aquestes submatrius. La selecció de les diverses restriccions presenta, en aquest

a

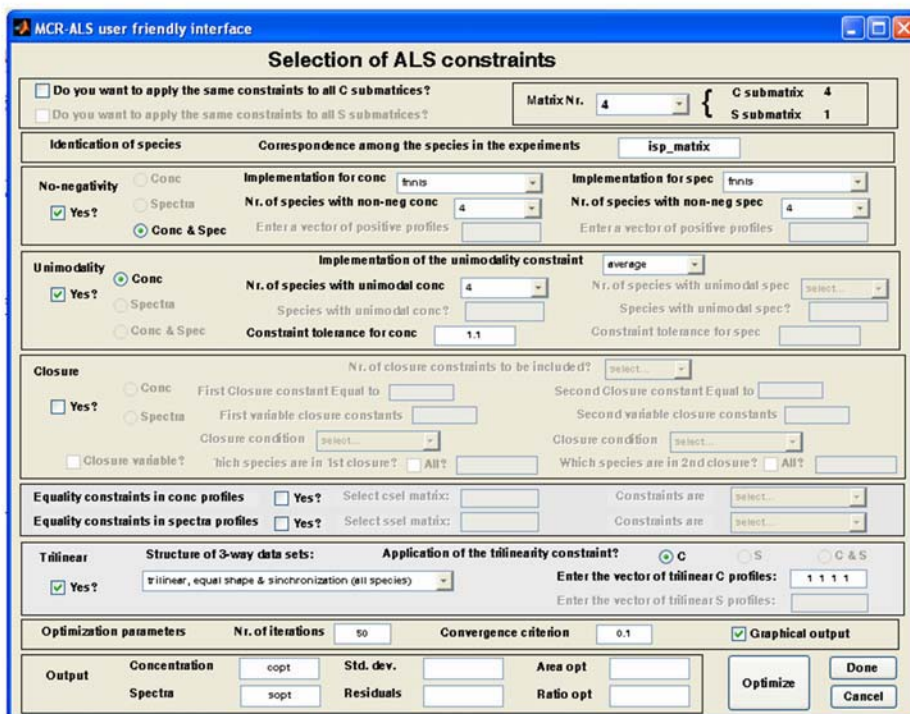


b

The screenshot shows the 'Selection of ALS constraints' window. It is divided into several sections:

- No-negativity:** Includes radio buttons for 'Conc', 'Spectra', and 'Conc & Spec'. 'Yes?' is checked. 'Implementation for conc' and 'Implementation for spec' are set to 'fnvis'. 'Nr. of species with non-neg conc' and 'Nr. of species with non-neg spec' are both set to '4'. There are input fields for 'Enter a vector of positive profiles'.
- Unimodality:** Includes radio buttons for 'Conc', 'Spectra', and 'Conc & Spec'. 'Yes?' is checked. 'Implementation of the unimodality constraint' is set to 'average'. 'Nr. of species with unimodal conc' is set to '4'. 'Nr. of species with unimodal spec' is set to 'select...'. There are input fields for 'Species with unimodal conc?' and 'Species with unimodal spec?'. 'Constraint tolerance for conc' is set to '1.1'. There is an input field for 'Constraint tolerance for spec'.
- Closure:** Includes radio buttons for 'Conc' and 'Spectra'. 'Yes?' is unchecked. 'Nr. of closure constraints to be included?' is set to 'select...'. There are input fields for 'First Closure constant Equal to', 'Second Closure constant Equal to', 'First variable closure constants', and 'Second variable closure constants'. There are dropdown menus for 'Closure condition' and 'Closure condition'. There are checkboxes for 'Closure variable?' and 'Which species are in 1st closure?' (set to 'All?'). There are checkboxes for 'Which species are in 2nd closure?' (set to 'All?').
- Equality constraints in conc profiles:** 'Yes?' is checked. 'Select csel matrix:' is set to 'cse1_mat'. 'Constraints are' is set to 'equal than'.
- Equality constraints in spectra profiles:** 'Yes?' is unchecked. 'Select ssel matrix:' is empty. 'Constraints are' is set to 'select...'. There are input fields for 'cse1_mat' and 'sssel_mat'.
- Optimization parameters:** 'Nr. of iterations' is set to '50'. 'Convergence criterion' is set to '0.1'. 'Graphical output' is checked.
- Output:** Includes checkboxes for 'Concentration' (set to 'copt') and 'Spectra' (set to 'sopt'). There are input fields for 'Std. dev.', 'Area opt', and 'Ratio opt'. There are also input fields for 'Residuals'. 'Optimize', 'Done', and 'Cancel' buttons are at the bottom right.

C



d

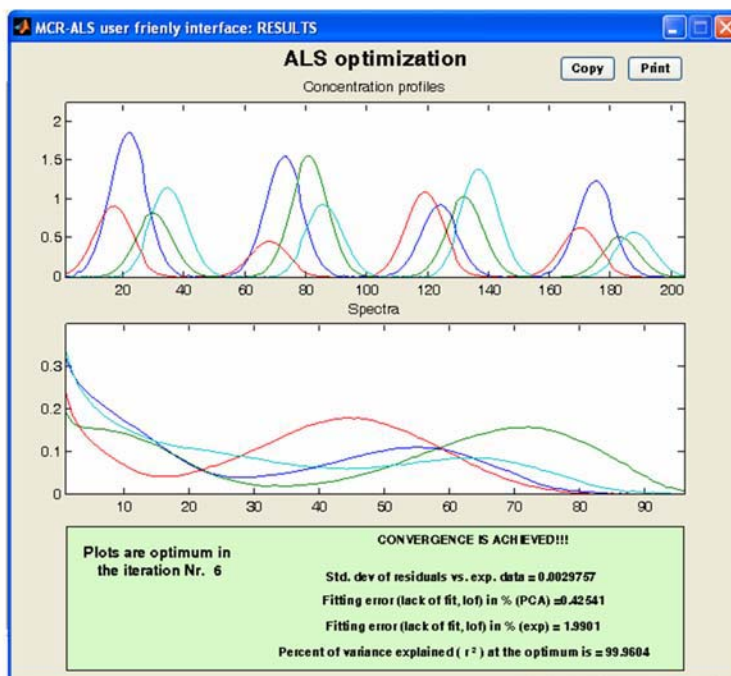


Figura 4.1. Captures de pantalla de la interfície gràfica. (a) Introducció de la informació, (b) selecció de restriccions en el cas d'un únic experiment, (c) selecció de restriccions en el cas de múltiples experiments i (d) presentació de resultats.

cas, diferències respecte la selecció de restriccions per a una única matriu. Així, en primer lloc, s'ha de decidir si es vol aplicar les mateixes restriccions per a totes les matrius o aplicar diferents restriccions a cada matriu (tant en el cas de que siguin diferents experiments com diferents tècniques). En el cas d'aplicar les mateixes restriccions a totes les matrius, el procés de selecció de les restriccions serà similar al cas d'una única matriu. Però, en el cas d'aplicar diferents restriccions a cada matriu el procés serà diferent. Així, caldrà seleccionar les restriccions per a la primera matriu i , posteriorment, canviar la matriu a la qual s'apliquen les restriccions fent servir el menú emergent que es troba a la part superior dreta de la figura. En cas que sigui necessari per a l'optimització tornar a seleccionar alguna restricció, aquesta reinicialitzarà el seu valor i l'usuari podrà tornar a seleccionar la forma d'implantar la restricció (per exemple, algoritme utilitzat en la no-negativitat) o el nombre d'espècies a les qual s'aplicarà. Finalment, cal destacar que hi ha la possibilitat d'aplicar la restricció de trilinearitat que únicament es pot aplicar si l'estructura de les dades ho permet.

En els dos casos a la part inferior de la finestra apareixen unes opcions comuns. Així, es té l'opció de canviar paràmetres de l'optimització (com, per exemple, el nombre d'iteracions i el criteri de convergència) i donar noms a les variables de sortida. Finalment, si es prem el botó d'inici *Optimize* s'iniciarà el procés d'optimització.

- Finestra per a la presentació dels resultats (Figura 4.1.d). Al final de l'optimització (o durant el procés si es selecciona l'opció de sortida gràfica) es mostren els perfils de concentració i espectrals resolts. A més, a la part inferior es mostra informació que permet avaluar la qualitat del procés d'optimització. Així, s'informa sobre si l'optimització és convergent o divergent i es mostren paràmetres com la quantitat de variància explicada o la manca d'ajust que permeten avaluar la bondat del procés.

Per tal de dur a terme aquesta interfície gràfica s'ha fet servir una eina per desenvolupar interfícies gràfiques de l'entorn de treball MATLAB®, la funció *guide*. A partir d'aquí es podran dissenyar les finestres de la interfície afegint els continguts de l'entorn MATLAB® que es considerin necessaris (eixos per tal de fer representacions

de les matrius, texts fixes, texts variables, caselles de selecció, botons de radio, menús emergents, etc.). Una vegada es té l'estructura física de la interfície s'ha omplert de contingut les accions referents als diferents elements. Així, cal tenir en compte, si al obrir una finestra, algun element (per exemple, un menú desplegable) necessita fer una lectura de dades o quines són les conseqüències de seleccionar una matriu o una restricció.

Per tal de facilitar la propagació i utilització d'aquesta interfície gràfica s'ha publicat a l'adreça web <http://www.ub.es/gesq/mcr/als2004.htm>. En aquesta, a més de poder descarregar els fitxers dels programes es poden trobar tutorials i exemples de conjunts de dades. L'acceptació de la interfície creada en aquesta Tesi Doctoral es pot demostrar a partir de les més de 150 descàrregues del programa registrades des de novembre de 2004 fins a finals de l'any 2005.

Finalment, cal destacar que hi ha altres eines que permeten dur a terme la resolució de matrius de dades mitjançant interfícies gràfiques. Entre aquestes hi ha productes comercials com la *PLS Toolbox* [22] per a MATLAB® o els programa *The Unscrambler*® [23]. A més, hi ha també programes de lliure distribució com les eines *GUIPRO* [24] o *CuBatch* [25].

4.5.2. Validació dels mètodes de resolució multivariant

En aquesta memòria s'han presentat tres treballs que realitzats amb l'objectiu d'analitzar les fonts d'incertesa en les solucions obtingudes mitjançant els mètodes de resolució multivariant. Es poden dividir aquests estudis en dos grups en funció de si tracten un mètode de modelització flexible com el MCR-ALS o un mètode de modelització rígida.

I. Validació del mètode de Resolució Multivariant de Corbes per Mínims Quadrats Alternats (MCR-ALS)

En un primer treball (apartat 4.2.) es va dur a terme aquesta avaluació en el cas de considerar un únic experiment mentre que en un segon treball (apartat 4.3.) es va dur a terme considerant més d'un experiment simultàniament.

Abans d'exposar individualment els dos efectes principals que poden afectar les solucions obtingudes mitjançant el mètode MCR-ALS, cal tenir en compte les causes de cadascun d'aquests efectes. Així, mentre que les anomenades bandes de solucions factibles (*feasible bands*) són degudes a les ambigüitats rotacionals, les bandes de confiança (*confidence bands*) de les solucions obtingudes mitjançant els mètodes de remostratge (*resampling*) són conseqüència tant de l'efecte de la propagació de soroll, com de l'ambigüitat rotacional que és inherent al procés de resolució. Una vegada s'ha aclarit aquest punt es pot passar a tractar individualment cadascun d'aquests efectes.

Efecte de les ambigüitats rotacionals

Les ambigüitats rotacionals són inherents als mètodes basats en l'anàlisi de factors com és el cas del MCR-ALS (veure apartat 1.2.2.). Per això, és important determinar les bandes de solucions factibles per a un determinat ajust de les dades i aplicant unes restriccions determinades. Per això s'han analitzat els efectes en els dos treballs esmentats anteriorment en el cas d'analitzar una única matriu (4.2. i 4.3.) i en el cas d'analitzar diverses matrius simultàniament (4.3.) emprant el mètode proposat per Tauler [26].

