

Università degli Studi di Firenze

DOTTORATO DI RICERCA IN CHIMICA E TECNOLOGIA DEL FARMACO

XXV CICLO

SETTORE SCIENTIFICO DISCIPLINARE

CHIM/01

***Sviluppo di metodologie analitiche di tipo elettroforetico
e loro impiego in problematiche farmaceutiche***

Candidato:

Dr. Benedetta PASQUINI

Docente supervisore:

Prof. Sandra FURLANETTO

Coordinatore del corso:

Prof. Elisabetta TEODORI

2010-2012

INDICE

Riassunto	1
1. Introduzione	2
2. Parte sperimentale	4
3. Tecniche chemiometriche.....	4
3.1. Disegno sperimentale.....	4
3.2. Disegno di miscele	15
3.3. Modelli misti e disegni sperimentali a matrice mista di variabili miscela-processo	21
3.4. Analisi delle componenti principali	24
3.5. Analisi dei cluster	28
3.6. Metodi di classificazione multivariata	31
4. Spettroscopia NMR e Tecniche computazionali di simulazione	33
4.1. Principi generali	33
4.2. NMR bidimensionale	42
4.3. Dinamica Molecolare.....	43
5. Elettroforesi Capillare.....	46
5.1. Principi generali	46
5.2. Modalità operative	53
5.3. Utilizzo di Ciclodestrine	63
5.4. Convalida	66
6. Quality by Design	68
6.1. Principi generali	68
6.2. Schema generale applicativo del QbD	70
7. Risultati e discussione.....	82
7.1. Studio alla base dei meccanismi di separazione in cromatografia elettrocinetica	82

7.2. Sviluppo di un metodo MEEKC per l'analisi del Ramipril e delle sue impurezze	87
7.3. Sviluppo di un metodo MEEKC per l'analisi di un integratore a base di CoQ10	100
7.4. Caratterizzazione geografica del miele mediante un approccio chemiometrico.....	116
8. Conclusioni.....	128
9. Bibliografia.....	129
10. Appendice.....	138

Riassunto

Il presente lavoro di tesi si inserisce nel settore della ricerca analitico farmaceutica ed ha previsto lo sviluppo di metodiche analitiche mediante approccio chemiometrico. In questo contesto sono state studiate fasi pseudo stazionarie da utilizzarsi in cromatografia elettrocinetica e sono state sviluppate metodiche analitiche da impiegarsi in ambito di assicurazione di qualità. Il lavoro di tesi ha in particolare riguardato lo sviluppo mediante strategia multivariata di metodi di elettroforesi capillare che hanno permesso di risolvere diverse problematiche farmaceutiche comprese quelle inerenti problemi di classificazione. L'approccio multivariato ha permesso di ottenere la massima quantità di informazione con un risparmio di tempo e costi rispetto, permettendo di ottenere informazioni aggiuntive circa le interazioni tra i fattori in studio.

Due lavori hanno riguardato l'ottimizzazione di metodi da applicarsi per la determinazione simultanea di principi attivi e delle loro impurezze nelle forme farmaceutiche. I farmaci presi in esame sono stati la Clemastina Fumarato e il Ramipril. Dopo la fase di ottimizzazione, i metodi studiati sono stati convalidati secondo le linee guida ICH per assicurare adeguate prestazioni del metodo e dati affidabili per la specifica applicazione a cui erano destinati. Per la prima volta in letteratura è stato investigato mediante tecniche di Molecular Modeling e NMR il meccanismo alla base della separazione degli analiti in presenza di fasi pseudostazionarie complesse quali le microemulsioni e le ciclodestrine. Ancora lo studio di fasi pseudostazionarie è stato approfondito conducendo uno studio multivariato di un metodo di cromatografia elettrocinetica micellare mediante una matrice mista di variabili di miscela e di processo e studiando le interazioni esistenti tra questi due tipi di variabili. Il metodo è stato applicato per la determinazione del Coenzima Q10 in un integratore alimentare in presenza di altre sostanze bioattive. Per concludere metodi di classificazione multivariata sono stati utilizzati per analizzare alcuni tipi di miele di diversa origine botanica e geografica caratterizzati dal contenuto di specie ioniche determinato mediante cromatografia a scambio ionico.

1. Introduzione

L'analisi farmaceutica riveste un ruolo fondamentale nella tutela della salute dell'uomo permettendo di risolvere molti aspetti legati allo sviluppo ed all'utilizzo di sostanze bioattive di sintesi o di origine naturale. Il suo ruolo chiave è inoltre svolto a livello di assicurazione di qualità che comprende diverse attività volte a garantire la sicurezza e l'efficacia di un prodotto. Nel settore farmaceutico il concetto di qualità di un prodotto si riferisce alle sue proprietà e caratteristiche che includono il confezionamento, gli aspetti organolettici e, soprattutto, il corretto contenuto nella formulazione farmaceutica del/dei principio/i attivo/i. Tutti questi concetti sono applicabili non solo ai farmaci, ma anche agli integratori o nutraceutici, cioè a prodotti contenenti principi attivi estratti da alimenti. In particolare, un nutraceutico è un "alimento-farmaco", cioè un prodotto che contiene all'interno di una idonea formulazione componenti nutrizionali selezionati in dosi appropriate. Il loro mercato è in continua crescita e risulta pertanto essenziale avere a disposizione metodi analitici convalidati applicabili al controllo di qualità. Metodologie analitiche convalidate per il dosaggio del principio attivo e delle sue impurezze sono inoltre richieste dagli organi governativi per la commercializzazione delle specialità medicinali, ed hanno all'interno delle industrie farmaceutiche un ruolo chiave nel controllo di qualità [1]. Molto spesso il controllo di qualità viene effettuato attraverso HPLC, ma ormai anche l'elettroforesi capillare (*Capillary Electrophoresis*, CE) ha guadagnato popolarità come tecnica alternativa o complementare per le applicazioni farmaceutiche [2, 3, 4]. La CE presenta altissima efficienza, ridotto consumo di solventi organici e bassi costi di gestione, non essendo necessario il mantenimento di una varietà di colonne cromatografiche [3, 4]. Un altro importante vantaggio di questa tecnica risiede nella possibilità di sfruttare differenti meccanismi di separazione [3], per cui può essere applicata all'analisi di composti di natura molto diversa. Recentemente sono stati pubblicati diversi lavori sull'applicazione della CE nell'*impurity profiling* [5-11]. La determinazione delle impurezze rappresenta infatti una sfida alla selettività della tecnica, in quanto queste hanno spesso strutture e proprietà molto simili. In ogni caso i requisiti standard del metodo devono assicurare la completa risoluzione tra i picchi relativi ai diversi analiti e

la possibilità di determinare le impurezze ad un livello di concentrazione pari o inferiore allo 0.1% p/p rispetto al principio attivo [3]. Per potere garantire un'elevata selettività del metodo è necessario ottimizzare un numero elevato di variabili (tipo, concentrazione e pH del tampone, tipo e concentrazione di additivi, temperatura, voltaggio, lunghezza del capillare, e così via) che possono anche presentare interazioni tra loro. Queste ultime possono essere individuate solo applicando una strategia di ottimizzazione multivariata in cui i fattori pertinenti sono variati in modo simultaneo e sistematico in modo da ottenere la massima e la migliore qualità di informazione con il minimo sforzo [12]. Da questo deriva l'importanza dello sviluppo di strategie multivariate di disegno sperimentale in analisi farmaceutica ed in particolare lo sviluppo di metodiche analitiche secondo l'approccio "Quality by design". Attraverso tale approccio è infatti possibile definire un "Design space" all'interno del quale la variazione della risposta non è statisticamente significativa. In particolare con Quality by Design (QbD) si intende un nuovo modello di qualità, raccomandato dalle linee guida FDA e ICH, che incoraggia l'uso della chemiometria nello sviluppo di prodotti e processi farmaceutici con l'obiettivo di tenere costantemente sotto osservazione le misure e i parametri di processo per rilevare eventuali deviazioni nella qualità del prodotto prima ancora di rilevare eventuali anomalie mediante metodi tradizionali di campionamento e analisi.

