

Nicola Marchetti, Maria Chiara
Pietrogrande, Francesco Dondi
Dipartimento di Chimica
Università di Ferrara
mcp@unife.it



DECODIFICA SEPARAZIONI MULTIDIMENSIONALI

Applicazione alla proteomica

Il presente lavoro descrive un metodo statistico in grado di estrarre utili informazioni da separazioni bi-dimensionali (2D) multicomponente.

Il metodo, validato con una procedura di tipo simulativo, è stato applicato con successo a mappe generate numericamente che simulano mappe 2D PAGE (elettroforesi bi-dimensionale su gel di poliacrilamide) di miscele complesse di proteine, il cuore della proteomica.

Uno dei più complessi obiettivi della moderna scienza delle separazioni consiste nell'analisi dei componenti (anche a livello di tracce) di miscele complesse, formate cioè da un elevato numero di componenti (oltre 100) con strutture chimiche diverse, presenti con un ampio intervallo di concentrazione (oltre 1:1.000): è il caso di campioni di interesse biologico, ambientale, industriale. Nonostante i notevoli sviluppi tecnologici, i convenzionali sistemi mono-dimensionali presentano notevoli limiti nell'isolare tutti i singoli componenti. Pertanto negli ultimi anni si stanno sviluppando e sempre più utilizzando tecniche multidimensionali dove si realizza l'accoppiamento diretto di 2 o più

sistemi basati su processi separativi diversi completamente indipendenti (1-3). Alcuni esempi sono: LC-GC (cromatografia liquida accoppiata a cromatografia gassosa), GCxGC (comprehensive two dimensional GC), accoppiamenti di sistemi di cromatografia liquida con diverse combinazioni, quali HPLC in fase inversa, cromatografia ionica, cromatografia di affinità, elettroforesi capillare. Le analisi con queste tecniche, che si rivelano estremamente potenti per la separazione di centinaia-migliaia di componenti, producono una quantità enorme di dati, la cui interpretazione richiede l'impiego di strumenti idonei per poterne ricavare le informazioni analitiche richieste (2). Pertanto si assiste oggi ad un interesse crescente verso

la Scienza delle Separazioni teorica, per sviluppare modelli per la comprensione dei meccanismi di separazione, e computazionale per produrre nuovi strumenti teorico-computazionali di ampia validità ed applicabilità per interpretare i dati prodotti dall'analisi. Questo può essere anche esteso al caso di tecniche accoppiate, dove il sistema separativo cromatografico è accoppiato on-line a rivelatori spettroscopici per ottenere informazioni strutturali sui composti separati, per esempio GC-MS, LC-DAD UV (assorbimento UV con array di diodi), LC-MS-MS (4-5).

Un esempio di miscela complessa è il proteoma di una cellula, di un tessuto o di un fluido biologico, il cui studio (proteomica) è

di estrema attualità (6-8). Il proteoma viene definito come la totalità delle proteine specifiche complementari al genoma: esso comprende tutte le proteine che sono espresse in una cellula in un certo istante, incluse le isoforme e le proteine modificate. Gli obiettivi principali della proteomica sono l'identificazione delle proteine che sono coinvolte in specifici processi cellulari (*cell map proteomics*), oppure che mostrano un profilo di espressione alterato, a seguito di cambiamenti nella condizione fisiologica (*expression proteomics*).

Particolare interesse è rivolto all'identificazione di proteine che esplicano un ruolo di regolazione, coinvolte in processi patologici e pertanto candidate come target per interventi terapeutici-farmacologici. Si stima che il proteoma umano contenga da 10^5 a 2×10^6 proteine, delle quali 10^4 - 2×10^4 sono espresse contemporaneamente in un dato momento. In tipici esperimenti di *expression proteomics* è importante separare il più possibile ogni singola proteina per poi trattarla con un enzima proteolitico, generalmente tripsina: tale prodotto può poi essere analizzato tramite MS o LC-MS o LC-MSⁿ. Se si considera che la tripsina produce mediamente circa 20 peptidi per ogni proteina, la miscela che ne risulta può contenere approssimativamente 2×10^5 - 5×10^5 componenti. Data la complessità dei campioni, è quindi evidente la necessità di utilizzare per l'analisi tecniche con elevata capacità e potere separativo. Attualmente, il metodo maggiormente utilizzato nella proteomica per la separazione di miscele proteiche complesse è l'elettroforesi bi-dimensionale su gel di poliacrilamide (2D-PAGE) (7-12); in aggiunta, tecnologie altamente sofisticate, come ad esempio la determinazione della massa di polipeptidi e l'identificazione della loro struttura primaria tramite spettrometria MALDI-TOF oppure nanoESI MS/MS e LC-MS/MS, hanno avuto notevole sviluppo per per-

mettere l'identificazione di prodotti proteici (6, 9). La tecnica 2D-PAGE, anche se presenta alcuni limiti nella separazione dei componenti, fornisce informazioni strutturali sulle proteine (punto isoelettrico, pI, e peso molecolare relativo, M_r) utili per la ricerca e l'identificazione in database.