En el cas d'analitzar una única matriu, la presència de les ambigüitats rotacionals es veu totalment afectada per les restriccions imposades durant el procés de resolució. Així, si durant la resolució no s'imposa cap restricció, l'anomenada banda de solucions factibles serà molt ampla provocant que un nombre gran de combinacions lineals dels perfils de les solucions sigui possible. En canvi, en augmentar el nombre de restriccions, es produeix una reducció de l'amplada d'aquestes bandes de solucions factibles. En alguns casos, sobretot si es pot aplicar la restricció de selectivitat (o de rang local) durant el procés de resolució, aquesta banda de solucions factibles es pot arribar a col·lapsar en una solució única. Exemples de les diferències que es poden observar en les bandes factibles de les solucions depenent de les restriccions aplicades es troben a la Figura 1 del treball de l'apartat 4.2. i a la Figura 5 de l'apartat 4.3.

Un dels principals avantatges de treballar amb més d'un experiment simultàniament és l'augment de la fiabilitat dels resultats obtinguts. Un exemple molt clar d'aquesta millora s'obté comparant les Figures 5 i 6 del treball de l'apartat 4.3. Com s'ha dit

abans, la Figura 5 correspon a la determinació de les bandes factibles de les solucions obtingudes a l'analitzar cada experiment individualment mentre que a la Figura 6 es mostren les bandes factibles obtingudes quan tant l'anàlisi com la determinació de les bandes factibles es va dur a terme per a tots els experiments simultàniament. A partir d'aquestes figures es pot concloure que, degut entre altres coses, a la sobredeterminació del sistema per la gran quantitat d'informació que proporciona analitzar experiments diferents i, especialment, al compliment més fàcil de les condicions de resolució total, les ambigüitats rotacionals en les solucions obtingudes tenen un efecte molt més petit, la qual cosa provoca un increment significatiu de la qualitat dels resultats.

Efecte de la propagació de soroll

L'efecte de la propagació del soroll experimental en les solucions obtingudes mitjançant el mètode MCR-ALS s'ha estudiat en els dos treballs presentats per a una única matriu (4.2.) i per a diverses matrius (4.3.). Com a punt de partida s'ha de tenir en compte que els resultats i les conclusions obtingudes en aquests treballs tenen una gran dependència de l'estructura de les dades i de les restriccions que s'han aplicat en cada cas. Això implica que malgrat les conclusions generals poden ser vàlides en una àmplia varietat de sistemes experimentals i dades procedents de diverses tècniques espectroscòpiques, l'extrapolació de conclusions concretes com, per exemple, els límits d'error, s'han de fer amb cura.

En el cas d'estudiar un únic experiment (veure en els dos treballs presentats), es va obtenir que les solucions obtingudes fins a un nivell de soroll afegit d'aproximadament un 2% es podien considerar òptimes, ja que pràcticament no existien diferències amb les obtingudes en absència de soroll. A un nivell de soroll afegit del 3% les solucions obtingudes comencen a mostrar els efectes de la propagació de soroll (diferències més grans entre els perfils teòrics i resolts, estimació de paràmetres amb errors més grans), però encara es poden considerar acceptables. A partir de nivells de soroll afegit d'un 5% les distorsions en els perfils resolts fan difícil la seva interpretació i, per tant, es pot considerar que en dades que presentin aquests nivells de soroll els resultats ja no es poden considerar fiables. A més, en aquests nivells de soroll, és evident que a l'efecte de la propagació de soroll se li ha de sumar l'efecte de les ambigüitats rotacionals.

En el cas d'estudiar diverses matrius simultàniament (apartat 4.3.) es va observar una millora substancial en els resultats obtinguts. D'aquesta forma, a nivells d'error afegit baixos, els resultats són bons com en el cas anterior però, la principal diferència es troba a nivells de soroll afegit intermedis. Així, mentre que en el cas d'un únic experiment a un nivell del 5% els resultats ja no es podien considerar fiables, en aquest cas el marge de confiança arriba fins a nivells d'aproximadament del 7%. Aquest fet torna a demostrar les avantatges de l'estratègia de treballar amb matrius de dades augmentades degut a l'increment d'informació i a la major sobredeterminació del sistema.

Una vegada s'han tractat separatament aquests dos aspectes de les possibles fonts d'error en resolució multivariant es pot concloure que únicament un anàlisi tenint en compte la informació procedent dels dos factors permet avaluar correctament la qualitat dels resultats obtinguts.

Finalment, es poden discutir altres resultats que s'ha obtingut en els treballs realitzats en aquesta Tesi Doctoral.

Comparació de mètodes de remostratge

L'estimació dels errors associats als perfils de concentració i espectrals obtinguts mitjançant el mètode MCR-ALS presenta una dificultat inherent al mètode. Així, el mètode MCR-ALS es pot considerar que és un mètode d'optimització no lineal d'un nombre molt elevat de paràmetres i en el qual s'apliquen, a més, restriccions també no lineals. Aquesta no-linealitat de l'optimització dificulta l'aplicació de fórmules exactes per a la determinació de les incerteses o errors en els paràmetres calculats, els quals són possibles d'obtenir en el cas dels mètodes lineals. Per aquest motiu, i donada la importància de l'avaluació de l'estimació dels errors i de la propagació del soroll en els mètodes quimiomètrics, s'han proposat diferents mètodes estadístics per poder fer aquestes estimacions.

En el treball presentat al punt 4.2. es va dur a terme la comparació de tres d'aquests mètodes: les simulacions de MonteCarlo, el mètode de l'addició de soroll i el mètode *Jackknife* (veure Introducció, capítol 1, per a una descripció detallada dels diferents

mètodes). A la Taula 4.1. es presenta un resum dels principals resultats quantitius i qualitius obtinguts pels diferents mètodes.

Taula 4.1. Comparació dels resultats obtinguts pels diferents mètodes de remostratge.

	Simulacions de MonteCarlo	Addició de soroll	Jackknife
Tipus de dades	Simulades	Experimentals	Experimentals
Temps necessari i Recursos informàtics	Petit	Petit	Gran
Tipus de soroll	Conegut	Conegut i desconegut	Conegut i desconegut
Remostratge	No	Sí	Si
Valors de manca d'ajust i desviació estàndard (nivell de soroll 1%)	2,113 (0,030)	2,398 (0,036)	2,386 (0,043)
Desviació estàndard dels valors estimats del pK_{a1} (nivell de soroll 1%)	0,007	0,006	0,009
Desviació estàndard dels valors estimats del pK_{a2} (nivell de soroll 1%)	0,013	0,026	0,024

La comparació dels resultats obtinguts pels diferents mètodes es va dur a terme amb l'objectiu de seleccionar el mètode més eficient per a treballs futurs. A partir dels resultats quantitius (Taula 4.1), s'ha determinat que el mètode que presenta millors resultats són les simulacions de MonteCarlo. Malgrat això, aquest mètode presenta l'inconvenient de treballar amb dades simulades, la qual cosa pot no ser útil en intentar extrapolat els resultats obtinguts a dades experimentals. Els dos mètodes de remostratge van proporcionar resultats quantitativament molt similars. Tanmateix, la comparació va resultar molt favorable pel mètode d'addició de soroll ja que, a més de proporcionar resultats del mateix ordre, presenta uns requeriments informàtics (i en conseqüència de temps) molt menors que el mètode de *Jackknife*. Per tant, es va decidir seleccionar el mètode d'addició de soroll com a mètode de remostratge en treballs futurs com, per exemple, el que es presenta al punt 4.3.

Influència de les restriccions en les solucions obtingudes

Una de les principals virtuts del mètode MCR-ALS és la varietat i flexibilitat que presenta en la imposició de restriccions. Per tal de conèixer millor l'efecte que tenen les restriccions sobre el procés de resolució i sobre els paràmetres de qualitat a partir dels quals es determina la bondat de l'optimització, al treball del punt 4.2. es va dur a terme un estudi comparatiu sobre unes dades simulades, els resultats del qual es mostren a la Taula II del treball.

A partir dels resultats obtinguts, es pot concloure que l'aplicació d'un major nombre de restriccions implica una millor resolució del sistema i, en el cas estudiat, una estimació més precisa dels valors de pK_a . Al mateix temps, s'observa un lleuger increment del valor de manca d'ajust en augmentar el nombre de restriccions aplicades la qual cosa ens indica que la solució òptima des d'un punt de vista químic (la que compleix les restriccions aplicades) pot ser que no sigui la solució òptima matemàticament. De la mateixa manera, els resultats obtinguts ens mostren que els valors de manca d'ajust són molt petits quan no s'imposen restriccions o les restriccions imposades es compleixen a les dades experimentalment. Aquests resultats obtinguts en el cas ideal de treballar amb dades simulades i amb una quantitat de soroll mínima ens permeten treure conclusions dels valors de manca d'ajust quan s'analitzen dades experimentals.

II. Validació dels mètodes de modelització rígida per a l'ajust de processos cinètics

Els principals avantatges dels mètodes de modelització rígida respecte als mètodes de modelització flexible són l'eliminació de les ambigüitats rotacionals o d'intensitat inherents als mètodes d'anàlisi de factors, i el menor efecte del soroll en les solucions obtingudes. No obstant això, i tal i com s'ha indicat a l'apartat de "Limitacions" (veure punt 1.3.4), hi ha una sèrie de situacions on això no és totalment cert i que no es poden menysprear. Així, a part de l'aspecte clau que fa referència a la necessitat de conèixer *a priori* el sistema químic que s'està estudiant per tal de proposar el mecanisme adient, en el cas de l'anàlisi de dades procedents de reaccions cinètiques hi ha una possible ambigüitat (de natura similar a l'ambigüitat rotacional) que provoca l'aparició de diferents solucions que presenten exactament el mateix ajust. D'aquesta forma, si no es seleccionen unes estimacions inicials suficientment properes a les solucions finals, el procés d'optimització pot convergir cap a una solució incorrecta. En alguns casos, aquesta optimització dolenta pot ser detectada pels valors negatius que apareixen als perfils espectrals [27-29], però altres vegades no es pot detectar tant

fàcilment. Aquest fet és el que provoca que l'anàlisi de dades cinètiques mitjançant mètodes de modelització rígida s'hagi de fer amb especial cura per tal de seleccionar un model adequat i proposar unes estimacions inicials bones.

A la bibliografia es poden trobar diferents mètodes per eliminar aquest tipus d'ambigüitat en el cas de treballar amb dades univariants [27, 28] però, per a dades multivariants únicament, existeixen mètodes que impliquen una aplicació de restriccions externament al procés d'optimització [30-32]. En el treball presentat a l'apartat 4.4. s'estudia la detecció i eventual eliminació d'aquesta ambigüitat i com fer la selecció de les estimacions inicials dels paràmetres per iniciar l'optimització. Per aconseguir aquest objectiu es va implementar una restricció de no-negativitat a l'interior del bucle d'optimització utilitzat en el procediment d'optimització per mínims quadrats no lineals implementat al programa GUIPRO [24].

Es demostra, que en mecanismes amb més d'una etapa, s'acostuma a tenir més d'una solució que ajusta de manera òptima les dades experimentals i que habitualment es troba relacionada amb l'intercanvi de les constants cinètiques associades a cada pas del mecanisme. La demostració de l'equivalència de les solucions és fàcilment demostrable en el cas de tenir el mecanisme $A \rightarrow B \rightarrow C$ (veure el treball 4.4.) i de la mateixa manera es pot realitzar la demostració per a casos més complexos. Una forma il·lustrativa de certificar l'existència d'aquesta ambigüitat és mitjançant el càlcul de la superfície de resposta associada als valors de la suma de quadrats dels residuals quan es consideren de forma sistemàtica molts dels valors possibles per a cadascun dels paràmetres (constants) que es volen optimitzar tal, i com es mostra a la Figura 1 del treball 4.4. En casos més complexos, on el nombre de paràmetres a ajustar és més gran, la representació gràfica no és tan fàcil d'obtenir i d'interpretar, com es mostra a la Figura 4.2. on es té el volum de resposta corresponent a l'optimització de tres constants de velocitat.