In questo lavoro di tesi sono stati sviluppati vari metodi analitici di elettroforesi capillare che hanno permesso di risolvere le più diverse problematiche farmaceutiche grazie all'aiuto di un approccio multivariato di disegno sperimentale e inoltre sono stati approfonditi vari metodi di classificazione chemiometrica. Due lavori hanno riguardato l'ottimizzazione di metodi per la simultanea determinazione di principi attivi nelle forme farmaceutiche e delle loro impurezze, rispettivamente Clemastina Fumarato e Ramipril. Dopo la fase di ottimizzazione, i metodi studiati sono stati convalidati secondo le linee guida ICH per assicurare adatte prestazioni del metodo e dati affidabili. Studi di Molecular Modeling e NMR sono stati effettuati per capire meglio i meccanismi alla base della separazione dei nostri analiti. Lo sviluppo di un metodo per la determinazione di Coenzima Q10 in un composto nutraceutico in presenza di altre sostanze bioattive ha invece richiesto l'utilizzo di una matrice mista delle variabili miscela-processo. Per concludere metodi di classificazione multivariata sono stati utilizzati per classificare alcuni tipi di miele di diversa origine botanica e geografica

caratterizzati dal contenuto di ioni determinati mediante cromatografia a scambio ionico.

2. Parte Sperimentale

Le tecniche ed i metodi discussi nella presente tesi sono stati oggetto o sono in corso di pubblicazione, pertanto per la relativa parte sperimentale si rimanda agli estratti allegati nel cap.10 “Appendice”.

3. Tecniche chemiometriche

3.1. Disegno sperimentale

Nell'ambito delle scienze sperimentali la selezione degli esperimenti da condurre riveste grande importanza; una strategia ideale dovrebbe:

- fornire i risultati più velocemente possibile;
- evitare esperimenti non necessari;
- assicurare che i risultati siano i più affidabili possibile;
- fornire un modello di ottimizzazione dei fenomeni studiati [13].

Il metodo classico nello studio di un fenomeno è l'approccio OVAT (*One Variable at A Time*), in cui viene effettuata la variazione di una sola variabile alla volta, tenendo tutte le altre ad un livello fisso. Nonostante questo sia un approccio semplice, facile da controllare e interpretare, le sue limitazioni principali sono la perdita di informazioni sull'interazione tra variabili e l'alto numero di esperimenti necessari [14]. Come alternativa all'approccio OVAT si ha il *disegno sperimentale*, che utilizza un approccio multivariato in cui i valori di tutte le variabili sono modificati simultaneamente in ogni esperimento in modo programmato e razionale [13]. Esso include la definizione preliminare delle condizioni sperimentali, attraverso la definizione delle variabili potenzialmente influenti sull'esperimento e dei loro livelli a cui sono studiate. Sono quindi definite le risposte che rappresentano la traduzione numerica del criterio di ottimizzazione e viene scelto un modello matematico (normalmente una funzione polinomiale al massimo di secondo ordine) che correli le variabili alle risposte. Sulla base del modello selezionato vengono quindi selezionati gli

esperimenti da condurre.

Il ricercatore si interessa in primo luogo delle variabili, supposte o certe, che agiscono sul fenomeno in studio. Tali variabili sono chiamate *Fattori* e possono essere di tipo qualitativo o quantitativo. I fattori quantitativi possono essere continui cioè possono assumere qualsiasi valore all'interno di un insieme determinato di valori, o discreti: in questo caso possono assumere solo alcuni valori prefissati. Gli effetti sulla/e risposta/e di ogni variabile sono determinati imponendo a queste delle variazioni all'interno di un dominio detto *Dominio sperimentale*, all'interno del quale ogni variabile può assumere diversi valori o *livelli*.

Il *dominio sperimentale* [15] è il dominio entro cui si muove lo sperimentatore. La scelta di tale dominio è estremamente importante; può accadere infatti che le prove iniziali non forniscano sufficienti informazioni, ma i risultati ottenuti possono essere sfruttati dallo sperimentatore per definire un nuovo dominio in cui c'è una buona probabilità di trovare la soluzione del problema, e sarà appunto quest'ultimo il dominio d'interesse.

Per rendere comparabili gli effetti delle variabili naturali, che non sono espresse dalle stesse unità di misura, queste vengono sottoposte a due trasformazioni [15].

La prima consiste nel *cambiamento in origine*, per cui la nuova origine è il centro del dominio sperimentale in modo che le variabili risultino centrate.

La seconda porta invece al *cambiamento in unità di misura*, per cui le variabili naturali sono trasformate in variabili ridotte adimensionali.

Variabili *centrate* e *ridotte* sono dette *codificate* e ne risultano di conseguenza stabiliti centro e passo di variazione (step).

$$X_{ij} = \frac{U_{ij} - \bar{U}_j}{\Delta U_j}$$

dove:

X_{ij} è il valore della variabile codificata j nella i -esima prova

U_{ij} è il valore della variabile naturale j nella i -esima prova

\bar{U}_j è il valore della variabile naturale j al centro del dominio d'interesse

ΔU_j è il passo di variazione della variabile naturale j .

Convenzionalmente, nei disegni a due livelli il livello basso di un fattore è

codificato con -1 o -, il livello alto con +1 o + e quello intermedio tra i due (livello centrale) con 0. L'insieme delle diverse condizioni sperimentali è dato dalla *matrice sperimentale* che è un oggetto matematico che rappresenta sotto forma codificata gli esperimenti da effettuare [15]. Essa consiste in una tabella di dati costituita da N righe, corrispondenti a n esperimenti, e K colonne, corrispondenti alle k variabili studiate. L'elemento x_{ij} della matrice così formata corrisponde al valore che assume la j -esima variabile nell' i -esimo esperimento.

Il *piano sperimentale* è invece la traduzione di una matrice sperimentale in variabili naturali: si tratta di una tabella contenente dei dati espressi nelle loro unità di misura, direttamente utilizzabile dallo sperimentatore.

L'*effetto* di un fattore è il cambiamento nella risposta osservato in seguito al cambiamento di livello.

Una *risposta*, indicata con y , è il risultato sperimentale di un'esperienza (tempo di analisi, resa di una reazione e così via), cioè è una manifestazione misurabile di ciò che si osserva facendo variare i fattori studiati. Essa dipenderà quindi da k fattori e si potrà scrivere

$$y = f(x_1, x_2, \dots, x_k)$$

dove la funzione f è chiamata *funzione di risposta* ed è una funzione matematica di tipo polinomiale che descrive la risposta come funzione dei fattori. In generale è possibile rappresentare un problema di questo tipo in forma matriciale:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

dove \mathbf{y} è il vettore colonna delle n risposte, $\boldsymbol{\beta}$ è il vettore delle stime dei coefficienti, \mathbf{X} è la matrice $n \times k$ che definisce i punti sperimentali ed $\boldsymbol{\varepsilon}$ è il vettore colonna degli errori.

L'obiettivo dello sperimentatore è trovare la dipendenza che esiste tra una o più risposte sperimentali e i fattori selezionati, cioè un *modello matematico* che corredi la risposta alle variabili [15].

Viene quindi ipotizzato un modello sulla base delle conoscenze del ricercatore e viene selezionata una matrice sperimentale che permetta di stimare correttamente i coefficienti del modello.

Il modello utilizzato potrà essere di diverso tipo:

-lineare senza interazioni

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad \text{dove } \varepsilon \text{ è l'errore sperimentale}$$

-lineare con interazioni di ordine $k-1$ (con k variabili)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \beta_{12} x_1 x_2 + \dots + \beta_{1k} x_1 x_k + \beta_{23} x_2 x_3 + \dots + \beta_{2k} x_2 x_k + \dots + \beta_{1jk} x_1 x_j x_k + \dots + \varepsilon \quad \text{con } j \neq k$$

-non lineare, comprendente cioè anche termini quadratici.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \beta_{12} x_1 x_2 + \dots + \beta_{1k} x_1 x_k + \beta_{11} x_1^2 + \dots + \beta_{kk} x_k^2 + \varepsilon$$

La scelta del modello è legata strettamente al dominio sperimentale: più il dominio è piccolo, più il modello sarà semplice. Dalla scelta del modello dipenderanno il tipo di informazione ottenuta, il numero e il tipo di esperimenti da condurre e quindi il tipo di pianificazione sperimentale. La matrice sperimentale dovrà infatti contenere come minimo un numero di esperimenti pari al numero di coefficienti da determinare.

Si hanno inoltre a disposizione dei *criteri di qualità* indipendenti dai risultati sperimentali che permettono di conoscere a priori con quale accuratezza saranno stimati i coefficienti e di apprezzare la qualità di una matrice sperimentale. Di tali criteri si parlerà più diffusamente in seguito.

Dalla matrice sperimentale, nota l'equazione del modello, si ottiene la *matrice del modello*: questa è costituita da tante righe quanti sono gli esperimenti e da tante colonne quanti sono i termini presenti nell'equazione del modello. Nel caso di un modello lineare senza interazioni la matrice del modello differisce dalla matrice sperimentale per la sola colonna relativa al termine noto b_0 , la quale è costituita da tutti +1. Nel caso di un modello con interazioni si ha inoltre l'aggiunta di colonne relative a ciascuna interazione ottenute moltiplicando fra loro le colonne relative alle variabili di cui si vuole conoscere l'interazione.