Teoria

Sovrapposizione delle macchie in separazioni multicomponente

Ogni separazione bi-dimensionale può essere rappresentata da un grafico bi-dimensionale in cui ogni asse rappresenta il tempo di ritenzione in un sistema separativo. Su tale superficie ogni componente della miscela appare come una macchia, la cui posizione nel piano rappresenta la ritenzione nei 2 sistemi e la cui intensità rappresenta la quantità del componente. Le mappe 2D-PAGE vengono fisicamente ottenute su una lastra di gel di poliacrilamide e opportunamente digitalizzate da un sistema di scansione ottico; nei sistemi bi-dimensionali in colonna il segnale registrato dal rivelatore è trasformato da un opportuno software in una matrice di dati e riportato in un grafico bi-dimensionale (mappa) (1, 2). È noto che le separazioni bi-dimensionali di miscele altamente eterogenee sono caratterizzate da una disposi-

zione casuale delle macchie all'interno dello spazio separativo disponibile: i componenti si dispongono in modo diradato in alcune regioni della separazione, mentre si affollano in altre, cosicché non si possono separare completamente ed alcuni di essi si sovrappongono (13). Tale situazione, che è una conseguenza della complessità stessa della miscela, determina una diminuzione della qualità dell'informazione analitica contenuta nella separazione. Come conseguenza della sovrapposizione fra le macchie, il loro numero, p , rivelabili nella mappa è sempre minore del numero m di componenti presenti nella miscela (13-19): la perdita di informazione analitica è data dal grado di separazione

$$\gamma = p/m \quad (1)$$

proporzionale al grado di sovrapposizione presente nella mappa 2D (13, 14). Così solo alcune macchie risulteranno essere pure (singoletti, s), formate, cioè, da un solo componente per macchia; molte altre saranno il risultato della sovrapposizione fra 2 (doppietti, d), 3 (tripletti, t), 4 (quadrupletti, q) o più (n -pletti) singoli componenti. Quindi, il numero totale di macchie rivelabili, p , è dato da:

$$p = s + d + t + q + \dots \quad (2)$$

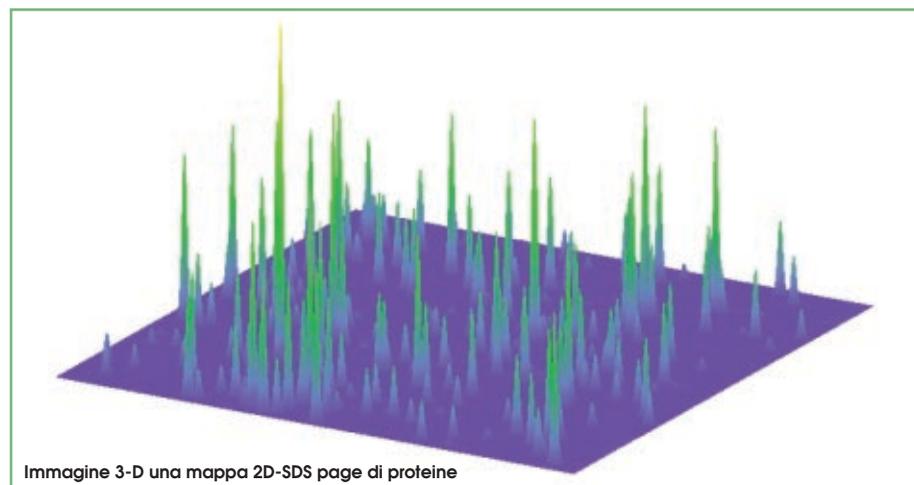


Immagine 3-D una mappa 2D-SDS page di proteine

mentre il numero totale di singoli componenti (SC), m , è:

$$m = s + 2d + 3t + 4q + \dots \quad (3)$$

Questo significa che per avere un'idea del numero reale di componenti presenti in un campione, è necessaria una stima quantitativa della sovrapposizione fra le macchie.