Es tracta, doncs, d'eliminar aquesta ambigüitat ja que pot impedir la determinació correcta de la solució òptima des d'un punt de vista químic. En els darrers anys han aparegut diverses modificacions de mètodes de modelització flexible en les quals s'imposen restriccions de model químic que forcen a les solucions obtingudes mitjançant aquests mètodes a seguir un determinat model proposat, ja sigui cinètic (llei de la velocitat de reacció) [33] o d'equilibri químic (llei d'acció de masses) [34].

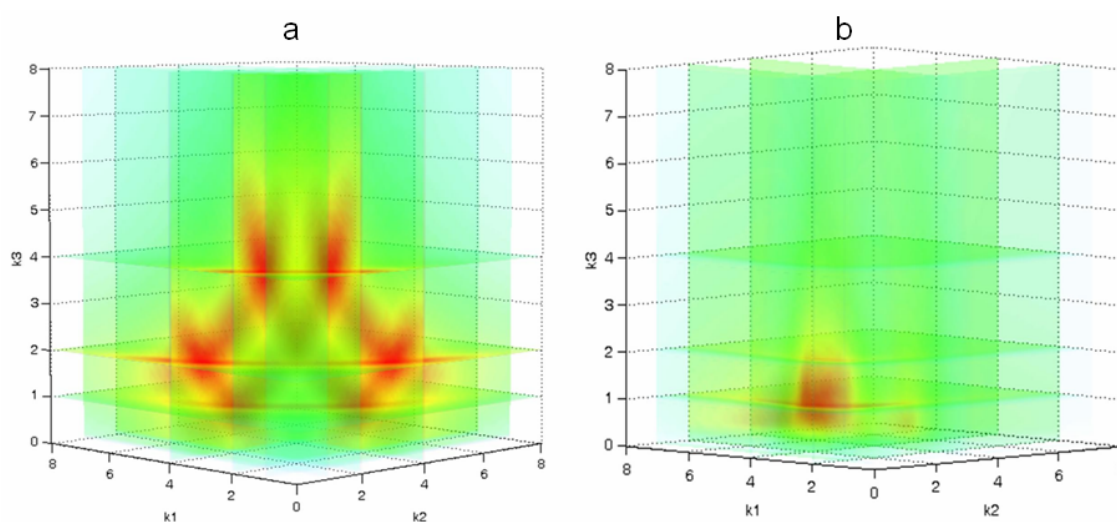


Figura 4.2. Representació de volum de resposta de la suma de quadrats dels residuals per a l'ajust del model $A \rightarrow B \rightarrow C \rightarrow D$ al conjunt de dades D_2 del treball presentat a l'apartat 4.4. (a) Sense restricció de no-negativitat on s'observen 6 mínims de resposta diferents, (b) Amb restricció de no-negativitat on únicament s'observa un mínim de resposta. La gradació de color indica la magnitud del valor associat de la suma de quadrats dels residuals segons l'ordre de major a menor verd – groc – vermell.

Per tal d'eliminar aquesta ambigüïtat en el cas de modelització rígida es va seguir el camí invers, ja que s'aplica una restricció com la de no-negativitat pels perfils espectrals resolts durant el procés d'optimització no lineal per mínims quadrats. D'aquesta forma s'eviten totes les solucions que inclouen perfils espectrals negatius i s'aconsegueix guiar el procés d'optimització cap a la solució correcta fent desaparèixer els mínims locals equivalents a l'òptim (veure Figura 4.2.b) però amb solucions amb valors negatius o poc raonables des d'un punt de vista químic. Així, l'aplicació d'aquesta restricció proporciona una major robustesa i fiabilitat a la resolució, al mateix temps que disminueix la dependència dels resultats amb les estimacions inicials. A més, es fa una comprovació molt simple per verificar si la solució obtinguda és la òptima consistent en continuar l'optimització dels paràmetres obtinguts sense aplicar la restricció de no-negativitat.

Finalment, cal tenir en compte, a més, que es poden eliminar algunes de les ambigüïtats existents a l'anàlisi de processos cinètics a partir de les alternatives següents:

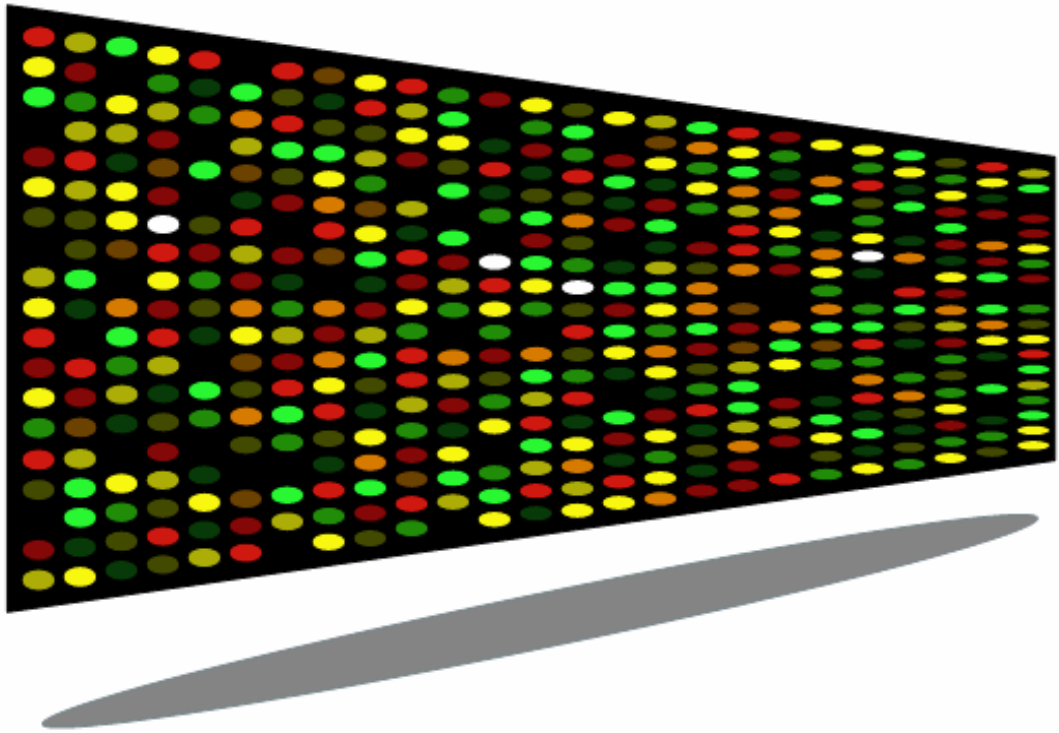
- Anàlisi global de diferents experiments mitjançant mètodes de modelització rígida [35, 36]. En aquest cas es tracta de dur a terme experiments canviant les condicions experimentals. D'aquesta forma s'aconsegueix una sobredeterminació del sistema la qual serveix per eliminar una part important de les solucions ambigües. Els fonaments d'aquest tipus d'anàlisi són molt similars als que s'han descrit pel mètode de resolució MCR-ALS en el cas de treballar amb vàries matrius (veure apartat 1.2.2.).
- Anàlisi de les dades corresponents a un únic experiment mitjançant un mètode de modelització híbrida rígida-flexible com, per exemple, el mètode MCR-ALS amb restriccions de models cinètics [33] (veure apartat 1.2.2.). L'avantatge de la utilització d'aquests tipus de mètodes és deguda a que s'elimina la necessitat de realitzar unes estimacions inicials dels paràmetres que es volen estimar. Així, s'utilitza la part de l'algoritme de modelització flexible per estimar els perfils de concentració, de forma que és la informació pròpia del sistema experimental la que guia el procés d'optimització de modelització rígida.

4.6. Bibliografia

- (1) <http://portal.isiknowledge.com/>. ISI Web of Knowledge [v3.0]
- (2) Garrido, M.; Larrechi, M. S.; Rius, F. X.; Tauler, R. (2005) Calculation of band boundaries of feasible solutions obtained by Multivariate Curve Resolution Alternating Least Squares of multiple runs of a reaction monitored by NIR spectroscopy. *Chemometrics and Intelligent Laboratory Systems*, **76**, 111-120.
- (3) Pasamontes, A.; Callao, M. P. (2004) Determination of amoxicillin in pharmaceuticals using sequential injection analysis and multivariate curve resolution. *Analytica Chimica Acta*, **515**, 159-165.
- (4) Rodriguez-Cuesta, M. J.; Boque, R.; Rius, F. X.; Vidal, J. L. M.; Frenich, A. G. (2005) Development and validation of a method for determining pesticides in groundwater from complex overlapped HPLC signals and multivariate curve resolution. *Chemometrics and Intelligent Laboratory Systems*, **77**, 251-260.
- (5) Antunes, M. C. G.; da Silva, J. C. G. E. (2005) Multivariate curve resolution analysis excitation-emission matrices of fluorescence of humic substances. *Analytica Chimica Acta*, **546**, 52-59.
- (6) Richards, S.; Ropic, M.; Blackmond, D.; Walmsley, A. (2004) Quantitative determination of the catalysed asymmetric transfer hydrogenation of 1-methyl-6,7-dimethoxy-3,4-dihydroisoquinoline using in situ FTIR and multivariate curve resolution. *Analytica Chimica Acta*, **519**, 1-9.
- (7) Ruckebusch, C.; Duponchel, L.; Huvenne, J. P.; Caudron, A.; Boilet, L.; Cornard, J. P.; Merlin, J. C.; de Juan, A. (2005) Chemometric strategies for the study of the complexation of Al(III) ions with model molecule of humic substances from UV-vis data sets. *Analytica Chimica Acta*, **544**, 337-344.
- (8) Gallagher, N. B.; Shaver, J. M.; Martin, E. B.; Morris, J.; Wise, B. M.; Windig, W. (2004) Curve resolution for multivariate images with applications to TOF-SIMS and Raman. *Chemometrics and Intelligent Laboratory Systems*, **73**, 105-117.
- (9) Sanehez-Ponce, R.; Rutan, S. C. (2005) Steady state kinetic model constraint for Multivariate Curve Resolution-Alternating Least Squares analysis. *Chemometrics and Intelligent Laboratory Systems*, **77**, 50-58.
- (10) Van Benthem, M. H.; Keenan, M. R.; Haaland, D. M. (2002) Application of equality constraints on variables during alternating least squares procedures. *Journal of Chemometrics*, **16**, 613-622.
- (11) 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.
- (12) Van Gorkom, L. C. M.; Hancewicz, T. M. (1998) Analysis of DOSY and GPC-NMR experiments on polymers by multivariate curve resolution. *Journal of Magnetic Resonance*, **130**, 125-130.