Nota la matrice del modello è possibile calcolarne la sua trasposta \mathbf{X}' (scambiandone le righe con le colonne) e quindi la *matrice di informazione*, data da $\mathbf{X}'\mathbf{X}$, la cui inversa è la *matrice di dispersione* $(\mathbf{X}'\mathbf{X})^{-1}$, che risulta utile per il calcolo delle caratteristiche della matrice.

La matrice di dispersione compare anche nel calcolo dei coefficienti del modello (i coefficienti del modello β non sono affetti da errore, mentre i coefficienti b stima dei β sono comprensivi dell'errore sperimentale), che si ottengono per mezzo dell'analisi di

regressione lineare multivariata applicando la formula seguente:

$$\mathbf{B}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

dove \mathbf{B} è il vettore delle stime b_j , \mathbf{Y} è il vettore risposte contenente le risposte ottenute dagli esperimenti effettuati secondo un'opportuna matrice sperimentale, \mathbf{X}' è la trasposta della matrice del modello e $(\mathbf{X}'\mathbf{X})^{-1}$ è la matrice di dispersione.

Una volta che gli esperimenti sono stati effettuati e i coefficienti calcolati i risultati possono essere analizzati ed interpretati soprattutto attraverso test statistici. In particolare l'applicazione dei metodi statistici ai dati raccolti è resa possibile dal fatto che gli esperimenti vengono eseguiti secondo un'opportuna progettazione iniziale e quindi secondo adatte matrici.

L'importanza di una stima accurata dei coefficienti consiste nel fatto che questa consente di valutare l'effetto che ciascuna variabile ha singolarmente sulla risposta, oppure, nel caso di un'interazione tra variabili, dell'effetto di tale interazione, od ancora, nel caso di un termine quadratico, dell'effetto quadratico della variabile. Si parla in particolare di *effetti principali*, suddivisibili in *lineari* e *quadratici*, e di *effetti di interazione* tra variabili. I primi danno informazioni sul peso esercitato dalla variabile a cui l'effetto è riferito sulla risposta (o sulla curvatura del modello se si tratta di un termine quadratico); i secondi sono indice della interazione tra variabili e quindi dell'effetto che una variabile ha sull'effetto esercitato sulla risposta da un'altra variabile. Dal segno dei coefficienti si hanno invece informazioni su quale zona del dominio sperimentale offre i migliori risultati [12].

Modelli lineari con interazioni o modelli quadratici possono essere rappresentati graficamente e tale rappresentazione grafica dell'equazione del modello prende il nome di *superficie di risposta*. La superficie di risposta rappresenta la variazione della risposta nello spazio al variare dei diversi valori assunti dai fattori. La rappresentazione bidimensionale della superficie di risposta costituisce la *superficie di isorisposta* ed è rappresentata da linee concentriche lungo le quali la risposta si mantiene costante.

La *selezione della migliore strategia sperimentale* dovrà essere effettuata in conseguenza della definizione accurata e completa del problema [13], cioè valutando l'eventuale presenza di limitazioni sperimentali, il rischio di errore sistematico, il modo con cui viene stimato l'errore sperimentale, la necessità o meno di studiare tutti i fattori fin dall'inizio, e così via.

In generale nella prima fase del disegno sperimentale si potrà utilizzare un modello lineare senza interazioni in modo da valutare l'importanza di ogni variabile sulla risposta ed aggiustare il dominio sperimentale di ciascuna variabile, inizialmente piuttosto ampio, a seconda delle indicazioni ottenute. Grazie ad una fase di *screening* così concepita sarà quindi possibile attribuire il valore ottimale ad alcune variabili e restringere il dominio sperimentale delle variabili ancora in studio. In particolare le variabili fissate saranno quelle che in base ai valori dei coefficienti associati a ciascuna di esse risulteranno non influenti sulla risposta oppure influenti a tal punto da non avere dubbi circa il loro valore ottimale. Una volta ridotto il numero delle variabili si potrà condurre uno studio più approfondito del problema postulando un modello che descriva in modo più dettagliato il fenomeno e cioè che contenga termini rappresentativi delle interazioni tra variabili ed anche, nel caso che il numero delle variabili lo consenta, della curvatura del modello (termini quadratici) [12].

Una volta trovato il modello matematico, tale modello deve essere validato previsionalmente. Si procede dunque attraverso due fasi, e cioè una *fase di predizione*, nella quale si sostituiscono nell'equazione del modello i valori codificati delle variabili ottimizzate predicendo in tal modo la risposta secondo il modello costruito, ed una *fase di convalida*. In quest'ultima si conduce l'esperimento nelle condizioni ottimizzate e si confronta la risposta ottenuta con la risposta predetta. In particolare si calcola l'intervallo di confidenza della risposta predetta per un certo grado di probabilità ed un certo numero di gradi di libertà. Si verifica quindi che il valore misurato per la risposta ricada nell'intervallo di confidenza così costruito, evidenziando in questo modo un buon accordo tra la risposta ottenuta e quella calcolata dal modello.

3.1.1. Matrici

La scelta della matrice da utilizzare dipende dal modello ipotizzato. È possibile effettuare una prima distinzione tra matrici utili per stimare modelli lineari senza e con interazioni, utilizzate principalmente in fase di screening, e matrici adatte a stimare modelli quadratici, utilizzate essenzialmente per ottenere delle superfici di risposta.

Di seguito viene descritta la matrice che nella seguente tesi è stata utilizzata come punto di partenza per effettuare lo studio delle variabili di processo, poi opportunamente combinata con il disegno di miscele per formare una matrice mista di variabili miscela-

processo.

Central composite design (CCD)

Il *central composite design* è un disegno fattoriale a due livelli ampliato con ulteriori punti sperimentali al fine di studiare le variabili a più di due livelli [12]. Tale disegno permette la stima dei coefficienti di un modello quadratico del tipo:

$$y = b_0 + \sum b_i x_i + \sum b_{ii} x_i^2 + \sum \sum b_{ij} x_i x_j \quad (\text{con } i \neq j = 1, 2, \dots, k)$$

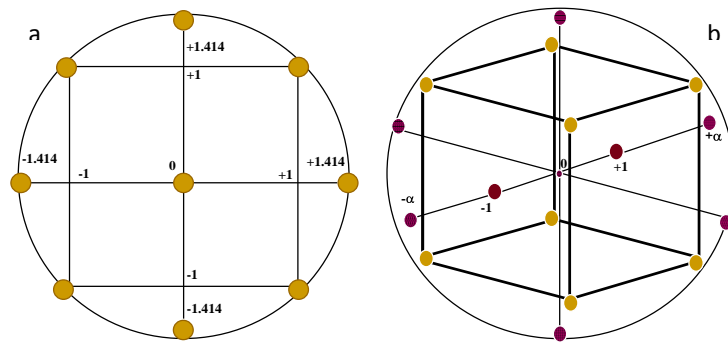
e richiede un numero di esperimenti pari a $2^k + 2k + n$, dove k rappresenta il numero delle variabili e n il numero di esperimenti condotti al centro del dominio sperimentale. Il numero degli esperimenti necessari aumenta rapidamente con il numero delle variabili studiate.

Il disegno consiste di tre parti:

-un disegno fattoriale saturo (2^k) o frazionario (2^{k-p}) usato per stimare i coefficienti dei termini lineari (b_i) e di interazione (b_{ij}) del modello (punti a distanza ± 1 dal centro del dominio sperimentale)

- n esperimenti al centro del dominio sperimentale che permettono di ottenere una stima della varianza sperimentale

- $2k$ esperimenti simmetricamente spazati a $\pm\alpha$ lungo l'asse relativo ad ogni variabile, cioè punti le cui proiezioni cadono al centro delle iperfacce dell'ipercubo che rappresenta le dislocazioni nello spazio dei fattori degli esperimenti del disegno fattoriale completo. Questi punti sperimentali, detti punti assiali o punti star, hanno quindi livello $\pm\alpha$ in un fattore e livello 0 negli altri. Gli esperimenti all'estremità della stella sono usati per stimare i coefficienti dei termini quadratici (b_{ii}) nel modello. In generale, il valore assunto da α dipende dal numero di esperimenti nella parte fattoriale del CCD. Più in particolare, nel caso di due variabili α è uguale a 1.414 (diagonale del quadrato di lato 1), nel caso di tre variabili è uguale a 1.627 (diagonale del cubo di lato 1).



Central Composite Design per 2 fattori (a) e per 3 fattori (b).

3.1.2. Analisi della varianza e analisi statistica dei coefficienti

L'analisi della varianza (ANOVA) è una tecnica statistica usata per analizzare i risultati che dipendono dalla simultanea variazione di un certo numero di fattori. Nel disegno sperimentale tale tecnica viene usata per valutare la significatività e la validità del modello di regressione assunto [16].