Proprietà delle separazioni multicomponente

Oltre al grado di separazione, γ , sono stati definiti altri attributi fondamentali per descrivere le separazioni lineari monodimensionali (1D) (13, 14): la capacità di picco, $n_{c,1D}$ ed il fattore di saturazione, α_{1D} . La capacità di picco, $n_{c,1D}$, esprime il potere separativo disponibile di una separazione: essa si riferisce all'estensione della separazione, X , e definisce il numero massimo di componenti che possono essere risolti nello spazio separativo:

$$n_{c,1D} = X/4\sigma R_s \quad (4)$$

dove R_s è il valore di risoluzione che si vuole considerare, definito come la distanza tra due picchi adiacenti, ΔX , rapportata alla larghezza comune dei picchi

$$R_s = \Delta X/4\sigma \quad (5)$$

e $4\sigma R_s$ viene definita, attraverso il simbolo x_0 , come interdistanza critica di risoluzione. Ovvero due o più SC adiacenti che si trovino ad interdistanza inferiore ad x_0 risultano sovrapposti e quindi considerati come un *cluster* di picchi.

Il fattore di saturazione rappresenta il grado relativo di separazione che si vuole raggiungere ed è definito da:

$$\alpha_{1D} = m/n_{c,1D} \quad (6)$$

Gli stessi parametri possono essere facilmente estesi a separazioni 2D (13, 15, 17). La capacità di macchia è definita come:

$$n_{c,2D} = A/A_0 \quad (7)$$

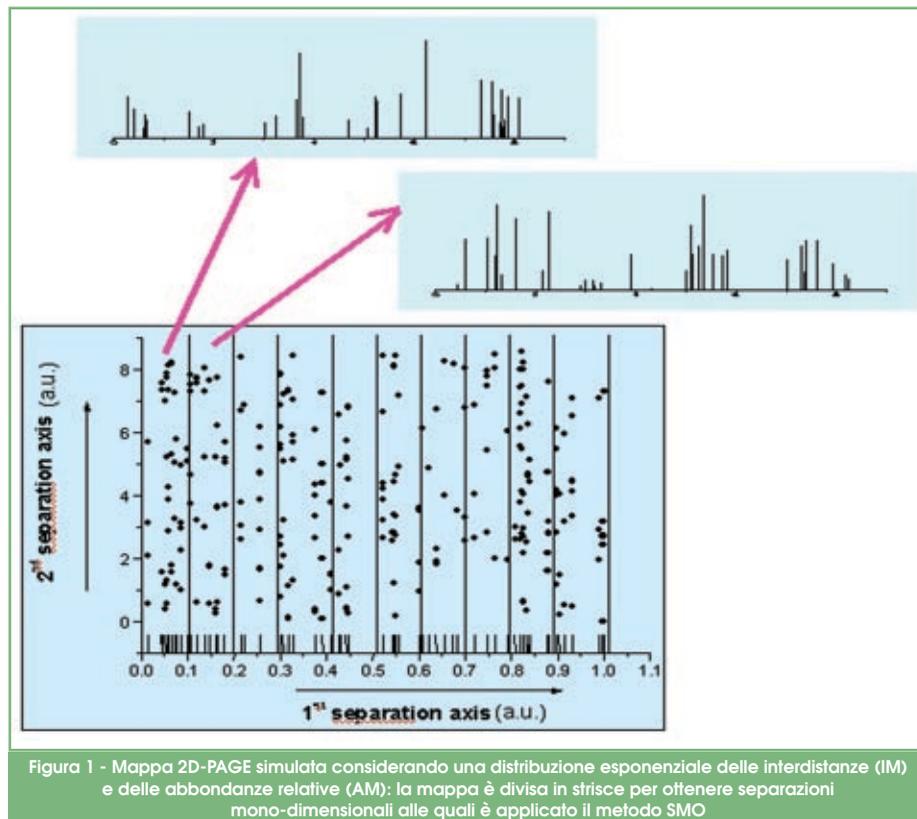


Figura 1 - Mappa 2D-PAGE simulata considerando una distribuzione esponenziale delle interdistanze (IM) e delle abbondanze relative (AM): la mappa è divisa in strisce per ottenere separazioni mono-dimensionali alle quali è applicato il metodo SMO

dove A rappresenta l'area dello spazio separativo bi-dimensionale e A_0 l'area delle macchie, che per semplicità viene assunta essere la stessa per tutti i SC. L'analogia tra $n_{c,1D}$ e $n_{c,2D}$ risulta chiara se si osserva che il denominatore rappresenta il valore critico in base al quale discriminare la sovrapposizione o meno tra i SC, mentre il numeratore è lo spazio (lunghezza in 1D, area in 2D) all'interno del quale avviene la separazione. Analogamente per il fattore di saturazione si ha:

$$\alpha_{2D} = m/n_{c,2D} \quad (8)$$

Oltre a questi parametri principali, altri attributi possono essere calcolati per caratterizzare completamente una separazione multicomponente. A causa della sovrapposizione sempre presente nella separazione, non tutti possono essere direttamente calcolati dal segnale sperimentale, pertanto possono essere suddivisi in tre gruppi principali: osser-