- (13) 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.
- (14) Vives, M.; Gargallo, R.; Tauler, R. (2001) Analytical characterization of the conformational transitions of polynucleotides by means of different molecular spectroscopies and multivariate curve resolution. *Analytical Biochemistry*, **291**, 1-10.
- (15) 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.
- (16) 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.
- (17) Mallat, E.; Barcelo, D.; Tauler, R. (1997) Degradation study of benomyl and carbendazim in water by liquid chromatography and multivariate curve resolution methods. *Chromatographia*, **46**, 342-350.
- (18) 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.
- (19) Martinez, M. J.; Aragon, A. D.; Rodriguez, A. L.; Weber, J. M.; Timlin, J. A.; Sinclair, M. B.; Haaland, D. M.; Werner-Washburne, M. (2003) Identification and removal of contaminating fluorescence from commercial and in-house printed DNA microarrays. *Nucleic Acids Research*, **31**, e18.
- (20) Timlin, J. A.; Haaland, D. M.; Sinclair, M. B.; Aragon, A. D.; Martinez, M. J.; Werner-Washburne, M. (2005) Hyperspectral microarray scanning: impact on the accuracy and reliability of gene expression data. *Bmc Genomics*, **6**, -.
- (21) MATLAB® versió (5.3 - 7.0). The Mathworks. Natick, MA, Estats Units.
- (22) PLS Toolbox® v. 3.5. for Matlab. Eigenvector Research. Wenatchee, WA, Estats Units.
- (23) The Unscrambler® v. 9.0. CAMO ASA, Oslo, Noruega.
- (24) Gemperline, P. J.; Cash, E. (2003) Advantages of soft versus hard constraints in self-modeling curve resolution problems. Alternating least squares with penalty functions. *Analytical Chemistry*, **75**, 4236-4243.
- (25) Gourvenec, S.; Tomasi, G.; Durville, C.; Di Crescenzo, E.; Saby, C. A.; Massart, D. L.; Bro, R.; Oppertheim, G. (2005) CuBatch, a MATLAB (R) interface for n-mode data analysis. *Chemometrics and Intelligent Laboratory Systems*, **77**, 122-130.

- (26) 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.
- (27) Vajda, S.; Rabitz, H. (1994) Identifiability and Distinguishability of General Reaction Systems. *Journal of Physical Chemistry*, **98**, 5265-5271.
- (28) Andraos, J. (1999) The problem of distinguishability of rate constants in the two-step consecutive sequence $A \rightarrow B \rightarrow C$. *Canadian Journal of Chemistry-Revue Canadienne De Chimie*, **77**, 565-576.
- (29) Purves, R. D. (1996) Multiple solutions, illegal parameter values, local minima of the sum of squares, and anomalous parameter estimates in least-squares fitting of the two-compartment pharmacokinetic model with absorption. *Journal of Pharmacokinetics and Biopharmaceutics*, **24**, 79-101.
- (30) Bijlsma, S.; Boelens, H. F. M.; Smilde, A. R. (2001) Determination of rate constants in second-order kinetics using UV-visible spectroscopy. *Applied Spectroscopy*, **55**, 77-83.
- (31) Bijlsma, S.; Smilde, A. K. (2000) Estimating reaction rate constants from a two-step reaction: a comparison between two-way and three-way methods. *Journal of Chemometrics*, **14**, 541-560.
- (32) Thurston, T. J.; Brereton, R. G. (2002) Estimation of second order rate constants using chemometric methods with kinetic constraints. *Analyst*, **127**, 659-668.
- (33) 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.
- (34) 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.
- (35) Dyson, R. M.; Kaderli, S.; Lawrance, G. A.; Maeder, M.; Zunderbuhler, A. D. (1997) Second order global analysis: the evaluation of series of spectrophotometric titrations for improved determination of equilibrium constants. *Analytica Chimica Acta*, **353**, 381-393.
- (36) Maeder, M.; Zuberbuhler, A. D. (1990) Nonlinear Least-Squares Fitting of Multivariate Absorption Data. *Analytical Chemistry*, **62**, 2220-2224.



Capítol 5

Aplicacions bioanalítiques

En aquest apartat, es presenten les aplicacions dels mètodes de resolució multivariant a l'estudi dels àcids nucleics que s'han dut a terme en aquesta Tesi Doctoral. Els treballs resultants s'han agrupat en tres blocs per tal de facilitar la seva organització.

En el primer bloc es presenten quatre treballs relacionats amb l'estudi dels canvis conformacionals dels àcids nucleics al variar condicions experimentals del medi com ara la temperatura, el pH o la força iònica.

- En el primer treball es presenta un revisió bibliogràfica de les darreres aplicacions de mètodes de resolució multivariant al seguiment de processos bioquímics i biofísics.
- En el segon treball es mostren els resultats obtinguts en l'estudi de les transicions conformacionals d'un oligonucleòtid cíclic. Així, es va investigar l'efecte de la temperatura, la força iònica o la concentració d'oligonucleòtid en aquestes transicions. Aquests experiments es van seguir mitjançant l'absorció molecular i el dicromisme circular a l'ultraviolat. A partir dels resultats obtinguts mitjançant l'anàlisi de les dades amb el mètode MCR-ALS es van determinar els paràmetres termodinàmics associats a aquests equilibris. Aquest treball es va dur a terme en col·laboració amb el Grup de Síntesi i Disseny d'Àcids Nucleics del Prof. Enrique Pedroso del Departament de Química Orgànica de la Universitat de Barcelona.
- En el tercer treball es mostren els resultats obtinguts en l'estudi dels equilibris de formació d'estructures triples en funció del pH i de la temperatura. A partir d'un oligonucleòtid base en forma de forqueta i de diferents cadenes complementàries es va investigar la formació d'estructures triples paral·leles i antiparal·leles. A partir dels resultats obtinguts mitjançant el mètode MCR-ALS es van determinar les constants de formació associades a aquests equilibris. Aquest estudi es va dur a terme mitjançant l'espectroscòpia d'absorció molecular a l'ultraviolat i es van confirmar els resultats obtinguts mitjançant electroforesi en gel.
- Finalment, en el quart treball es mostren els resultats obtinguts en l'estudi de les transicions conformacionals d'un oligonucleòtid que pot adoptar estructures quàdruples, i la competència entre aquestes estructures i la forma dúplex en

presència de la cadena complementària. Aquest treball presenta un interès addicional ja que s'ha fet servir la tecnologia de les alimares moleculars mitjançant la qual es marquen els extrems de la cadena de l'oligonucleòtid amb un fluoròfor i un amortidor de fluorescència (*quencher*) de forma que es poden seguir els processos utilitzant la tècnica de fluorescència molecular, a més de l'absorció molecular i el dicroisme circular. Aquests dos darrers treballs es van dur a terme en col·laboració amb el Grup de Química d'Àcids Nucleics del Prof. Ramon Eritja del Departament de Biologia Estructural de l'Institut de Biologia Molecular de Barcelona (C.S.I.C).

En el segon bloc es presenten dos treballs relacionats amb l'aplicació dels mètodes de resolució multivariant en l'anàlisi de dades de Ressonància Magnètica Nuclear.

- En el primer treball es va tractar el problema de l'ambigüitat dels senyals làbils de Ressonància Magnètica Nuclear quan s'estudien equilibris on l'intercanvi entre dos o més components és més ràpid que el temps d'adquisició de l'espectre. Aquest fet provoca que únicament es pugui observar un senyal mitjà dels senyals de les espècies present en equilibri, de forma que per a senyals làbils no es poden aplicar mètodes basats en la descomposició bilineal de les dades com, per exemple, el mètode MCR-ALS. En aquest treball es proposa un mètode que permet la transformació d'aquests senyals RMN làbils en senyals RMN inerts bilineals, i el seu anàlisi posterior amb mètodes de resolució multivariant.
- En el segon treball es van analitzar dades de RMN bidimensional que havien estat enregistrades prèviament [1] pel Grup de Síntesi i Disseny d'Àcids Nucleics de la Prof. Anna Grandas del Departament de Química Orgànica de la Universitat de Barcelona. En aquest treball es va estudiar la cinètica de la reacció entre el cisplatí i un oligopèptid. El principal objectiu del treball va ser l'anàlisi d'aquestes dades multidimensionals ja que, al tractar-se d'una reacció lenta, no es van trobar problemes d'ambigüitats de senyal làbil com en el cas anterior.

Finalment, en el tercer bloc es presenta un treball en el qual s'analitzen dades de micromatrius d'ADN. En aquest treball es comprova la capacitat que té el mètode MCR-ALS per analitzar aquest tipus de dades. Així, s'ha analitzat en primer lloc uns conjunts de dades simulades i, posteriorment, dos conjunts de dades experimentals

que es troben disponibles públicament i que es consideren dades model en el desenvolupament de mètodes per a l'anàlisi d'aquest tipus de dades.

Referències

- (1) Marchan, V.; Moreno, V.; Pedroso, E.; Grandas, A. (2001) Towards a better understanding of the cisplatin mode of action. *Chemistry-a European Journal*, **7**, 808-815.

5.1. “Application of multivariate resolution methods to the study of biochemical and biophysical processes ”

Analytical Biochemistry,
(2004),
327, 1-13

Joaquim Jaumot
Montse Vives
Raimundo Gargallo

5.2. “Multivariate curve resolution: a powerful tool for the analysis of conformational transitions in nucleic acids ”

Nucleic Acids Research,
(2002),
30, e92

Joaquim Jaumot
Núria Escaja
Raimundo Gargallo
Carlos González
Enrique Pedroso
Romà Tauler

5.3. “Resolution of parallel and antiparallel oligonucleotide triple helices formation and melting processes by multivariate curve resolution ”

*Journal of Biomolecular
Structure & Dynamics,*
(2003),
21, 267-278

*Joaquim Jaumot
Anna Aviño
Ramon Eritja
Romà Tauler
Raimundo Gargallo*

5.4. “Resolution of a structural competition involving dimeric G-quadruplex and its C-rich complementary strand ”

Nucleic Acids Research,
(2006),
34, 206-216

Joaquim Jaumot
Ramon Eritja
Romà Tauler
Raimundo Gargallo

5.5. “Multivariate resolution of NMR labile signal by means of hard- and soft-modelling methods ”

Analytica Chimica Acta,
(2003),
490, 253-264

Joaquim Jaumot
Montse Vives
Raimundo Gargallo
Romà Tauler

5.6. “Multivariate curve resolution applied to the analysis and resolution of two-dimensional [^1H , ^{15}N] NMR reaction spectra ”

Analytical Chemistry,
(2004),
76, 7094-7101

Joaquim Jaumot
Vicente Marchán
Raimundo Gargallo
Anna Grandas
Romà Tauler

5.7. “Exploratory data analysis of DNA microarrays by Multivariate Curve Resolution”

(enviat per a la seva publicació)

*Joaquim Jaumot
Romà Tauler
Raimundo Gargallo*

5.8. Discussió dels resultats

En aquest capítol es presenta una discussió global dels treballs com a complement de la que es pot trobar en els articles. La discussió s'ha dividit en dos subapartats seguint l'esquema inicial del capítol. En el primer apartat es realitza la discussió en termes quimiomètrics dels avantatges de l'aplicació de mètodes de resolució multivariant a aquests tipus de dades. En el segon apartat es duu a terme la discussió des d'un punt de vista biofísic.

5.8.1. Discussió dels resultats des d'un punt de vista quimiomètric

En primer lloc es va realitzar un estudi sobre l'aplicació de mètodes de resolució multivariant a l'anàlisi de dades espectroscòpiques obtingudes en el seguiment de processos bioquímics i biofísics. Amb aquesta finalitat es va fer una recerca bibliogràfica per conèixer els sistemes biològics més estudiats, i les tècniques instrumentals i quimiomètriques més utilitzades. La informació recollida i ordenada sistemàticament va ser publicada en forma d'article de revisió (apartat 5.1.).

Així, es va constatar que els sistemes bioquímics més estudiats són les proteïnes entre les quals l'hemoglobina, la mioglobina i el citocrom c són les més àmpliament citades. En canvi, es van trobar menys estudis sobre l'estudi de les transicions conformacionals dels àcids nucleics. A més, es comprova que els processos cinètics han estat els més estudiats amb els mètodes de resolució multivariant. Per acabar, cal destacar la gran varietat de tècniques instrumentals emprades en els estudis dels sistemes bioquímics.