La varianza totale SS , espressa come somma dei quadrati (*Sum of Squares*) delle deviazioni dalla media, viene suddivisa nelle componenti corrispondenti alle varie sorgenti di variazione presenti nei dati, SS_{regr} e SS_{resid} , la prima dovuta alla regressione e la seconda ai residui [17]:

$$SS = SS_{regr} + SS_{resid}$$

dove $SS = \sum (y_i - \bar{y})^2$, $SS_{regr} = \sum (y_i - \hat{y}_i)^2$, $SS_{resid} = \sum (y_i - \hat{y}_i)^2$, y_i sono le risposte, \bar{y} è la media delle risposte, \hat{y}_i sono le risposte predette dal modello.

Se sono disponibili delle repliche degli esperimenti, la componente dei residui può essere ulteriormente suddivisa in SS_{lof} e in SS_{pe} , rispettivamente correlate alla perdita di aderenza rispetto al modello (*lack of fit*) e al puro errore sperimentale [17]:

$$SS_{resid} = SS_{lof} + SS_{pe}$$

dove $SS_{pe} = \sum (y_i - \bar{y}_i)^2$ e \bar{y}_i è la media delle risposte ottenute nelle stesse condizioni sperimentali in cui sono state effettuate le repliche.

Le varie somme dei quadrati relative alle diverse sorgenti di variazione sono quindi divise per i gradi di libertà corrispondenti, in modo da ottenere i quadrati medi.

Infine, i quadrati medi vengono confrontati tra di loro attraverso opportuni test statistici. In generale in un test statistico viene definita un'ipotesi nulla H_0 ed un'ipotesi alternativa H_1 specificando il livello di significatività; quindi si prende la decisione di

rifiutare l'ipotesi nulla H_0 o di accettarla in base a regole che si fondano sul livello di significatività.

In particolare nell'ANOVA viene effettuato l' F -test, detto test del rapporto della varianza (*variance ratio test*). In questo caso vengono calcolati i rapporti dei quadrati medi tra diverse sorgenti di variazione ottenendo i rapporti F , che vengono confrontati (test unilaterale destro) con un F critico tabulato (F_{crit}). Se $F > F_{crit}$ l'ipotesi nulla è rifiutata con una probabilità di rifiutare un'ipotesi nulla vera pari a α ; se $F < F_{crit}$ l'ipotesi nulla è accettata con una probabilità di assumere un'ipotesi nulla falsa pari a β [14].

Riassumendo, la tabella dell'ANOVA è del tipo sotto riportato, dove p è il numero di coefficienti da stimare, n è il numero totale di esperimenti effettuati, f è il numero delle differenti combinazioni dei fattori.

Sorgente di variazione	Somma dei quadrati	Gradi di libertà	Quadrati medi	Rapporto F
Regressione	$SS_{regr} = \sum (y_i - \bar{y})^2$	$p-1$	$SS_{regr} / p - 1$	$\frac{SS_{regr}}{p-1} / \frac{SS_{resid}}{n-p}$
Residui	$SS_{resid} = \sum (y_i - \hat{y}_i)^2$	$n-p$	$SS_{resid} / n - p$	
Validità	$SS_{lof} = SS_{resid} - SS_{pe}$	$f-p$	$SS_{lof} / f - p$	$\frac{SS_{lof}}{f-p} / \frac{SS_{pe}}{n-f}$
Errore	$SS_{pe} = \sum (y_i - \bar{y}_i)^2$	$n-f$	$SS_{pe} / n - f$	

Significatività del modello

Nel test di significatività l'obiettivo è decidere se i coefficienti del modello sono significativi. Secondo l'ipotesi nulla H_0 i coefficienti del modello non apportano informazioni alla descrizione del fenomeno, mentre secondo l'ipotesi alternativa H_1 i coefficienti del modello apportano informazioni (senza evidenziare però quali siano i coefficienti significativi) [14]. In questo caso l'ipotesi nulla H_0 prevede che la varianza dei dati non sia spiegata dai fattori presenti nel modello, per cui i rapporti F relativi a regressione e residui non sono significativamente diversi tra loro.

$$H_0: \beta_1 = \dots = \beta_p = 0$$

Affinché il modello sia significativo occorre che $F > F_{crit}$, in modo da accettare l'ipotesi alternativa H_1 secondo la quale uno o più coefficienti del modello sono significativi.

Validità del modello

Per valutare la validità del modello, ovvero valutare se il modello scelto descrive bene il fenomeno in studio, il numero di esperimenti deve essere maggiore del numero dei coefficienti del modello ed inoltre devono essere effettuate delle repliche per ottenere una stima della varianza sperimentale [16]. Questa stima, effettuata ad un dato set di livelli dei fattori, è in realtà una misura della varianza sperimentale in tutto il dominio studiato; infatti, per poter applicare la regressione lineare multipla occorre essere in condizioni di omoscedasticità, cioè di varianza sperimentale costante nel dominio. Un altro metodo per ottenere questa stima è quello di duplicare ogni esperimento del disegno [18].

Occorre quindi verificare l'ipotesi nulla H_0 secondo la quale il rapporto $\frac{SS_{lof}}{f-p} / \frac{SS_{pe}}{n-f}$ è uguale a 1, cioè il *lack of fit* è paragonabile all'errore puramente sperimentale. Il modello risulta valido se $F < F_{crit}$; se $F > F_{crit}$ H_0 è rifiutata, ovvero esiste una significativa perdita di aderenza tra modello e dati [17].

Analisi statistica dei coefficienti

Lo scopo dell'analisi statistica dei coefficienti è capire se l'effetto b_k del fattore k è significativo o no; in altre parole, verificare se le risposte ai diversi livelli di k sono significativamente diverse. In questo caso, l'ipotesi nulla H_0 prevede $\beta_k=0$, mentre l'ipotesi alternativa H_1 prevede $\beta_k \neq 0$.

È possibile costruire un intervallo di confidenza (noto il t di Student) intorno a b_k in modo che ci sia una data probabilità che l'intervallo contenga il valore vero β_k . Per far questo occorre valutare la deviazione standard della stima del coefficiente b_k , chiamata errore standard s_{b_k} , la quale può essere calcolata dalla matrice di dispersione $(\mathbf{X}'\mathbf{X})^{-1}$ se è disponibile una stima della varianza sperimentale, σ_{pe}^2 .

$$\text{Var}(\mathbf{B}) = [\sigma_{pe}^2 (\mathbf{X}'\mathbf{X})^{-1}] \quad \sigma_{b_k}^2 = \sigma_{pe}^2 c^{kk}$$

dove c^{kk} è il coefficiente di varianza.

Calcolati b_k e s_{b_k} , se tale intervallo include il valore zero viene accettata l'ipotesi

nulla H_0 secondo la quale $\beta_k=0$. In caso contrario si accetta l'ipotesi alternativa H_1 che β_k sia significativamente diverso da zero [16].

In altre parole, un coefficiente b_k è significativo se il rapporto tra il coefficiente b_k e il suo errore standard s_{b_k} è maggiore del valore del t di Student tabulato ad un certo numero v di gradi di libertà e ad una data probabilità [18].

3.1.3. Criteri di qualità

I criteri di qualità delle matrici consentono di conoscere a priori, cioè prima di condurre gli esperimenti, con quale accuratezza saranno stimati i coefficienti e quindi di valutare la qualità delle matrici per stabilire quale sia più adatta allo scopo.

Criteri legati alla stima dei coefficienti del modello

In generale, le stime dei coefficienti devono essere accurate e non correlate tra loro. La stima b_i del coefficiente β_i è definita accurata se l'intervallo centrato su b_i , nel quale con un certa probabilità si trova il valore vero del coefficiente, è il più piccolo possibile; la varianza della stima $var(b_i)$ misura tale intervallo. Due stime b_i e b_j sono non correlate se sono calcolate in maniera indipendente.

È quindi possibile assegnare dei limiti di confidenza ad ogni coefficiente e prendere in considerazione tutti i fattori contemporaneamente [19], per cui i limiti della regione di confidenza di tutti i coefficienti stimati (*joint confidence region*) sono definiti da un iperellissoide nello spazio dei coefficienti. Tale iperellissoide di confidenza è centrato sul vettore \mathbf{B} , cioè sul vettore stima dei coefficienti b . L'iperellissoide si riduce ad un'ellisse nel caso che i coefficienti siano solo due. Le caratteristiche di questo iperellissoide, quali la sua orientazione ed estensione, possono essere usate come criteri per valutare la qualità della stima dei coefficienti. Tali aspetti dipendono a loro volta dalle proprietà della matrice di dispersione $(\mathbf{X}'\mathbf{X})^{-1}$.