vabili, intrinseci e nascosti (20). Gli attributi *intrinseci* sono quelli che descrivono le caratteristiche della miscela e le proprietà del sistema di separazione utilizzato. Quelli *osservabili* sono quelli sperimentalmente ricavabili dalla separazione, sia direttamente dalla separazione stessa, quale risultato dell'analisi, ma anche con calcoli senza alcuna ipotesi *a priori*. Infine, gli attributi *nascosti* raggruppano tutti quegli effetti e grandezze che sono il risultato dell'incompletezza della separazione. È perciò di fondamentale importanza disporre di metodi di calcolo opportuni per elaborare i dati sperimentali (attributi osservabili) al fine di ricavare tutta l'informazione analitica in essi nascosta, cioè gli attributi intrinseci e nascosti.

Modello Statistico di Sovrapposizione (SMO)

Tra i diversi modelli sviluppati per descrivere le separazioni complesse, (15, 17, 19), gli Autori hanno sviluppato la *teoria quantitativa di sovrapposizione dei picchi* (teoria SMO

quantitativa) (21, 22). Tale procedura, traendo origine da modelli sviluppati da Giddings per studiare il *grado statistico di sovrapposizione* (SDO) (23, 24), è in grado di descrivere le proprietà della separazione considerata come un insieme di eventi (la ritenzione dei SC) le cui proprietà medie possono essere stimate attraverso opportune funzioni statistiche. Il modello è stato originariamente sviluppato per cromatogrammi monodimensionali e successivamente esteso a separazioni 2D (25, 26).

La teoria SMO quantitativa si basa sulla descrizione del cromatogramma in termini di funzioni di probabilità che rappresentano due diverse distribuzioni: il *modello delle interdistanze* (IM), che descrive la posizione dei singoli componenti lungo l'asse dei tempi di ritenzione, e il *modello delle abbondanze* (AM) che rappresenta la distribuzione delle abbondanze dei componenti. Partendo da parametri osservabili, come il tempo di ritenzione e l'area/altezza dei picchi, vengono selezionati diversi valori di interdistanza critica (correlati alla risoluzione cromatografica richiesta) e per ognuno di essi vengono contati i picchi visibili nel cromatogramma e calcolato il valore medio di area osservata, \bar{v}_{obs} .

Nel caso più semplice e più generale, quando sia IM e sia AM mostrano una distribuzione esponenziale, che corrisponde ad una distribuzione casuale delle posizioni e delle abbondanze dei componenti, si può derivare una relazione lineare tra \bar{v}_{obs} e x_0 (21):

$$\ln \bar{v}_{obs} = \ln \bar{v} + m x_0 / X \quad (9)$$

dove \bar{v} corrisponde all'intensità media dei picchi separati. I punti sperimentali \bar{v}_{obs} e x_0 possono così essere fittati al modello di Eq. 9, che descrive una retta, la cui pendenza rappresenta una stima statistica di m , il numero di SC. È possibile estendere il modello a diverse funzioni IM e AM, in grado di descrivere la grande variabilità dei cromatogrammi sperimentali (21, 27). Con queste funzioni si sono sviluppate espressioni analitiche per calcolare il grado statistico di sovrapposizio-

ne (in termini di frazione di p , n -picchi formati da componenti in rapporto alla capacità di picco) in funzione del fattore di saturazione:

$$p/n_{c,1D} = \alpha \exp(-\alpha) \quad (10)$$

$$n/n_{c,1D} = \alpha \exp(-2\alpha)(1-\exp(-\alpha))^{n-1} \quad (11)$$

per $n=1,2,3,4$ si hanno rispettivamente i rapporti $s/n_{c,1D}$, $d/n_{c,1D}$, $t/n_{c,1D}$, $q/n_{c,1D}$.

Si è dimostrato matematicamente e verificato sperimentalmente che la funzione esponenziale è il modello che meglio descrive la distribuzione delle interdistanze e delle abbondanze relative per miscele estremamente complesse (21).

Metodo di calcolo

Il metodo proposto è stato validato mediante un approccio di tipo simulativo: utilizzando calcoli di simulazione numerica si sono generate separazioni 2D con proprietà note, ad esse si è applicato il metodo e si è testata la validità dei risultati ottenuti (25, 26).