Com a conclusió d'aquest treball cal destacar que el nombre de publicacions sobre la utilització dels mètodes de resolució multivariant està creixent de forma molt ràpida. Tradicionalment, en aquest camp s'han fet servir majoritàriament els mètodes quimiomètrics de modelització rígida ja que els seus resultats es consideraven més fiables. Malgrat això la complexitat creixent de les dades a analitzar i, per tant, dels models a ajustar està donant lloc a un ràpid increment en l'ús de mètodes quimiomètrics de modelització flexible. Això és degut a que l'anàlisi simultània de més d'un experiment i l'aplicació de restriccions durant l'optimització per mínims quadrats alternats proporciona solucions del mateix grau de fiabilitat que els mètodes de

modelització rígida però sense la necessitat d'aquests últims de postular un model químic determinat. Així, per exemple, en els treballs que recull aquesta memòria s'han fet servir ambdues aproximacions pel tractament de diferents tipus de problemes amb resultats comparables. Els requeriments dels mètodes basats en models flexibles són menors i més fàcils d'assolir en l'estudi dels sistemes químics i bioquímics en condicions generals.

A continuació, es discutirà l'aplicació d'aquests diferents tipus de modelitzacions depenent del tipus de dades que s'estigui analitzant.

I. Resolució de processos

Mètodes de modelització rígida

Els mètodes de modelització rígida només s'han aplicat en els casos on es tenia un coneixement previ del sistema i, a més, el model a ajustar no presentava una gran complexitat. Per exemple, s'ha utilitzat el mètode de modelització rígida EQUISPEC [1] per a l'anàlisi de dades obtingudes al llarg de valoracions àcid-base en dos casos ben diferenciats. En el primer cas (treball de l'apartat 5.5.), es van estudiar els equilibris àcid-base de dues molècules petites com són els desoxinucleòtids dCMP i dGMP mitjançant Ressonància Magnètica Nuclear. Els dos sistemes havien estat descrits prèviament a la literatura [2, 3] i es tenia un coneixement del model a ajustar, consistent en tres protonacions en l'interval de pH estudiat. En el segon cas (apartat 5.3.), es van analitzar els equilibris de protonació de l'oligonucleòtid h26 que presenta una estructura de forqueta i tres oligonucleòtids (s11AG, s11CT i s11GT) de cadena senzilla. L'anàlisi de les dades de la valoració d'aquests oligonucleòtids individualment es va estudiar també mitjançant el mètode EQUISPEC [1]. En tots tres oligonucleòtids només es va observar una protonació. A més, al tractar-se d'oligonucleòtids d'un nombre petit de bases, l'absència d'efectes polifuncionals i polielectrolítics permet postular models senzills basats en el compliment de la llei d'acció de masses [4].

En canvi, en el cas d'analitzar experiments de valoració àcid-base més complexos o en el cas de tractar dades procedents de desnaturalitzacions tèrmiques, es va optar per utilitzar mètodes de modelització flexible i, en concret, el mètode de Resolució Multivariant de Corbes per Mínims Quadrats Alternats (MCR-ALS).

Mètodes de modelització flexible

El mètode MCR-ALS s'ha fet servir en la majoria dels estudis presentats en aquesta memòria, tant per a l'estudi de transicions conformacionals (apartats 5.2., 5.3., 5.4., 5.5. i 5.6.) com per a l'anàlisi de dades de micromatrius d'ADN (apartat 5.7.).

En els casos on es segueixen transicions conformacionals, l'aplicació del mètode MCR-ALS ha permès donar una possible interpretació biofísica d'aquestes transicions. Tradicionalment, l'estudi d'aquestes transicions es duia a terme mitjançant mètodes d'anàlisi univariant els quals presenten força inconvenients si es volen analitzar sistemes complexos [5]. Un exemple d'això és el treball on s'han estudiat els equilibris entre les estructures *dumbbell* i *bi-loop* de l'oligonucleòtid cíclic d<pTGCTCGCT> (veure treball de l'apartat 5.2.). En la majoria dels casos s'ha fet servir el coneixement que es tenia del sistema en la forma de restriccions durant l'optimització ALS. El factor més important, però, ha estat l'ús de l'estratègia d'anàlisi del sistema sota condicions diverses, ja sigui mitjançant l'anàlisi simultani de mesures obtingudes amb més d'una tècnica espectroscòpica, o mitjançant l'anàlisi simultani d'experiments obtinguts en condicions diferents. A continuació es comentaran els factors que donen una major fiabilitat als resultats obtinguts pel mètode MCR-ALS:

A) Flexibilitat en la imposició de restriccions

A l'apartat 1.2.2. s'han mostrat les restriccions que es poden imposar durant el procés d'optimització ALS per dotar a les solucions un significat químic més fàcilment interpretable que la solució matemàtica obtinguda, per exemple, per PCA. A continuació es discuteix l'aplicació d'aquestes restriccions:

- **No-negativitat.** Aquesta és una de les restriccions que més s'ha fet servir ja que en la majoria dels casos es complia en les dades experimentals. Així, pels perfils de concentració aquesta restricció s'ha imposat en tots els casos ja que no té sentit químic permetre concentracions amb valors negatius. En canvi, en el cas de perfils espectrals, aquesta restricció s'ha aplicat només on es complia, per exemple en l'absorció molecular a l'ultraviolat o a la fluorescència. En canvi, en el cas de l'anàlisi de dades de dicromisme circular no s'ha imposat aquesta restricció, ja que aquest senyal tant pot ser positiu com negatiu.

- **Unimodalitat.** Aquesta restricció només s'ha fet servir en alguns casos pels perfils de concentració, als quals obliga a presentar una única banda o màxim. La imposició d'aquesta restricció als perfils espectrals no és adequada en els casos aquí estudiats ja que els senyals espectroscòpics presentaven més d'una banda o màxim local. La imposició d'aquesta restricció als perfils de concentració requereix d'una certa habilitat. El mètode MCR-ALS permet controlar la flexibilitat amb la qual s'imposa aquesta restricció. En el cas d'aplicar aquesta restricció amb poca flexibilitat el procés d'optimització pot divergir, mentre que si s'aplica amb més flexibilitat (es permet una certa tolerància respecte en l'aparició de petits màxims locals) el procés d'optimització no es veu afectat negativament.
- **Sistema tancat.** Aquesta restricció també s'ha fet servir ja que es compleix en molts sistemes químics. Aquesta restricció obliga a que la suma de les concentracions de les espècies escollides sigui, com a màxim, igual a un determinat valor que pot ser la concentració total analítica (en el cas que sigui coneguda) o al 100% de la concentració relativa. Malgrat això, s'ha d'anar amb compte amb els sistemes que no compleixin estrictament aquesta restricció. Per exemple, en els experiment de valoració àcid-base els volums afegits han de ser menyspreables respecte al volum inicial de solució, per tal que es compleixi aquesta restricció. Si aquest no és el cas, aleshores el mètode MCR-ALS permet la imposició d'una restricció de sistema tancat variable si es coneixen els canvis en la concentració total analítica durant el procés. Finalment, si no es pot aplicar cap d'aquestes dues opcions és recomanable imposar durant l'optimització una normalització d'alçada o d'àrea (generalment als perfils espectrals) per tal de controlar els efectes de l'ambigüitat d'intensitat.
- **Selectivitat.** La restricció de selectivitat, tant en els perfils de concentració com en els perfils espectrals, es pot considerar com la restricció més important per tal d'eliminar els efectes de l'ambigüitat rotacional. Aquesta restricció només es pot imposar quan es té un coneixement previ del sistema i es sap amb certesa si una espècie és present o no en unes determinades condicions experimentals, o si una determinada espècie presenta o no senyal espectroscòpic a una determinada longitud d'ona. Així, en experiments on es coneix que una espècie no es troba present en l'inici del procés, no es permet que el seu perfil de concentració sigui

diferent de zero en aquesta regió. En el cas de que el sistema presenti selectivitat espectral aquesta informació també pot ser utilitzada. Per exemple, si es coneix que una de les espècies no presenten senyal d'absorció a una determinada longitud d'ona, es pot aplicar aquesta restricció per eliminar les interferències espectrals d'aquesta espècie a aquesta longitud d'ona.

La restricció de trilinearitat no s'ha fet servir ja que no s'ha treballat amb dades experimentals que tinguessin aquest tipus d'estructura.

B) Estratègia d'augmentació de matrius

A l'apartat 1.2.2. s'ha presentat l'estratègia de dur a terme l'optimització per mínims quadrats alternats fent servir dades espectroscòpiques obtingudes mitjançant diverses tècniques diferents, sistemes estudiats en diferents condicions experimentals o la combinació d'aquestes estratègies.

Estratègia d'anàlisi simultània de conjunts de dades obtingudes mitjançant diferents tècniques espectroscòpiques

L'estratègia d'analitzar conjuntament les dades obtingudes mitjançant múltiples tècniques espectroscòpiques diferents ha estat especialment útil en els casos que la informació proporcionada per a cadascuna de les tècniques era complementària. Un exemple d'això es descriu al treball 5.2. per a l'experiment de desnaturalització tèrmica de l'oligonucleòtid d<pTGCTCGCT> en medi aquós i força iònica pràcticament zero, on la suma de la informació proporcionada per les espectroscòpies d'absorció molecular a l'ultraviolat i de dicroisme circular va permetre la millor resolució del sistema. Aquesta estratègia ajuda a disminuir l'efecte de l'ambigüitat rotacional en augmentar la determinació del sistema i, a més, també és útil per eliminar problemes com la deficiència de rang. Aquest darrer problema es pot resoldre si s'ajunten una matriu deficient de rang i una matriu que presenta rang complet. Per exemple, es pot analitzar una matriu de dades obtingudes mitjançant fluorescència molecular (amb $n-1$ components) i una matriu de dades de dicroisme circular (amb n components). D'aquesta forma, el rang de la matriu ampliada serà n . Aquest és el cas, per exemple, de les dades de fluorescència molecular i de dicroisme circular en el treball 5.4., on

s'analitzen les diferents conformacions adoptades per l'oligonucleòtid d(TAG GGT TAG GGT).

Estratègia d'anàlisi simultània de conjunts de dades obtingudes en experiments diferents amb espècies comunes

Una altra opció per augmentar la fiabilitat dels resultats obtinguts pel mètode MCR-ALS i, en alguns casos, la millor opció per obtenir resultats amb significat químic és utilitzar l'estratègia d'analitzar simultàniament dades obtingudes en diferents condicions experimentals per a un mateix sistema.

Seguint amb l'exemple del treball 5.2., en el cas de tractar la desnaturalització tèrmica de l'oligonucleòtid d<pTGCTCGCT> en medi salí no s'obtenien unes solucions que expliquessin totalment els fets experimentals. Així, únicament s'observaven dues espècies de les tres que es podrien esperar tenint en compte els fets experimentals. Això feia indicar que en la resolució d'aquest experiment es presentava el problema de la deficiència de rang. Com a conseqüència d'això es van realitzar experiments alternatius per tal d'aconseguir eliminar aquesta deficiència de rang. Així, es van analitzar simultàniament els experiments de desnaturalització tèrmica en medi d'aigua pura, en medi salí i una valoració on es canviava la força iònica del medi. De la mateixa forma, aquesta estratègia d'augmentació de matrius va estar utilitzada amb èxit en els treballs on s'estudiaven els equilibris àcid-base d'estructures triples (apartat 5.3.).

L'estratègia d'augmentació de matrius analitzant simultàniament diferents experiments permet, a més, resoldre aquelles contribucions espectroscòpiques d'espècies que es poden considerar interferents. Un clar exemple d'aquest fet és l'eliminació de la contribució del blanc (treball 5.2.) En aquest cas, l'amortidor de pH utilitzat, PIPES, presentava un senyal d'absorció molecular a l'ultraviolat no menyspreable respecte al senyal espectroscòpic de l'oligonucleòtid i, a més, el senyal variava amb la temperatura. Per tal de resoldre aquest problema es va dur a terme un experiment mesurant únicament el senyal del blanc de l'amortidor PIPES a cada temperatura. Finalment, es va dur a terme la resolució global del sistema incloent els experiments amb presència de l'oligonucleòtid i sense, per tal de modelitzar les contribucions de l'oligonucleòtid i de l'amortidor.