Il *volume* dell'iperellissoide traduce l'accuratezza globale della stima dei coefficienti; per un data probabilità, più l'iperellissoide è piccolo, migliori sono le stime. La *forma* dell'iperellissoide è legata all'accuratezza con la quale è determinata ogni stima. Perché l'accuratezza nella stima dei coefficienti sia la stessa per tutti i coefficienti, la forma dell'iperellissoide deve essere quella di un'ipersfera. L'*orientazione* dell'iperellissoide indica il grado di correlazione fra le stime dei

coefficienti. Le stime dei coefficienti risultano indipendenti se gli assi principali dell'iperellissoide sono paralleli agli assi dei coefficienti [19].

In particolare, per quanto riguarda il volume dell'iperellissoide e quindi l'accuratezza nella stima dei coefficienti, questo è proporzionale alla radice del determinante della matrice di dispersione $D(\xi) = |(XX')^{-1}|$

$$V = a |(XX')^{-1}|^{1/2}$$

Per una coppia modello/matrice sperimentale data, questo criterio è definito come *criterio del determinante*: più il determinante della matrice di dispersione è piccolo, minore è il volume dell'iperellissoide e quindi più accurata è la stima dei coefficienti.

Il *D-optimal design* si basa sul criterio del volume dell'iperellissoide. La matrice sperimentale è detta *D-ottimale* se fra tutte le matrici del dominio è quella che conduce al più piccolo valore del determinante della matrice di dispersione. È quindi la matrice sperimentale per la quale l'ellissoide di confidenza dei coefficienti è il più piccolo possibile.

Criteri legati alla qualità previsionale del modello

Nel caso in cui lo sperimentatore ricerchi un buon modello previsionale, vale a dire un modello che permetta di predire il valore della risposta in qualsiasi punto del dominio sperimentale, l'obiettivo è ottenere la miglior qualità possibile nella previsione della risposta [12].

Tra i criteri di qualità che indicano la qualità previsionale del modello si trova il criterio *I-optimal*, basato sul calcolo della varianza media delle predizioni del modello nella regione sperimentale. L'*I-optimal design* dunque, minimizzando tale varianza media, produce una stima più precisa della risposta predetta. Questa a sua volta porta ad una migliore collocazione della risposta ottimale [20].

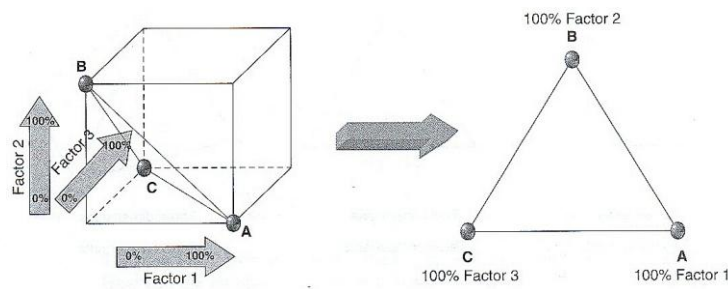
3.2. Disegno di miscele

Il disegno sperimentale di miscele viene utilizzato quando si assume che la risposta misurata dipenda dalle proporzioni relative dei componenti presenti nella miscela [21]. Tale disegno è uno speciale tipo di studio della superficie di risposta in cui i fattori sono i costituenti della miscela e la risposta è funzione delle percentuali di

ognuno di questi costituenti. In particolare, in campo farmaceutico questo approccio è adatto per diversi problemi di miscelazione (ad esempio, progettazione di una nuova formulazione) e permette sia la valutazione con il minimo numero di esperimenti degli effetti dei cambiamenti nella composizione delle miscele sia la scelta della composizione ottimale [21].

La caratteristica peculiare del disegno di miscele consiste nel fatto che non è possibile variare le percentuali dei singoli componenti in modo indipendente l'uno dall'altro, dato che la loro somma deve essere pari al 100%. Di conseguenza i fattori sono espressi come frazione dell'ammontare totale e i loro range sperimentali variano tra 0% e 100% [12,21].

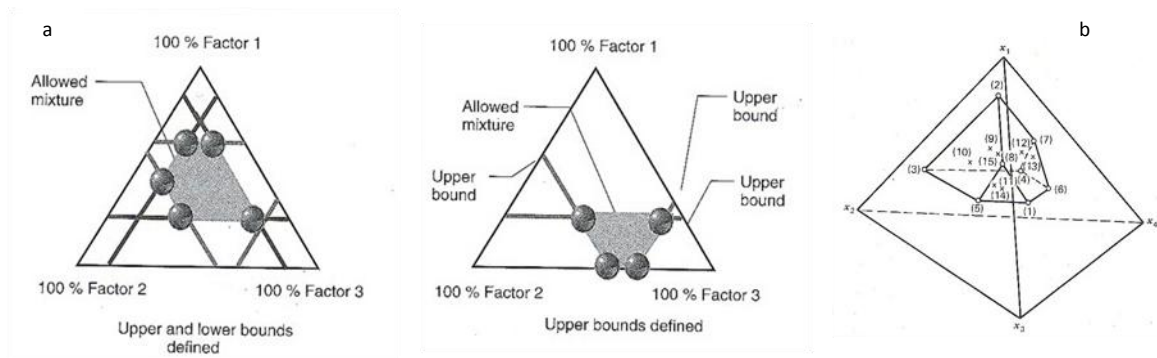
Ad esempio, una miscela a tre componenti viene rappresentata da un triangolo, sezione bidimensionale di uno spazio tridimensionale, a sua volta rappresentato da un cubo [22]. Il triangolo mostra appunto la regione permessa in cui la somma delle proporzioni dei tre componenti è pari al 100%. I punti del triangolo rappresentano le possibili miscele: i tre vertici corrispondono ai singoli componenti; i punti sul perimetro corrispondono a miscele binarie; i punti interni al triangolo corrispondono a miscele ternarie. Il centro del triangolo corrisponde ad una uguale proporzione di tutti i componenti [22].



Rappresentazione spaziale di una miscela a tre componenti.

Con l'aumento del numero dei componenti, aumenta la dimensionalità dello spazio; infatti, per una miscela di n componenti la dimensionalità dello spazio è pari a $n-1$. Quindi, se per tre componenti lo spazio delle miscele è un triangolo, per quattro componenti è un tetraedro. Queste figure geometriche sono definite *simplex* e corrispondono ai più semplici oggetti possibili in uno spazio di data dimensionalità [22].

Anche se teoricamente ogni componente può variare tra 0 e 100%, in pratica spesso vengono applicate delle limitazioni alle proporzioni dei costituenti della miscela per attenersi alle percentuali dei componenti effettivamente utilizzate. In generale, le limitazioni imposte delimitano, all'interno della figura definita dai componenti, una regione sperimentale rappresentata da un poliedro (spazio dei fattori limitato) [12,21].



Esempi di particolari situazioni riscontrabili nel caso di miscele costituite da tre componenti (a) e quattro componenti (b).

Nel disegno sperimentale di miscele vengono utilizzati specifici modelli matematici e specifiche matrici sperimentali [12,21].

I possibili modelli che correlano la risposta con i componenti sono diversi. Considerando per semplicità tre componenti, è possibile usare un modello di primo ordine in cui vengono valutati gli effetti dei costituenti puri:

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

Con un modello più complesso come quello di secondo ordine vengono valutate anche le miscele binarie:

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \varepsilon$$

È possibile scegliere un modello di ordine ancora superiore, normalmente di terzo ordine. La forma canonica del modello di terzo ordine completo è relativamente complessa e il numero di coefficienti da stimare è alto:

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \gamma_{12} x_1 x_2 (x_1 - x_2) + \gamma_{13} x_1 x_3 (x_1 - x_3) + \gamma_{23} x_2 x_3 (x_2 - x_3) + \beta_{123} x_1 x_2 x_3 + \varepsilon$$

Comunque, è possibile omettere alcuni termini ed ottenere un modello, detto cubico speciale, in cui oltre all'effetto dei componenti puri sono valutati gli effetti di miscele binarie e ternarie:

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_1 x_2 x_3 + \varepsilon$$

Questo modello per un sistema a tre componenti è costituito da sette termini. I primi tre rappresentano il modello lineare, ciascun coefficiente indica l'effetto che ogni singolo componente ha sulla risposta. I tre termini successivi rappresentano invece gli effetti delle miscele binarie e, insieme ai termini lineari, formano il modello quadratico. Infine è presente il termine relativo all'effetto della miscela ternaria, solitamente importante per sistemi che hanno i valori di massimo o di minimo all'interno del triangolo delle concentrazioni [23].