Nella simulazione numerica si sono prese in considerazione separazioni 2D-PAGE e si sono sviluppati opportuni algoritmi per generare separazioni con le caratteristiche delle mappe 2D-PAGE di miscele di proteine, di rilevanza fondamentale nella proteomica. Ciascuna macchia di ogni SC viene considerata come un punto nella mappa 2D descritto da due coordinate di posizione (p ed M_r del baricentro delle macchie) e da una terza coordinata che ne descrive l'intensità. L'intera mappa 2D è stata poi suddivisa in diverse strisce, per le quali viene considerata solo una coordinata di separazione (Figura 1): in tale modo l'intera separazione 2D viene ridotta ad una serie di separazioni 1D sui quali applicare il metodo sviluppato. Infine, i risultati ottenuti per ogni singola striscia sono stati sommati in modo da ottenere la stima per l'intera mappa 2D.

Allo scopo di estrarre un campione statisticamente valido dalla popolazione delle infinite possibili mappe 2D generate, per ogni gruppo di parametri sperimentali la procedura di simulazione è stata replicata 500 volte ottenendo 500 mappe generate in modo casuale: i risultati riportati costituiscono i valori medi di queste prove ripetute. Usando questa procedura si sono calcolati il numero di macchie visibili (p), di singoletti (s), doppietti (d), triplete (t) e quadruplette (q) per diversi valori di interdistanza critica x_0 .

Per generare i dati relativi alle posizioni delle macchie nella mappa, si è utilizzato il modello più generale che descrive IM, cioè quello esponenziale. Inoltre si sono generate mappe con caratteristiche simili a quelle di mappe 2D reali: si è sviluppata una metodologia di simulazione basata sullo stesso modello IM presente nelle mappe sperimentali. Poiché non è nota alcuna formula funzionale di questa distribuzione, si è ricostruita una distribuzione discontinua sulla base di dati sperimentali: si sono presi in considerazione i dati di 1.956 proteine da campioni di tessuti umani ottenute dal SWISS 2D-PAGE database (28). Questa distribuzione costituisce il riferimento per la generazione casuale ed indipendente di coordinate p ed M_r , attraverso un opportuno algoritmo basato sul *Metodo di Reiezione*: con questo algoritmo è possibile generare in modo casuale mappe in cui la posizione delle macchie pre-

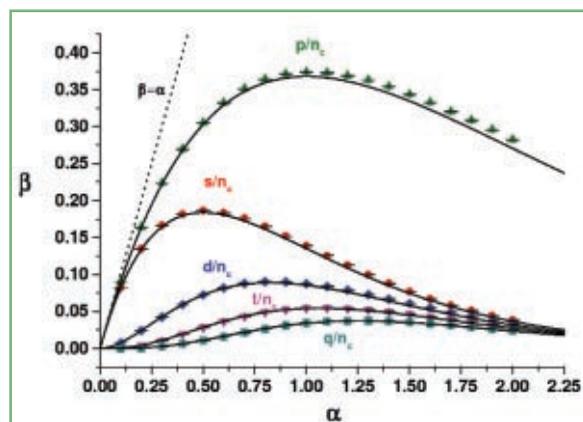


Figura 2 - Funzioni beta confrontate con dati simulati (2000 SC) in un ampio intervallo del fattore di saturazione, α . La retta tratteggiata indica l'andamento della funzione β in assenza di sovrapposizione

Tabella 1 - Calcolo di m su mappe simulate utilizzando il modello IM esponenziale per la prima e la seconda dimensione di separazione

SC	Condizione di simulazione	m
2.000	random	1.964±44
1.000	random	966±31
500	random	465±22
250	random	217±15
2.000	non-random	1.963±44
1.000	non-random	963±31
500	non-random	463±22
250	non-random	218±15

sentia la stessa distribuzione del modello di riferimento (29). Infatti nelle mappe ottenute, come nelle maggior parte delle mappe sperimentali di tessuti umani, si nota un elevato affollamento di macchie nella parte centrale della mappa, a valori di pI compresi tra 5 e 7 pH e di M_r tra 20.000 e 80.000 Da (6-8).

Risultati e discussione

Applicabilità delle distribuzioni teoriche per IM e AM

Su mappe simulate è possibile stimare quantitativamente il grado di sovrapposizione presente, calcolando numericamente i parametri $s/n_{c,1D}$, $d/n_{c,1D}$, $t/n_{c,1D}$, $q/n_{c,1D}$. I risultati ottenuti sono riportati in Figura 2, dove per confronto sono mostrate le curve teoriche descritte dalle Eq. 10 e 11. I dati riportati sono un valore medio calcolato su tutte le strisce in cui è stata suddivisa la mappa, replicando la procedura su 500 cicli di simulazione della stessa mappa e mediando successivamente questi risultati. Poiché le equazioni 10 e 11 sono state ottenute sulla base dell'ipotesi di distribuzione esponenziale per IM, il buon accordo che esse dimostrano con i dati ottenuti da simulazione è una dimostrazione che un modello IM di tipo esponenziale descrive in modo adeguato il pattern di separazione. Le curve riportate in Figura 2 forniscono anche un significato fisico concreto al concetto di sovrapposizione: esaminando il grafico si