Per concloure, aquesta estratègia permet l'eliminació de problemes de deficiència de rang en el cas de tenir un disseny d'experiments apropiat i, tal com s'ha demostrat a l'apartat de validació del mètode MCR-ALS, pot servir per eliminar gairebé totalment l'efecte de les ambigüitats rotacionals.

Estratègia d'anàlisi simultània de conjunts de dades obtingudes en diferents experiments i mitjançant tècniques espectroscòpiques diferents

A més, es poden analitzar simultàniament dades corresponents a més d'un experiment obtingudes mitjançant més d'una tècnica instrumental. D'aquesta manera, es pot augmentar la matriu de dades tant en la direcció de les files (diferents tècniques instrumentals) com de les columnes (diferents experiments estudiats). Aquesta estratègia combina els avantatges de les dues estratègies descrites prèviament.

Aquesta serà la millor opció possible en els casos on es pot disposar de més d'una tècnica instrumental i es pot estudiar el sistema en diferents condicions experimentals (concentracions inicials, temperatura, pH, ...). Aquesta ha estat la opció escollida per a la resolució dels sistemes estudiats en els treballs 5.2. i 5.4. Per exemple, en aquest darrer treball es va mostrar que l'estratègia d'augmentar la matriu tant en la direcció de les files (tècniques instrumentals) com de les columnes (condicions experimentals) permetia la completa resolució del sistema.

La fiabilitat dels resultats obtinguts seguint aquesta estratègia d'augmentació de matrius de dades en les dues direccions també serà més gran que amb les estratègies anteriorment comentades on només s'augmentaven les matrius en una direcció (files o columnes). Això és degut a que es facilita molt més l'eliminació tant dels problemes de deficiència de rang com d'ambigüitats que podrien afectar la qualitat de la resolució.

II. Anàlisi de dades RMN làbils i multidimensionals

En el treball 5.5. s'ha proposat un algoritme de transformació matemàtica dels senyals de RMN làbils en senyals inerts. Com s'ha explicat a la descripció de la tècnica, la RMN presenta coalescència del senyal si la velocitat de l'equilibri estudiat és major que la freqüència de la radiació que es fa servir a l'experiment per excitar els nuclis.

L'algoritme de transformació proposat es basa en la descomposició de cada senyal de RMN en dues senyals, corresponents a les seves contribucions àcida i bàsica pura. Mitjançant aquest algoritme s'aconsegueix la linealització de les dades originals (no lineals) fins obtenir unes dades lineals que es poden descompondre seguint un model bilineal com el de la llei de Lambert-Beer generalitzada. Els resultats obtinguts indiquen que aquest mètode pot ser utilitzat amb uns bons resultats en el cas d'analitzar àcids polipròtics que tinguin una separació entre els valors de pK_a consecutius suficientment gran per tal que totes les espècies arribin a ser majoritàries en algun pH de la valoració. Com a punt negatiu s'ha de tenir en compte que si no es compleix aquesta condició, dur a terme les estimacions de les espècies pures per tal de descompondre les senyals intermèdies és més problemàtica i pot presentar un major grau d'ambigüitat a les solucions obtingudes.

Per una altra banda, en el treball presentat a l'apartat 5.6. es mostra l'aplicació del mètode MCR-ALS per a l'anàlisi de dades obtingudes mitjançant espectroscòpia $^1H,^{15}N$ -HSQC RMN 2D. Així, s'ha analitzat la cinètica de la interacció entre un oligopèptid i el compost anticancerígen cisplatí. En aquest treball s'han obtingut una sèrie de resultats que, des de el punt de vista quimiomètric, han de ser comentats.

En els últims anys han aparegut multitud de dades multidimensionals en les quals a cada punt discret t (temps, pH, ...) s'obté una matriu de dades completa de dimensions $i \times j$. Així, al considerar tot el procés es tindrà l'estructura d'un cub de dades de dimensions $i \times j \times t$. L'anàlisi MCR-ALS d'aquest tipus de dades tridimensionals fa necessari desdoblar aquest cub de dades i transformar-lo en una matriu augmentada. Això s'aconsegueix mitjançant el desdoblament de cadascuna de les t matrius espectrals obtingudes en un únic vector (veure Esquema 2 en el treball). Aquesta nova matriu augmentada es pot analitzar amb el mètode clàssic MCR-ALS.

En segon lloc, cal diferenciar entre els dos tipus d'anàlisi que es poden dur a terme al tractar dades RMN 2D. D'una banda es pot analitzar cada espectre de $^1H,^{15}N$ -RMN individualment. D'aquesta forma es pot descompondre un espectre de RMN 2D en els seus espectres de correlació 1D de 1H i de ^{15}N de forma que es poden assignar amb una major facilitat els senyals de RMN. D'altra banda, es pot analitzar la matriu augmentada en forma obtinguda pel desdoblament del cub de dades original. L'anàlisi

d'aquesta nova matriu augmentada permet conèixer l'evolució cinètica del procés, durant la reacció d'interacció entre el cisplatí i l'oligopèptid.

III. Anàlisi de dades de micromatrius d'ADN

En el treball presentat a l'apartat 5.7. es mostra l'aplicació del mètode MCR-ALS a dades de micromatrius d'ADN. En aquest treball s'han obtingut una sèrie de resultats que, des de el punt de vista quimiomètric, han de ser comentats.

En primer lloc, s'ha de comentar com són les dades de micromatrius d'ADN ja que presenten unes característiques totalment diferents a les altres dades d'origen espectroscòpic que s'han tractat en aquesta Tesi. Així, en els experiments de micromatrius d'ADN no es segueix un procés sinó que es mesura l'expressió de determinats gens en diferents mostres. A més, com s'ha descrit a l'apartat 2.4. de la Introducció, aquests experiments presenten múltiples fonts de variabilitat que són difícilment controlables (per exemple, cada experiment està format a la seva vegada per múltiples experiments independents). Per tot això, la quantitat de variància que es pot explicar per a cada component extret és, en general, molt petita. Així, per exemple, es poden necessitar fins a deu components per arribar a explicar una variància total al voltant del 50%. Aquest fet implica que la selecció adequada del nombre de components sigui difícil i que existeixi el perill de sobreajustar el model durant l'optimització.

En segon lloc, cal destacar la utilització dels anomenats centroides obtinguts pel mètode de classificació no jeràrquica *k-means* com a estimacions inicials pel procés d'optimització per mínims quadrats alternats. Aquest mètode ha permès tenir una estimació inicial dels perfils gènics per a un determinat nombre de grups i, aconseguir una separació de les mostres més significatives. L'aplicació dels mètodes utilitzats habitualment per a les estimacions inicials en el cas de la resolució de processos, com el mètode EFA o el de detecció de variables pures, no proporcionaven, en aquest cas, bons resultats degut a la natura no evolutiva de les dades i a problemes en la detecció de les variables més diferents en presència d'una contribució molt gran del soroll experimental.

Finalment, en el procés de resolució d'aquestes dades de micromatrius s'han aplicat dues restriccions. D'una banda la restricció de no-negativitat pel perfil de mostres ja que, per exemple, una mostra podrà estar associada, o no, a un determinat tipus de malaltia però no té sentit que presenti una contribució negativa. D'altra banda, s'ha implementat una nova restricció als perfils gènics resolts consistent en obligar que cadascun d'aquests perfils compleixi la restricció d'autoescalat, és a dir, que tingui una mitjana dels seus valors igual a zero i una desviació estàndard igual a la unitat. Aquestes restriccions ens permeten obtenir unes solucions amb més sentit biològic i disminuir l'efecte de les ambigüitats rotacionals i d'intensitat.

5.8.2. Discussió dels resultats des d'un punt de vista biofísic

Des d'un punt de vista biofísic el treball dut a terme en aquesta Tesi Doctoral es pot organitzar en dos apartats. En un primer apartat es mostra el treball relacionat amb l'anàlisi de processos evolutius en els quals es produeix un canvi conformacional dels àcids nucleics o bé la interacció d'aquests amb metalls presents en el medi. En un segon apartat es mostra el treball relacionat amb l'anàlisi de dades de micromatrius d'ADN a partir de les quals es pretén obtenir informació biomèdica. A continuació es presentarà una discussió detallada dels resultats obtinguts en aquests dos apartats.

I. Anàlisi de processos evolutius

Primer, s'analitzen els resultats obtinguts en l'estudi de processos evolutius. Dins d'aquest grup es descriuran els resultats en funció de la variable experimental: temperatura, pH, força iònica o temps.

Temperatura

La temperatura és la variable experimental més emprada per induir transicions conformacionals en els oligonucleòtids estudiats en aquesta Tesi. Així, tal i com s'ha explicat a l'apartat 2.2., la variació de la temperatura afavoreix la desorganització d'un àcid nucleic; és a dir, a temperatures baixes es troben estructures bàsicament ordenades mentre que a temperatures altes es troben estructures fonamentalment desordenades.

En primer lloc es discuteixen els resultats dels experiments de desnaturalització tèrmica obtinguts a l'apartat 5.2. La motivació inicial d'aquest treball va ser intentar explicar la corba de fusió anòmala que presentava l'oligonucleòtid d<pTGCTCGCT> en medi salí (10 mM de Mg^{2+}) on s'observa un descens inicial de l'absorbància i un posterior augment d'ella a partir de 60 °C. L'anàlisi simultània de les dades d'aquest experiment de desnaturalització juntament amb d'altres dades corresponents a experiments de desnaturalització en condicions diferents i amb dades d'experiments de canvi de la força iònica del medi va permetre explicar plausiblement aquest corba. Així, en medi salí i a temperatures baixes, es va trobar únicament l'estructura dimèrica *bi-loop* que, en incrementar la temperatura, desapareix donant lloc tant a la estructura monomèrica *dumbbell* com a la conformació desordenada.

L'anàlisi comparatiu dels valors obtinguts de la temperatura de desnaturalització (T_m) a diferents concentracions de l'oligonucleòtid va permetre obtenir informació sobre la molecularitat del sistema. Així, es va observar que la T_m en variar la concentració de l'oligonucleòtid (de 5 μM a 70 μM) en medi d'aigua pura es mantenia pràcticament constant (les T_m obtingudes es troben a l'interval 47,6 – 48,4 °C), la qual cosa indicava clarament que l'equilibri estudiat era unimolecular. En canvi, al variar la concentració de l'oligonucleòtid en el mateix interval de concentracions però en medi salí es va observar un augment continu de la T_m entre 51,2 °C a una concentració d'oligonucleòtid de 5 μM i 64,7 °C a una concentració d'oligonucleòtid de 70 μM . Aquest fet és característic d'un equilibri dimèric amb una forta dependència amb la concentració [6]. A partir dels perfils de concentració resolts es van obtenir els paràmetres termodinàmics relacionats amb aquests equilibris.