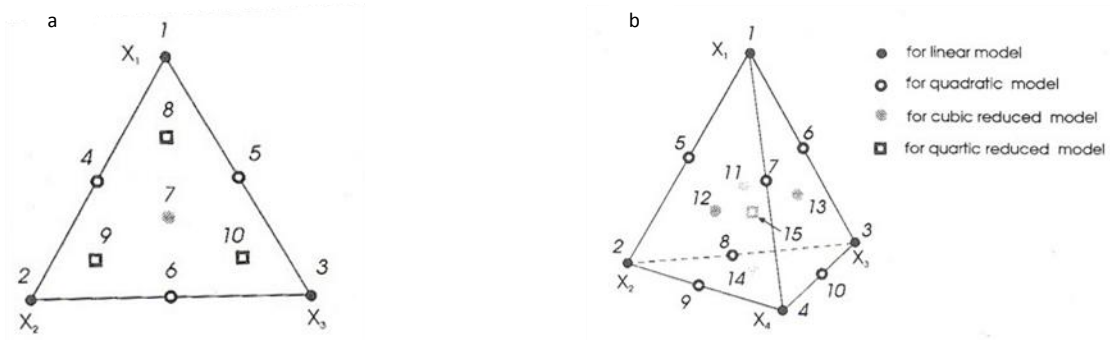
Per selezionare un modello che descriva bene il fenomeno in studio, la strategia comunemente seguita è il metodo *step-wise*, che consiste nel considerare inizialmente modelli più semplici e se necessario progressivamente modelli sempre più complessi. Dapprima si ipotizza quindi un modello di primo ordine che viene convalidato attraverso i cosiddetti punti test, esperimenti che in questo caso corrispondono a miscele binarie. Se le risposte misurate in questi punti non sono statisticamente diverse dalle risposte predette usando il modello ipotizzato, il modello è valido e rappresenta bene il fenomeno. In caso contrario, il modello non è valido e deve essere preso in considerazione un modello di secondo ordine. L'aggiunta dei punti test al disegno originario fornisce un disegno che risulta ottimale per stimare un modello di secondo ordine. Questo viene a sua volta convalidato utilizzando in genere come punto test una miscela ternaria con uguali proporzioni dei tre componenti. Se anche il modello di secondo ordine non rappresenta bene i dati, si può utilizzare un modello speciale cubico. La convalida di quest'ultimo prevede altre miscele ternarie. Gli esperimenti relativi ai punti test devono essere condotti in ordine randomizzato insieme con gli esperimenti del disegno in modo da ottenere una distribuzione casuale di eventuali piccoli e non noti errori sistematici [12].

Mentre la validità del modello viene quindi in generale valutata attraverso l'analisi dei residui usando i punti test, la significatività del modello viene valutata attraverso l'analisi della varianza [12]. Anche questa, come i modelli, mostra alcune differenze da quella per variabili indipendenti, dato che vi sono particolari limitazioni al modello e che i coefficienti non sono indipendenti [12].

Il disegno più utilizzato per stimare i coefficienti del modello è il disegno di Scheffé simplex centroide, che per k fattori definisce 2^{k-1} esperimenti. Nel caso di tre

componenti, questo può essere costruito sequenzialmente con tre esperimenti che corrispondono alle tre sostanze pure che servono alla stima del modello di primo ordine, tre miscele binarie (per il modello di secondo ordine), una miscela ternaria (per il modello cubico speciale), per un totale di 7 esperimenti a cui vengono aggiunti gli eventuali punti test.

Nella figura successiva è riportata la posizione degli esperimenti del disegno di Scheffé simplex centroide per tre componenti e per quattro componenti [12].



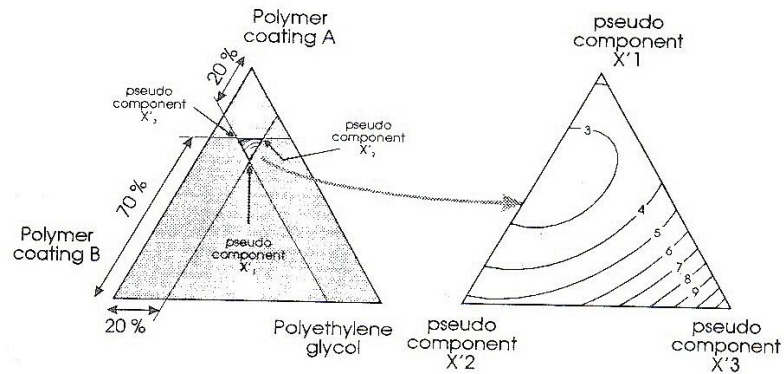
Rappresentazioni della posizione degli esperimenti per un disegno di Scheffé simplex centroide per 3 componenti (a) e per 4 componenti (b).

Se vi sono delle limitazioni nelle proporzioni dei componenti, le coordinate del poliedro originale definito dai componenti originali (x_i) devono essere ridefinite in termini di pseudocomponenti (x'_i), che sono combinazioni dei componenti originali. Alcuni vantaggi del loro uso consistono nel fatto che la costruzione dei disegni risulta più semplice e la qualità descrittiva (*fitting*) dei modelli risulta migliore che nel sistema dei componenti originali. Questa trasformazione è simile all'introduzione delle variabili codificate al posto delle variabili originali nel classico disegno sperimentale per fattori indipendenti [21]. La formula usata per trasformare i componenti in pseudocomponenti è la seguente:

$$x'_i = \frac{x_i - a_i}{1 - \sum a_i}$$

dove x'_i è il valore dello pseudocomponente, x_i è il valore del componente nella matrice sperimentale e a_i è il limite inferiore.

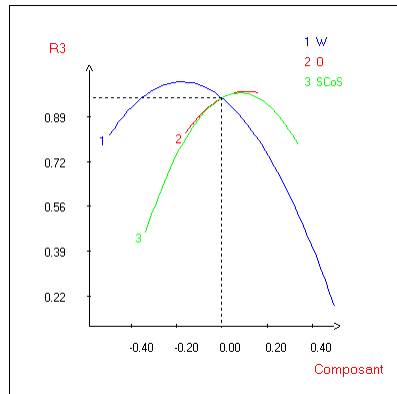
Nella figura successiva è riportato un esempio di trasformazione in pseudocomponenti. Una volta ottenuto un modello valido e significativo, questo può



Esempio di trasformazione da componenti a pseudocomponenti nel caso dell'ottimizzazione di un rivestimento per compresse.

essere usato in modo predittivo attraverso l'esame di superfici di risposta bidimensionali o tridimensionali. Da questi grafici, che coprono in modo completo lo spazio sperimentale della miscela, è possibile trovare la miscela ottimale per il problema considerato. In particolare, l'uso delle curve di isorisposta permette una interpretazione dei risultati molto semplice, come evidenziabile dalla figura sopra riportata.

Inoltre, per valutare meglio il contributo di ognuno dei componenti è possibile applicare il metodo della traccia della risposta [21]. Si tratta di un grafico in cui viene mostrato il cambiamento nella risposta stimata dovuto al cambiamento della percentuale di un singolo componente, mantenendo gli altri componenti ad un valore fissato. Una volta scelta la miscela di riferimento (normalmente il centroide della regione sperimentale), il grafico mostra la variazione della risposta stimata allontanandosi dalla miscela di riferimento lungo gli assi del componente considerato.



Metodo della traccia della risposta

3.3. Modelli misti e disegni sperimentali a matrice mista di variabili miscela-processo (MPV)

In certi casi i fenomeni in studio possono dipendere sia da variabili di miscela (*mixture components*, MC) che da variabili di processo (*process variables*, PV). L'ottimizzazione riguarda quindi sia MC che PV, con modelli misti di variabili di miscela e di processo. Usualmente, per studiare situazioni di questo tipo viene utilizzato un approccio che si articola in due fasi distinte [21]. Nella prima fase si effettua un disegno di miscele [21] per fissare la proporzione ottimale degli MC. Nella seconda fase, attraverso uno studio della superficie di risposta [12] vengono ottimizzati i fattori indipendenti. La maggiore limitazione di questo approccio a due fasi consiste nel fatto che non è in grado di fornire informazioni circa le eventuali interazioni tra i componenti di miscela e le variabili di processo. Per risolvere questo problema può venire utilizzata una matrice mista di variabili miscela-processo (*mixture-process variable*, MPV), che prevede la simultanea variazione di MC e PV. In questo modo vengono stimati modelli misti che mettono in relazione le prestazioni del sistema in studio sia con i componenti della miscela sia con le variabili di processo, e vengono individuate le eventuali interazioni tra MC e PV [21].