può notare come è pesante l'effetto della sovrapposizione all'aumentare di α , grado di saturazione della separazione. La linea retta tratteggiata rappresenta il caso ideale in cui la sovrapposizione è assente (tutti gli m componenti appaiono come una macchia pura nella mappa, perciò $m=p=s$ e quindi $\beta=\alpha$). Nella separazione reale la

curva $\beta=p/n_c$ si scosta dalla retta $\beta=\alpha$ dopo un breve tratto rettilineo, già da $\alpha=0,025$ a causa della sovrapposizione $m>p>s$ e questa differenza aumenta con l'aumento della saturazione. È proprio a causa di questa divergenza che i dati ottenibili dalla mappa (numero di macchie p) non sono indice della reale complessità della miscela (numero dei componenti m), ma è necessaria un'opportuna elaborazione matematico-statistica per ottenere questo valore.

Stima del numero dei componenti in separazioni 2D

Per testare l'accuratezza della procedura SMO nella stima del numero di componenti, m , il metodo è stato applicato a mappe simulate tramite calcolo numerico in cui il numero di componenti è noto. Le mappe sono state generate in base a una distribuzione esponenziale delle interdistanze: i

Tabella 2 - Calcolo di m su mappe simulate che rappresentano separazioni 2D PAGE di miscele di proteine, ottenute utilizzando il Metodo di Reiezione

SC	N° strisce	SC minimo per striscia	m
2.000	30	-	2.088±46
2.000	40	-	1.960±44
2.000	50	-	1.940±44
2.000	50	27	1.972±44
2.000	50	30	1.991±45
2.000	70	43	1.994±45
1.000	10	-	1.068±33
1.000	20	-	1.026±32
1.000	30	-	1.012±32
1.000	50	10	1.005±32
1.000	60	15	991±31
1.000	60	20	998±32
500	10	-	509±23
500	15	-	485±22
500	10	10	511±23
500	20	20	499±22
250	10	-	212±15
250	5	-	245±16
250	10	20	242±16

risultati ottenuti sono riportati in Tabella 1. Ogni mappa è stata suddivisa in 10 strisce lungo l'asse della prima dimensione. Si sono simulate due diverse condizioni di separazione: il termine *random* si riferisce a distribuzioni spaziali dei SC totalmente casuali (ogni SC è stato separato lungo la prima dimensione, poi sviluppato lungo la seconda dimensione); *non-random* sta ad indicare separazioni maggiormente ordinate in cui solo il 30% dei SC è stato separato nella prima dimensione. Dai risultati ottenuti si può notare che la procedura è in grado di stimare correttamente il numero di SC. E che l'accuratezza e la precisione della stima aumentano con l'aumentare del numero di componenti nella miscela.

La procedura è stata applicata con successo a mappe che simulano quelle sperimentali 2D PAGE di proteine. Si sono utilizzate mappe generate con il Metodo di Reiezione contenenti numeri diversi di SC: la Tabella 2 riporta i risultati ottenuti. Si sono utilizzate 2 diverse procedure di calcolo: in un caso l'intera mappa è stata suddivisa in un numero fisso di strisce (seconda colonna), nell'altro il numero di strisce (seconda colonna) è variabile in funzione del numero minimo di SC fissato per ogni striscia (terza colonna). I risultati ottenuti indicano che è possibile stimare m con buona accuratezza e precisione (confronta prima e quarta colonna), anche quando il numero di componenti presenti nella mappa è basso; in particolare i migliori risultati si sono ottenuti con la procedura con il numero di strisce variabile.