En el segon treball (apartat 5.3.), en el qual s'estudien els equilibris de formació d'estructures triples, es van dur a terme experiments de desnaturalització tèrmica per tal de comprovar l'estabilitat de les estructures formades. D'aquest treball hi ha diferents aspectes que han de ser destacats. En primer lloc, es pot observar l'existència d'una transició comuna a tots els experiments, que és la desnaturalització de l'oligonucleòtid amb estructura de forqueta (h26) que serveix com a base per a la formació de les estructures triples. Aquest oligonucleòtid presenta una T_m força elevada (71 °C) degut a la presència d'onze parells de bases tipus Watson-Crick (sis d'ells G·C), i a que l'estructura de forqueta proporciona una estabilitat addicional ja que

és una única cadena la que es doblega sobre sí mateixa. En segon lloc cal destacar l'estabilitat relativa de les diferents estructures triples formades observades en la primera desnaturalització. Així, l'estructura triple paral·lela formada pels oligonucleòtids h26 i s11CT presenta una T_m de 44 °C. Aquesta estructura es troba estabilitzada per la formació de tríades de bases tipus Hoogsteen $C^+(G\cdot C)$ i $T^*(A\cdot T)$. La necessitat de que les citosines de la cadena s11CT estiguin protonades té com a conseqüència que aquesta estructura triple només es formi a valors de pH lleugerament àcids (per exemple, només es va poder determinar el valor de la temperatura de desnaturalització entre $pH = 3,8$ i $pH = 6,0$). En canvi, les estructures triples antiparal·leles formades per la interacció de l'oligonucleòtid h26 i les cadenes s11AG i s11GT formen l'estructura triple en tot el rang de pH estudiat entre $pH = 3,9$ i $8,2$, ja que formen aquesta estructura per formació de tríades de bases tipus Hoogsteen invertit $G^*(G\cdot C)$, $A^*(A\cdot T)$ i $T^*(A\cdot T)$. És interessant destacar la diferència entre els valors de les T_m de les estructures triples formades. Així, mentre l'estructura triple formada pels oligonucleòtids h26 i s11AG té un valor de T_m de 44 °C (molt semblant a l'obtingut en el cas de l'estructura triple paral·lela), l'estructura formada pels oligonucleòtids h26 i s11GT té una T_m sensiblement més baixa ($T_m=28$ °C). Aquesta diferència en les T_m es pot explicar donat que la triada $G^*(G\cdot C)$ és més semblant a les triada $A^*(A\cdot T)$ que a la triada $T^*(A\cdot T)$.

Finalment, s'han analitzat les dades corresponents als experiments de desnaturalització tèrmica en els quals s'estudiaven les transicions conformacionals d'estructures quàdruples, i de la competició d'aquestes estructures quàdruples i de les estructures dobles que es poden formar si es troba la cadena complementària en solució (apartat 5.4.). En primer lloc cal considerar únicament l'oligonucleòtid SG (F – d(TAG GGT TAG GGT) – Q) que pot formar dues estructures quàdruples intermoleculares diferents: paral·lela i antiparal·lela. La resolució mitjançant el mètode MCR-ALS de les dades de l'experiment de desnaturalització tèrmica a pH neutre i en medi del catió potassi ha permès determinar que l'estructura quàdruple paral·lela és majoritària enfront de l'estructura antiparal·lela. A més, la comparació de les T_m obtingudes indica que l'estructura paral·lela és més estable ja que presenta una T_m aproximadament 20°C més gran que la T_m de l'estructura antiparal·lela ($T_{m,paral\cdot lela} = 65$ °C i $T_{m,antiparal\cdot lela} = 46$ °C). A partir dels perfils de concentració obtinguts es van determinar alguns paràmetres termodinàmics, com ara la constant d'equilibri de les estructures paral·lela i antiparal·lela i ΔG° , ΔH° i ΔS° associats.

En segon lloc, cal comentar els experiments de desnaturalització on es troben els oligonucleòtids SG i la seva cadena complementària SC. En el cas que els dos oligonucleòtids es trobin en igual concentració es pot observar que l'espècie dúplex és majoritària. Les estructures quàdruples de l'oligonucleòtid SG són minoritàries al llarg de tot l'experiment. En cas que l'oligonucleòtid SG es trobi en excés, s'observa la competició entre les dues estructures de forma que es té en solució l'estructura dúplex SG-SC i les estructures quàdruples paral·leles i antiparal·leles en una proporció similar a la descrita al primer experiment.

pH

Com s'ha citat a l'apartat 2.2., el pH és una variable experimental de gran interès per l'efecte que pot tenir sobre les conformacions adoptades pels àcids nucleics.

En el primer treball (apartat 5.5.), es van estudiar els equilibris àcid-base dels desoxinucleòtids dCMP i dGMP mitjançant RMN 1D. En el cas del desoxinucleòtid dCMP es van observar tres valors de pK_a corresponents a les desprotonacions d'un protó del grup fosfat ($pK_{a1} = 0,8$), d'un protó del nitrogen 3 de la base nitrogenada ($pK_{a2} = 4,2$) i del segon protó del grup fosfat ($pK_{a3} = 6,7$) [3]. En el cas del desoxinucleòtid dGMP el valor de pK_{a1} (3,0) correspon a la desprotonació del nitrogen N7 de la guanina, el valor de pK_{a2} (6,0) a la desprotonació del segon protó del grup fosfat i el valor de pK_{a3} (9,9) correspon al nitrogen N1 de la guanina [2].

En el segon treball (apartat 5.3.) es van dur a terme valoracions àcid-base dels oligonucleòtids independentment i, posteriorment, amb dos oligonucleòtids en solució (l'oligonucleòtid h26 i un oligonucleòtid de cadena senzilla). Aquests dos tipus d'experiments diferents permeten obtenir informació relativa als equilibris àcid-base dels oligonucleòtids i a la dependència amb el pH de la formació d'estructures triples. Així, en primer lloc, es van determinar els valors de pK_a dels oligonucleòtids analitzats a partir de les dades obtingudes de les valoracions dels quatre oligonucleòtids independentment. Dels resultats obtinguts cal destacar la similitud entre els resultats entre el desoxinucleòtid dCMP (descrit al treball anterior) i l'oligonucleòtid s11CT ($pK_{a1} = 4,3$). Aquesta semblança és força lògica si es té en compte que, al dur a terme la valoració de l'oligonucleòtid, s'està valorant cadascuna de les bases que el formen i

que, en aquest cas, en aquest rang de pH, només es protonen les citosines en el seu nitrogen N3. A més s'ha de destacar també que, al ser un oligonucleòtid de només 11 bases nitrogenades, els efectes secundaris (efectes polifuncionals i polielectrolítics) són menyspreables. En el cas de l'oligonucleòtid s11GT, el valor de pK_a obtingut és d'aproximadament 3,7 la qual cosa indica que s'està observant la protonació/desprotonació del nitrogen N7 de la guanina. En el cas de l'oligonucleòtid s11AG es produeix simultàniament la protonació/desprotonació de les adenines i de les guanines per la qual cosa el valor de pK_a determinat de 3,7 correspon a un valor mitjana dels equilibris corresponents a les dues bases nitrogenades. Finalment, en el cas de l'oligonucleòtid h26 també es va determinar un únic valor de pK_a que correspon a la protonació/desprotonació global de les citosines, les guanines i les adenines al voltant de 3,9.

En segon lloc, una vegada es van tenir caracteritzats aquests sistemes individualment es va estudiar la influència del pH en la formació d'estructures triples. Així, en el cas de la interacció entre els oligonucleòtids h26 i s11CT, es va observar una clara dependència de la formació de l'estructura triple amb el pH, ja que les tríades de bases tipus Hoogsteen $C^+(G \cdot C)$ necessiten que la citosina de la cadena senzilla es trobi protonada per poder formar l'estructura triple paral·lela. A causa d'aquest fet, les estructures triples paral·leles només apareixen en un petit interval de pH entre 3,5 i 6,0. En canvi, en el cas de la interacció entre els oligonucleòtids h26 i s11AG o s11GT no s'observa aquesta dependència per la qual cosa es poden formar les estructures triples antiparal·leles a pH neutres. Aquesta possibilitat de formar estructures triples a pH neutre (com per exemple, el pH fisiològic) té una gran importància degut a les possibles implicacions que pot tenir el disseny d'oligonucleòtids en les teràpies antigèniques per tal que interaccionen amb seqüències d'ADN presents a les cèl·lules [7].

Força iònica

A l'apartat 2.2. s'ha comentat que la força iònica és una de les variables que més afecta a la conformació que adopta un oligonucleòtid. En aquesta Tesi s'han presentat dos treballs en els quals es mostra l'efecte de la força iònica en les conformacions adoptades per l'oligonucleòtid.

En primer lloc, anteriorment s'han descrit les transicions conformacionals de l'oligonucleòtid d(pTGCTCGCT). En medi d'aigua pura aquest oligonucleòtid adopta una estructura de tipus *dumbbell*, mentre que en medi salí (10 mM Mg^{2+} i 100 mM Na^+) adopta una estructura dimèrica coneguda com a *bi-loop* [8]. Per tal d'estudiar les transicions conformacionals es van dur a terme dos tipus d'experiments. D'una banda, experiments de desnaturalització tèrmica en condicions salines variables, que ja s'han presentat anteriorment. D'altra banda, experiments en els quals es va variar la concentració salina de forma similar a una valoració. Aquest segon tipus d'experiments es va dur a terme a diferents temperatures. A temperatura baixa (21°C) es va observar una transició entre l'estructura *dumbbell* present en el medi d'aigua pura i l'estructura *bi-loop* present en el medi salí. Aquesta transició es produeix a concentracions salines baixes (aproximadament 3 mM Mg^{2+}) i, a partir de 5 mM de Mg^{2+} , ja únicament és present l'estructura *bi-loop*. Els resultats obtinguts a temperatura elevada (54 °C) presenten un comportament similar, malgrat que en aquest cas es té com a espècie predominant l'estructura monomèrica desestructurada. Aquest comportament s'explica per l'efecte d'apantallament que exerceixen els cations presents en el medi salí i que disminueixen l'efecte de les repulsions entre els grups fosfat presents a la cadena de l'oligonucleòtid. A més, com indiquen els resultats, aquest efecte és molt més important per a cations divalents com el Mg^{2+} que per a cations monovalents com el Na^+ . Així, concentracions baixes de Mg^{2+} permeten la formació de l'estructura *bi-loop* mentre que es necessiten concentracions molt més grans de ions Na^+ per produir el mateix efecte apantallant. Aquest fet està d'acord amb la relació càrrega/radi d'aquest dos ions, que és molt més gran en el cas del catió Mg^{2+} , divalent i petit, que pel Na^+ i el K^+ , que són cations monovalents i amb una radi iònic igual o més gran.

En segon lloc es van estudiar els efectes produïts per la natura del catió en l'estabilitat de les estructures quàdruples. Això es va realitzar mitjançant l'estudi de la desnaturalització d'aquestes estructures en presència de K^+ i Na^+ . A la bibliografia es troba que les estructures quàdruples es troben estabilitzades per cations monovalents i amb una lleugera preferència del K^+ respecte del Na^+ . Els experiments presentats al treball 5.4., estudien els equilibris d'estructures quàdruples intermoleculares en presència del ió potassi. Aquests experiments es van repetir en presència del ió sodi i es van obtenir unes estabilitats molt més baixes (aproximadament 20 °C menys en els valors de T_m). Aquests resultats es poden explicar degut a que els espais o cavitats que queden lliures entre les quatre bases de guanina que formen el G-tètrada són

d'unes dimensions aproximadament iguals al radi del ió potassi (1,3 Å). Aquest radi és similar al del ió divalent Sr^{2+} però és molt més gran que els altres ions utilitzats en aquest tipus d'estudis, com el Na^+ o el Mg^{2+} .

Temps

El treball 5.6. descriu el seguiment de la reacció entre un oligopèptid (híbrid entre nucleòtid i pèptid) i el compost amb propietats anticancerígenes cisplatí [9]. Aquest treball ja s'ha comentat extensivament a l'apartat anterior com una aplicació nova del mètode MCR-ALS per a l'anàlisi de dades obtingudes al llarg d'un procés seguit per una tècnica multidimensional, en aquest cas, la RMN bidimensional.