I modelli e i disegni sperimentali MPV sono in genere molto complessi, poiché comprendono un numero molto elevato di coefficienti da stimare e quindi di esperimenti da effettuare. Infatti, questo numero aumenta molto rapidamente con l'aumentare del

numero degli MC e delle PV, dato che modelli e disegni MPV sono spesso ottenuti sommando modelli di miscela a modelli per variabili di processo.

Considerando ad esempio tre MC e tre PV, come nel caso del problema che viene preso in esame nella presente tesi, e ipotizzando come modelli rispettivamente un modello cubico speciale e un modello quadratico, si avrebbero come punto di partenza i modelli seguenti.

Per gli MC, il modello di miscela speciale cubico è

$$y_{MC} = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_1 x_2 x_3 + \varepsilon \quad (1)$$

dove i termini β_i , β_{ij} , e β_{123} descrivono gli effetti lineari, quadratici e cubici speciali degli MC, e x_i rappresenta le proporzioni dei componenti (o loro trasformazioni in pseudocomponenti). Le proporzioni x_i soddisfano il requisito $x_1 + x_2 + x_3 = 1$.

Per le PV, il modello quadratico è

$$y_{PV} = \alpha_0 + \alpha_1 z_1 + \alpha_2 z_2 + \alpha_3 z_3 + \alpha_{12} z_1 z_2 + \alpha_{13} z_1 z_3 + \alpha_{23} z_2 z_3 + \alpha_{11} z_1^2 + \alpha_{22} z_2^2 + \alpha_{33} z_3^2 + \varepsilon \quad (2)$$

dove i termini α_i , α_{ij} , e α_{ii} descrivono gli effetti lineari, di interazione e quadratici delle PV.

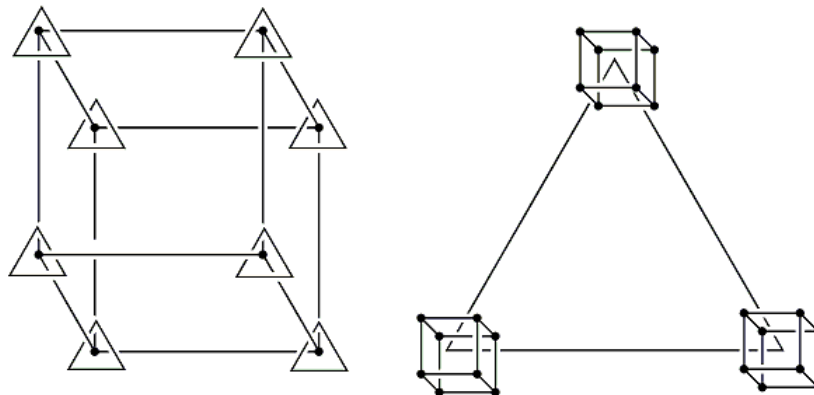
L'unione dei modelli (1) e (2) comporta la moltiplicazione di tutti i termini di un modello per tutti i termini dell'altro modello, ottenendo così un modello MPV (modello di miscela speciale cubico)x(modello quadratico PV) di questa forma:

$$\begin{aligned} y_{SC-Q} = & g_1^0 x_1 + g_2^0 x_2 + g_3^0 x_3 + g_{12}^0 x_1 x_2 + g_{13}^0 x_1 x_3 + g_{23}^0 x_2 x_3 + g_{123}^0 x_1 x_2 x_3 \\ & + (g_1^1 x_1 + g_2^1 x_2 + g_3^1 x_3 + g_{12}^1 x_1 x_2 + g_{13}^1 x_1 x_3 + g_{23}^1 x_2 x_3 + g_{123}^1 x_1 x_2 x_3) z_1 \\ & + (g_1^2 x_1 + g_2^2 x_2 + g_3^2 x_3 + g_{12}^2 x_1 x_2 + g_{13}^2 x_1 x_3 + g_{23}^2 x_2 x_3 + g_{123}^2 x_1 x_2 x_3) z_2 \\ & + (g_1^3 x_1 + g_2^3 x_2 + g_3^3 x_3 + g_{12}^3 x_1 x_2 + g_{13}^3 x_1 x_3 + g_{23}^3 x_2 x_3 + g_{123}^3 x_1 x_2 x_3) z_3 \\ & + (g_1^{12} x_1 + g_2^{12} x_2 + g_3^{12} x_3 + g_{12}^{12} x_1 x_2 + g_{13}^{12} x_1 x_3 + g_{23}^{12} x_2 x_3 + g_{123}^{12} x_1 x_2 x_3) z_1 z_2 \\ & + (g_1^{13} x_1 + g_2^{13} x_2 + g_3^{13} x_3 + g_{12}^{13} x_1 x_2 + g_{13}^{13} x_1 x_3 + g_{23}^{13} x_2 x_3 + g_{123}^{13} x_1 x_2 x_3) z_1 z_3 \\ & + (g_1^{23} x_1 + g_2^{23} x_2 + g_3^{23} x_3 + g_{12}^{23} x_1 x_2 + g_{13}^{23} x_1 x_3 + g_{23}^{23} x_2 x_3 + g_{123}^{23} x_1 x_2 x_3) z_2 z_3 \\ & + (g_1^{11} x_1 + g_2^{11} x_2 + g_3^{11} x_3 + g_{12}^{11} x_1 x_2 + g_{13}^{11} x_1 x_3 + g_{23}^{11} x_2 x_3 + g_{123}^{11} x_1 x_2 x_3) z_1^2 \\ & + (g_1^{22} x_1 + g_2^{22} x_2 + g_3^{22} x_3 + g_{12}^{22} x_1 x_2 + g_{13}^{22} x_1 x_3 + g_{23}^{22} x_2 x_3 + g_{123}^{22} x_1 x_2 x_3) z_2^2 \\ & + (g_1^{33} x_1 + g_2^{33} x_2 + g_3^{33} x_3 + g_{12}^{33} x_1 x_2 + g_{13}^{33} x_1 x_3 + g_{23}^{33} x_2 x_3 + g_{123}^{33} x_1 x_2 x_3) z_3^2 + \varepsilon \end{aligned} \quad (3)$$

In tutti i modelli (1), (2), (3) sopra riportati, il termine ε indica l'errore sperimentale. Si assume che gli errori siano statisticamente indipendenti e distribuiti con stessa media e varianza. Quando gli errori seguono una distribuzione normale (gaussiana), possono essere applicati molti metodi statistici di trattamento dei dati.

Il modello MPV (3) contiene un elevato numero di termini, pari a $7 \times 10 = 14$. Un disegno sperimentale che permetta di stimare un modello di questo tipo necessita di almeno 14 esperimenti più altri esperimenti per determinare l'errore sperimentale e quindi valutare la validità del modello in termini di aderenza ai dati (*lack of fit*).

Per stimare il modello (1) nel caso che alle proporzioni degli MC siano imposte delle limitazioni, come nello studio presentato in questa tesi, occorre utilizzare un disegno di miscele che consiste dei 4 vertici, dei 4 punti medi dei lati e del centroide della figura a quattro lati definita dai limiti superiori e inferiori degli MC, per un totale di 9 esperimenti. Per stimare il modello (2) occorre un disegno adatto per lo studio della superficie di risposta, ad esempio un *central composite design* [24], che comprende 8 vertici, 6 punti assiali, e il centro del cubo definito dai livelli -1 e +1 delle PV. L'incrocio di questi due disegni porta quindi a un disegno MPV a $9 \times 15 = 135$ punti, che comporta un numero di esperimenti estremamente elevato.



Un esempio di disegno MPV:

incrocio di un disegno fattoriale saturo per 3 fattori (2^3) con un disegno simplex per una miscela a tre componenti.

Un'alternativa per generare disegni MPV che contengano un numero minore di esperimenti è l'utilizzo dell'approccio "*optimal experimental design*" [20]. In questo caso il ricercatore individua il modello MPV più adeguato per approssimare le relazioni tra le variabili risposta e gli MC e le PV, escludendo interazioni di ordine elevato che si ritiene possano essere non significative. Poi, sono utilizzati *softwares* di disegno sperimentale Design-Expert [25], JMP [26], Minitab [27], che mediante un algoritmo appropriato generano un disegno ottimale. Possono essere utilizzati diversi criteri di ottimalità tra cui il *D*-ottimale e l'*I*-ottimale [20]. Il criterio *D*-ottimale si basa sulla

minimizzazione della varianza delle stime dei coefficienti del modello. Il criterio *I*-ottimale, utilizzato in questo studio, minimizza la varianza della predizione del modello nella regione sperimentale, producendo una stima più precisa della risposta predetta [20].