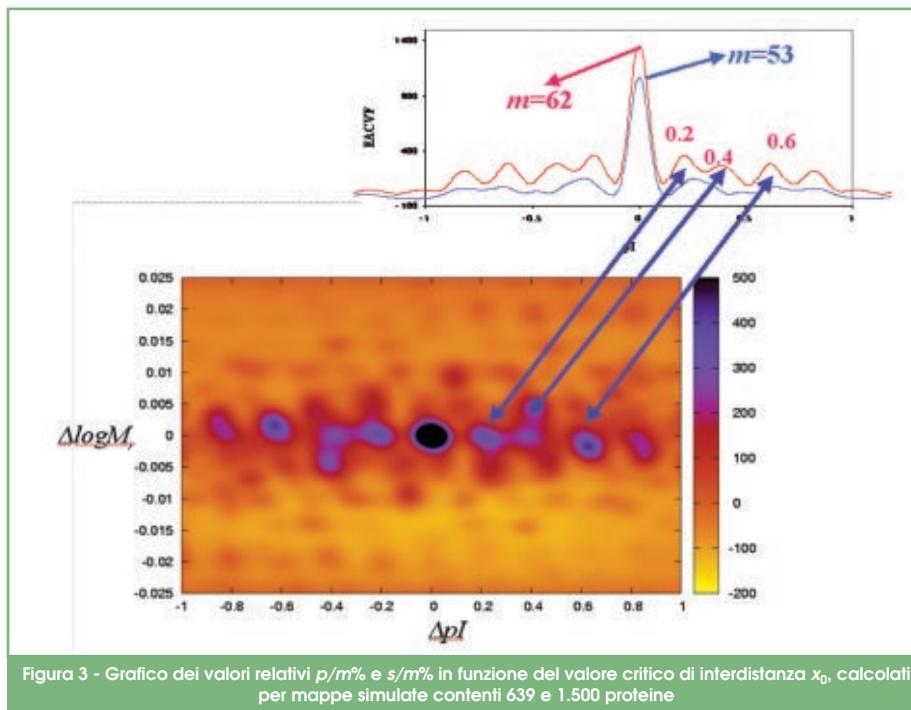


Figura 3 - Grafico dei valori relativi $p/m\%$ e $s/m\%$ in funzione del valore critico di interdistanza x_0 , calcolati per mappe simulate contenenti 639 e 1.500 proteine

Valutazione della sovrapposizione

Applicando il metodo descritto a mappe che simulano mappe reali è possibile stimare quantitativamente il grado di sovrapposizione da aspettarsi nelle mappe 2D-PAGE comunemente utilizzate negli studi di proteomica. In questo settore è fondamentale massimizzare il numero di macchie rivelabili, p , correlato al numero di proteine che possono essere identificate, e il numero di macchie pure, s , dal momento che un elevato grado di separazione e di purezza della macchia è il requisito necessario per un'identificazione più completa possibile della struttura delle proteine. Per mappe

simulate si sono calcolati il numero relativo di macchie, $p/m\%$, e di singoletti, $s/m\%$, che si possono contare in una mappa e si sono riportati in funzione di x_0 (Figura 3). Si può notare che, a parità di numero di proteine nel campione, il numero di picchi, $p/m\%$, e di picchi puri, $s/m\%$ diminuisce in modo significativo all'aumentare dell'interdistanza critica x_0 : infatti il grado di sovrapposizione nella mappa aumenta come conseguenza del maggiore affollamento nello stesso spazio separativo. La Tabella 3 riporta i valori di x_0 che corrispondono ad alcune condizioni sperimentali comuni in mappe 2D-PAGE di proteine (6-8). Perciò,

Decoding Multidimensional Separations. Application to Proteomics

ABSTRACT 

A statistical approach able to extract useful information contained in bi-dimensional multicomponent separations is here reported. It is based on a quantitative theory of the statistical degree of overlapping: it has been previously developed on monodimensional linear chromatographic separations, and then extended to bi-dimensional separation techniques. Theoretical expressions and computer algorithms were developed starting from simple mathematical models: the procedure has been validated on computer-generated separations and allows the accurate estimation of the intrinsic number of components in the sample. Application to separations simulating 2D PAGE maps of complex protein mixtures, the core of proteomics, is reported.

Tabella 3 - Valori di interdistanza critica, x_0 (in unità ΔpH), corrispondenti a diverse condizioni sperimentali della prima dimensione di separazione (IEF) di mappe 2D-PAGE

Lunghezza strip IEF (cm)	Broad-range (pH 3-10) gel	Narrow-range (1 ΔpH) gel
7	0,1	0,014
11	0,063	0,009
18	0,038	0,005

ipotizzato il numero di proteine presenti nel campione, si può stimare il numero di macchie e di macchie pure che si ottengono in una mappa quando si utilizza uno strip di una certa dimensione (7-18 cm) e si opera ad un certo intervallo di pH: in elettroforesi convenzionale, tra pH 3 e 10, o in narrow-range, utilizzando un intervallo di pH unitario (7). Per esempio, in Figura 3 è indicato il caso di una separazione convenzionale utilizzando uno strip di 11 cm ($x_0=0,063$ DpH): in una miscela con 639 proteine si ottiene una separazione completa (singoletti) di soltanto il 22% dei componenti (freccia rossa) e del 18% in una miscela con 1.500 proteine (freccia blu).