A partir dels espectres RMN resolts, que proporcionen una gran quantitat d'informació estructural i dels seus corresponents perfils de concentració, es pot fer una interpretació química de la reacció estudiada. Així, s'observa que, durant la primera hora de la reacció, es formen els adductes monofuncionals per l'enllaç entre el platí i el sulfur de la metionina i el nitrogen (N7) de la guanina. A partir dels perfils de concentració i de les constants de velocitat determinades, es pot deduir que existeix una lleugera preferència per l'adducte monofuncional Pt-S, ja que aquest compost es forma en una concentració, més elevada, que es pot explicar per la major afinitat que presenten els àtoms de Pt i S enfront dels àtoms de Pt i N. A temps de reacció més grans, a partir dels adductes monofuncionals, Pt-S i Pt-N, es forma el quelat que presenta enllaços del platí tant amb el sofre com amb el nitrogen de l'oligopèptid, donant lloc a una estructura tancada. Aquest producte és majoritari durant gran part de la reacció ja que presenta una estabilitat termodinàmica molt gran, probablement degut a l'estabilitat pròpia dels compostos quelats. Malgrat això, a temps de reacció encara més grans (a partir de cinc hores) aquest producte quelat es descompon i apareix l'anomenat producte final en el qual es perd el grup ammina en *trans* respecte l'àtom de sofre. Aquesta pèrdua del lligand no es pot considerar com inesperada ja que es coneix que l'àtom de sofre presenta un efecte *trans* molt elevat. A la literatura es van trobar altres exemples d'aquesta pèrdua del lligand en la posició *trans* a l'àtom de sofre però, encara, no es sap amb certesa quin grup substitueix aquest lligand malgrat que sembla que podria ser substituït per un lligand aquo.

II. Anàlisi de dades de micromatrius d'ADN

El treball de l'apartat 5.7. presenta una nova aplicació del mètode MCR-ALS que permet extreure informació biomèdica a partir dels perfils de mostres i gènics resolts quan s'analitzen dades de micromatrius d'ADN. En general, a partir dels perfils de mostres resolts, es pot obtenir informació sobre el tipus de malaltia que presenta una mostra d'un determinat teixit mentre que, a partir dels perfils gènics, es pot aconseguir una selecció dels gens candidats relacionats amb un determinat tipus de càncer, ja sigui per trobar-se sobreexpressats o infraexpressats.

En primer lloc es va analitzar el conjunt de dades generat per Golub *et al.* [10], format per 72 mostres de dos diferents tipus de leucèmies per a les quals es va mesurar l'expressió de 7120 gens. L'anàlisi mitjançant el mètode MCR-ALS d'aquest conjunt de dades ha permès la separació de les mostres en els dos tipus de leucèmies (Leucèmia linfofocítica aguda, LLA, i Leucèmia mielocítica aguda, LMA). Es va intentar la classificació de les mostres en un nombre més gran de components (per exemple, en el cas de la leucèmia de tipus LLA es coneix que es pot classificar en tipus B i T segons els limfòcits que es trobin en excés), però no es van obtenir uns resultats suficientment satisfactoris. A la Figura 1 del treball es mostra la separació aconseguida per a les mostres de leucèmia LLA i LMA. A més, s'han determinat un conjunt de gens, entre els quals hi podria haver alguns relacionats amb l'aparició d'aquests tipus de leucèmia. S'ha de destacar que alguns dels gens inclosos en els perfils gènics, com a sobreexpressats o infraexpressats, ja apareixien prèviament a la literatura biomèdica com a possibles causants de l'aparició d'aquests tipus de leucèmies [10-13]. Entre aquests gens cal destacar el gen conegut com *zyxin* àmpliament descrit com a causant de l'aparició de la LMA. Altres gens relacionats amb la leucèmia tipus LMA i seleccionats pel mètode MCR-ALS proposat són el descrits com *Fumarylacetoacetate* o el *CD33 antigen*. En el cas de la leucèmia tipus LLA cal destacar els gens descrits com *MB-1*, el *Cyclin D3* o el *Myosin light Chain*.

En segon lloc es va analitzar el conjunt de dades conegut com NCI60 generat per Ross *et al.* [14], format per 60 mostres de 9 tipus diferents de càncer (càncer de pit, tumor al sistema nerviós central, càncer de còlon, càncer de pulmó, leucèmia, melanoma, càncer d'ovari, càncer de pròstata i càncer de ronyó) per a les quals s'ha mesurat l'expressió de 1416 gens. En aquest cas es va considerar òptima la solució obtinguda amb MCR-ALS considerant 7 components, tot i que no es va poder donar

una interpretació biològica per a tots ells. No obstant, amb aquests set components resultats, es va poder distingir clarament en alguns components les mostres corresponents als càncers de melanoma, leucèmia, còlon i del sistema central nerviós. Per a aquests components es van poder determinar els perfils gènics associats de forma que per a aquests tipus de càncer també s'han pogut obtenir gens candidats que havien aparegut prèviament a la literatura [14-19]. Així, pel melanoma s'han seleccionat els gens identificats com *Erythropoietin receptor* o el *Transforming growth factor, beta receptor II* els quals es troben sobreexpressats i infraexpressats respectivament. En el cas del càncer al sistema nerviós central els gens *Collagen, type V, alpha Chr.2*, *THY-1 Membrane Glycoprotein Precursor Chr.11* o el *EST, Weakly similar to W09D10.2* es troben sobreexpressats. En el cas de la leucèmia s'ha determinat sobreexpressat un gen que es troba relacionat a la literatura biomèdica amb l'aparició de la leucèmia que és el *LCP1 Lymphocyte cytosolic protein 1 (L-plastin) Chr.13*. Per acabar en el cas del càncer de còlon s'ha trobat una sèrie de gens candidats sobreexpressats (*Cyclin-dependent kinase inhibitor 2A, H.sapiens E-MAP-115 mRNA Chr.6*) i infraexpressats (*MSN Moesin Chr.X.*, *GJA1 Cardiac gap junction protein Chr.X*, *Aldehyde reductase 1*).

Cal destacar doncs, que l'aplicació del mètode MCR-ALS ha permès la selecció d'aquests gens candidats, però únicament s'han pogut contrastar amb resultats prèviament existents a la literatura biomèdica anterior. Així, s'han determinat altres gens que es podrien considerar sobreexpressats o infraexpressats però que no es poden confirmar amb resultats existents prèviament a la literatura. Per tal de comprovar aquests gens s'haurien de dur a terme experiments específics per tal de comprovar la seva expressió, però aquest treball es troba fora de l'abast d'aquesta Tesi Doctoral.

5.9. Bibliografia

- (1) Dyson, R. M.; Kaderli, S.; Lawrance, G. A.; Maeder, M.; Zunderbuhler, A. D. (1997) Second order global analysis: the evaluation of series of spectrophotometric titrations for improved determination of equilibrium constants. *Analytica Chimica Acta*, **353**, 381-393.
- (2) Sigel, H.; Massoud, S. S.; Corfu, N. A. (1994) Comparison of the Extent of Macrochelate Formation in Complexes of Divalent Metal-Ions with Guanosine (Gmp(2-)), Inosine (Imp(2-)), and Adenosine 5'-Monophosphate (Amp(2-)) - the Crucial Role of N-7 Basicity in Metal Ion-Nucleic Base Recognition. *Journal of the American Chemical Society*, **116**, 2958-2971.
- (3) Song, B.; Feldmann, G.; Bastian, M.; Lippert, B.; Sigel, H. (1995) Acid-Base and Metal Ion-Binding Properties of 2'-Deoxycytidine 5'-Monophosphate (Dcmp(2-)) Alone and Coordinated to Cis-Diammine-Platinum(II) - Formation of Mixed-Metal Ion Nucleotide Complexes. *Inorganica Chimica Acta*, **235**, 99-109.
- (4) Saenger, W. (1988) *Principles of nucleic acid structure*. 2ona. ed. ed. Springer, New York, NY, USA.
- (5) Mergny, J. L.; Lacroix, L. (2003) Analysis of thermal melting curves. *Oligonucleotides*, **13**, 515-537.
- (6) Breslauer, K. J. (1995) Extracting thermodynamic data from equilibrium melting curves for oligonucleotide order-disorder transitions. *Energetics of Biological Macromolecules*, **259**, 221-242.
- (7) Orozco, E.; Gariglio, P. (1999) *Genética y biomedicina molecular*. 1era. ed. ed. Uteha: N oriega, Mèxic D.F., Mèxic.
- (8) Escaja, N.; Pedroso, E.; Rico, M.; Gonzalez, C. (2000) Dimeric solution structure of two cyclic octamers: Four-stranded DNA structures stabilized by A: T: A: T and G: C: G: C tetrads. *Journal of the American Chemical Society*, **122**, 12732-12742.
- (9) Marchan, V.; Moreno, V.; Pedroso, E.; Grandas, A. (2001) Towards a better understanding of the cisplatin mode of action. *Chemistry-a European Journal*, **7**, 808-815.
- (10) Golub, T. R.; Slonim, D. K.; Tamayo, P.; Huard, C.; Gaasenbeek, M.; Mesirov, J. P.; Coller, H.; Loh, M. L.; Downing, J. R.; Caligiuri, M. A.; Bloomfield, C. D.; Lander, E. S. (1999) Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science*, **286**, 531-537.
- (11) Antonov, A. V.; Tetko, I. V.; Mader, M. T.; Budczies, J.; Mewes, H. W. (2004) Optimization models for cancer classification: extracting gene interaction information from microarray expression data. *Bioinformatics*, **20**, 644-U145.

- (12) Thomas, J. G.; Olson, J. M.; Tapscott, S. J.; Zhao, L. P. (2001) An efficient and robust statistical modeling approach to discover differentially expressed genes using genomic expression profiles. *Genome Research*, **11**, 1227-1236.
- (13) Wang, Y.; Tetko, I. V.; Hall, M. A.; Frank, E.; Facius, A.; Mayer, K. F. X.; Mewes, H. W. (2005) Gene selection from microarray data for cancer classification - a machine learning approach. *Computational Biology and Chemistry*, **29**, 37-46.
- (14) Ross, D. T.; Scherf, U.; Eisen, M. B.; Perou, C. M.; Rees, C.; Spellman, P.; Iyer, V.; Jeffrey, S. S.; Van de Rijn, M.; Waltham, M.; Pergamenschikov, A.; Lee, J. C. E.; Lashkari, D.; Shalon, D.; Myers, T. G.; Weinstein, J. N.; Botstein, D.; Brown, P. O. (2000) Systematic variation in gene expression patterns in human cancer cell lines. *Nature Genetics*, **24**, 227-235.
- (15) Crescenzi, M.; Giuliani, A. (2001) The main biological determinants of tumor line taxonomy elucidated by a principal component analysis of microarray data. *Febs Letters*, **507**, 114-118.
- (16) Dang, H.; Le, T.; Segaran, T.; Levy, J. (2002) Integrating database information in microarray expression analyses: Application to melanoma cell lines profiled in the NCI60 data set. *J Biomol Tech*, **13**, 199-204.
- (17) Musumarra, G.; Barresi, V.; Condorelli, D. F.; Fortuna, C. G.; Scire, S. (2004) Potentialities of multivariate approaches in genome-based cancer research: identification of candidate genes for new diagnostics by PLS discriminant analysis. *Journal of Chemometrics*, **18**, 125-132.
- (18) Peterson, L. E. (2003) Partitioning large-sample micro array-based gene expression profiles using principal components analysis. *Computer Methods and Programs in Biomedicine*, **70**, 107-119.
- (19) Romualdi, C.; Campanaro, S.; Campagna, D.; Celegato, B.; Cannata, N.; Toppo, S.; Valle, G.; Lanfranchi, G. (2003) Pattern recognition in gene expression profiling using DNA array: a comparative study of different statistical methods applied to cancer classification. *Human Molecular Genetics*, **12**, 823-836.