Conclusioni

I risultati riportati nel presente lavoro indicano che, anche utilizzando sistemi separativi bi-dimensionali con un elevato potere separativo, come il caso di mappe 2D-PAGE, il potere risolutivo non è generalmente sufficiente per ottenere una separazione completa di tutti i componenti di una miscela complessa, anche sotto le migliori condizioni sperimentali. La sovrapposizione, infatti è una conseguenza inevitabile della complessità stessa della miscela analizzata. Pertanto in queste condizioni è di fondamentale importanza avere a disposizione metodi matematico-statistici per poter predire in modo quantitativo tale sovrapposizione ed interpretare la separazio-

ne complessa per estrarre tutte le informazioni analitiche sul campione contenute nella separazione ed in essa nascoste a causa della sovrapposizione. Solo in questo modo si può ottenere il massimo risultato dallo sforzo sperimentale di un sistema analitico complesso, come un sistema bi-dimensionale.

La teoria quantitativa di sovrapposizione dei picchi (teoria SMO quantitativa) si è dimostrata un metodo molto utile ed efficiente per caratterizzare la sovrapposizione presente in separazioni complesse, anche per lo studio di separazioni bi-dimensionali. Esistono attualmente altri modelli (19, 22, 30) ed altre procedure potranno essere sviluppate in futuro per studiare la complessità delle separazioni multicomponenti. Questi risultati evidenziano che le nuove direzioni di sviluppo della Scienza delle Separazioni devono essere rivolte sia alla produzione di strumentazioni più sofisticate ed efficienti per ottenere una maggior quantità di dati accurati, sia allo sviluppo di metodi per interpretare la grande quantità di dati analitici al fine di estrarre da essi il maggior numero di informazioni.

Bibliografia

- (1) T. Hyotylainen, M.L. Riekkola, *Journal of Chrom. A*, 2003, **1000**, 357.
- (2) J. Dalluge *et al.*, *Journal of Chrom. A*, 2003, **1000**, 69.
- (3) M.C. Pietrogrande *et al.*, *Chromatographia*, 2002, **55**, 189.
- (4) M.C. Pietrogrande *et al.*, *J. High Resolution Chromatogr.*, 1994, **17**, 839.
- (5) I.D. Wilson, U.A.T. Brinkman, *Journal of Chrom. A*, 2003, **1000**, 325.
- (6) Proteome Research: New Frontiers in Functional Genomics, M.R. Wilkins *et al.* (Eds.), Springer, New York, 1997.
- (7) P.G. Righetti *et al.*, *The Proteome Revisited: Theory and Practice of All Relevant Electrophoretic Steps*, Elsevier, Amsterdam, 2001, 275.
- (8) Proteomics: a Trend Guide, W. Blackstock, M. Mann. (Eds.), Elsevier, London, 2000.
- (9) S.E. Ong, A. Prandey, *Biomol. Engin.*, 2001, **8**, 195.
- (10) P.G. Righetti *et al.*, *Anal. Chem.*, 2001, **73**, 320A.
- (11) B. Kuster *et al.*, *Proteomics*, 2000, **1**, 641.
- (12) T. Rabilloud, *Proteomics*, 2002, **2**, 3.
- (13) J.C. Giddings, *Unified Separation Science*, Wiley, New York, 1991.
- (14) J.M. Davis, *Anal. Chem.*, 1991, **63**, 2141.
- (15) S.A. Roach, *The Theory of Random Clumping*, Methuen and Co., London, 1968.
- (16) W. Shi, J.M. Davis, *Anal. Chem.*, 1993, **6**, 482.
- (17) J.C. Giddings, *J. Chromatogr. A*, 1995, **703**, 3.
- (18) M. Shure, *J. Microcol. Sep.*, 1997, **9**, 169.
- (19) A. Felinger, *Data Analysis and Signal Processing in Chromatography*, Elsevier, Amsterdam 1998, 366.
- (20) F. Dondi *et al.*, *Chromatographia*, 1997, **45**, 435.
- (21) F. Dondi *et al.*, *Anal. Chem.*, 1998, **70**, 766.
- (22) A. Felinger, M.C. Pietrogrande, *Anal. Chem.*, 2001, **73**, 619A.
- (23) J.M. Davis, J.C. Giddings, *Anal. Chem.*, 1983, **55**, 418.
- (24) J.M. Davis, J.C. Giddings, *Anal. Chem.*, 1985, **57**, 2168.
- (25) M.C. Pietrogrande *et al.*, *Electrophoresis*, 2002, **23**, 283.
- (26) M.C. Pietrogrande *et al.*, *Electrophoresis*, 2003, **24**, 217.
- (27) M.C. Pietrogrande *et al.*, *Chemometrics Intel. Lab. Systems*, 1995, **28**, 239.
- (28) www.expasy.org.
- (29) W.H. Press *et al.*, *Numerical Recipes in Fortran*, Cambridge,