Il numero dei punti del disegno MPV deve essere comunque almeno pari al numero dei termini del modello MPV scelto per poi applicare il criterio di qualità selezionato. Inoltre, il disegno dovrebbe avere almeno 5 repliche per stimare l'errore sperimentale e 10 punti addizionali (non replicati) per valutare l'aderenza del modello ai dati. La strategia MPV si articola quindi in queste fasi principali [28]:

- 1) selezionare un modello MPV che approssimi adeguatamente le relazioni tra le risposte e gli MC e le PV;
- 2) generare un disegno sperimentale MPV utile per stimare il modello MPV e per valutare la validità del modello stesso;
- 3) usare i modelli MPV per trovare equazioni che individuino lo spazio del disegno e quindi determinare le migliori condizioni che portano a ottimizzare le risposte.

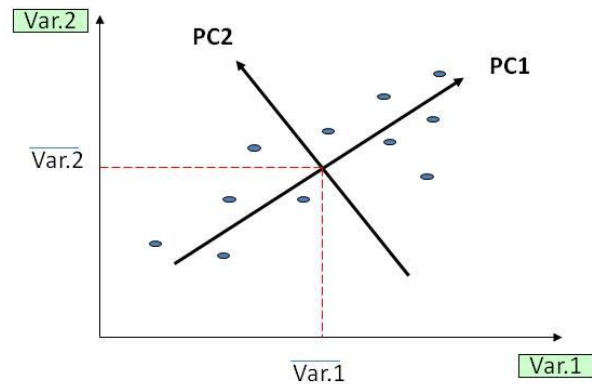
3.4. Analisi delle componenti principali

L'Analisi delle Componenti Principali (*Principal Component Analysis, PCA*) è una tecnica di analisi multivariata proposta da Karl Pearson nel 1901 di fondamentale importanza per l'esplorazione dei dati. In generale, le variabili che descrivono i dati sono trasformate in nuove variabili, chiamate componenti principali (PC). Le PC sono delle combinazioni lineari delle variabili originali e sono tra di loro ortogonali [1].

Mediante questa tecnica è possibile:

- Valutare le correlazioni tra le variabili e la loro rilevanza
- Visualizzare gli oggetti
- Sintetizzare la descrizione dei dati
- Ridurre la dimensionalità dei dati
- Ricercare proprietà principali

- Definire un modello di rappresentazione dei dati in uno spazio ortogonale

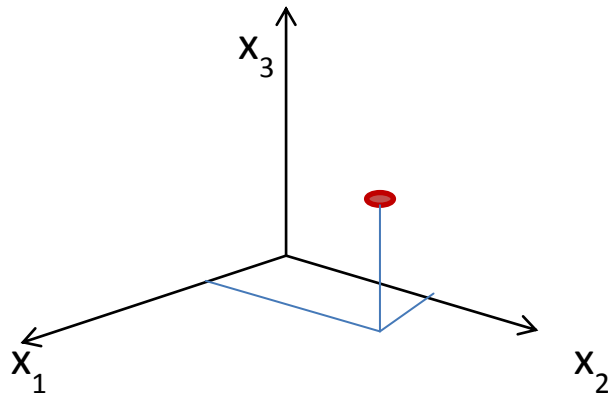


La PCA consiste in un processo di rotazione dei dati originali definiti da una matrice \mathbf{X} di dimensione $m \times n$, effettuato in modo che il primo nuovo asse sia orientato nella direzione di massima varianza dei dati, il secondo sia perpendicolare al primo e sia nella direzione della successiva massima varianza dei dati e così di seguito per tutti i p nuovi assi.

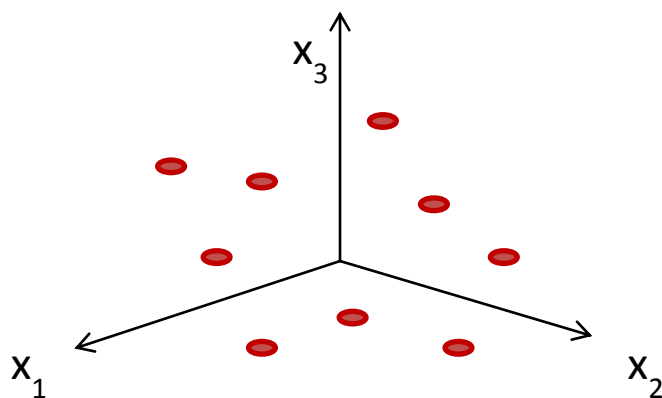
La figura mostra un esempio a due variabili: la prima componente principale (PC1) è nella direzione di massima varianza e la sua origine è situata nel valore medio della variabile; la varianza residua viene rappresentata dalla seconda componente principale (PC2), nella direzione perpendicolare alla prima componente.

In molti casi, come nella PCA, i metodi chemiometrici richiedono che i dati originali siano pretrattati. In particolare, quando si ricercano informazioni sulle relazioni tra le variabili, obiettivo tipico dell'analisi multivariata, è importante massimizzare la confrontabilità tra le variabili. I più comuni tipi di scalatura (*scaling*) sono la *Centratura* e l' *Autoscalatura*. La centratura consiste nel centrare i dati rispetto al valore medio di ciascuna variabile (centroide); l'autoscalatura consiste invece in una centratura seguita da una normalizzazione a varianza unitaria.

Nell'analisi multivariata, ogni variabile m definisce un asse e quindi le m variabili definiscono uno spazio M -dimensionale. Ogni oggetto è individuato in questo spazio da un punto, le cui coordinate sono date dal valore delle variabili corrispondenti. Nella figura seguente viene illustrato un esempio del caso di un campione descritto da tre variabili.



Gli n oggetti dello spazio delle m variabili definiscono un insieme di punti in uno spazio multidimensionale. Se gli oggetti sono simili tra di loro, questo insieme appare compatto, mentre se ci sono diversità si osserva una dispersione degli oggetti. Nella figura seguente viene mostrato l'esempio di un set di dati descritti da tre variabili.



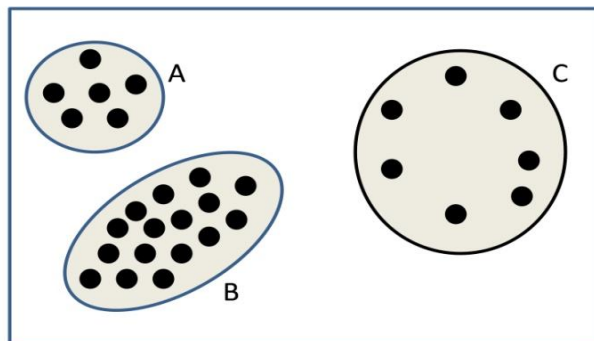
Dal punto di vista matematico la PCA ha lo scopo di diminuire il numero di variabili originali operando una rotazione ortogonale di assi cartesiani lungo la linea di

maggior variazione dei punti plottati nel piano secondo le proprie variabili naturali. Questa rotazione permette, partendo da un gran numero di variabili più o meno correlate, di ottenere un numero ristretto di variabili non correlate. La PCA dunque ricerca il modello di minima dimensionalità capace di descrivere in modo soddisfacente la struttura dei dati. Partendo dal dataset, viene ricercato il vettore, cioè la componente o autovettore, la cui direzione sia tale che la proiezione su di esso dei punti nello spazio M-dimensionale descriva nel miglior modo possibile la struttura dei dati nello spazio originale. Viene quindi definita la prima componente PC1 in modo tale che spieghi lungo la propria direzione la massima dispersione e dunque la massima varianza esistente fra i dati e intorno a cui si abbia la minima dispersione. Una volta individuato questo nuovo asse viene definita la PC2, ossia l'asse perpendicolare alla PC1. Le due rette così definite individuano un modello planare e la ricerca delle componenti successive può procedere ulteriormente fino all'esaurimento di tutta la varianza contenuta nei dati. L'informazione rilevante tuttavia è contenuta generalmente nelle prime 2 o 3 componenti principali ed è pertanto individuabile e visualizzabile.

Ogni autovettore viene ottenuto come combinazione lineare delle variabili originali, nel senso che ogni variabile dà il suo contributo più o meno grande per la costruzione di ciascun nuovo asse. Tale contributo è espresso attraverso un coefficiente detto loading. Gli scores invece rappresentano le nuove coordinate degli oggetti nello spazio delle componenti principali. La PCA permette rappresentazioni grafiche molto efficaci sia dei soli oggetti (*scores plot*) sia delle sole variabili (*loading plot*) sia di oggetti e variabili contemporaneamente (*biplot*). Il grafico dei *loadings* consente di analizzare il ruolo di ciascuna variabile nelle diverse componenti e la loro importanza. Il grafico degli *scores* consente invece di analizzare il comportamento degli oggetti nelle diverse componenti e la loro similarità. Infine il grafico *biplot* consente di rappresentare contemporaneamente oggetti e variabili al fine di poter valutare le relazioni che tra di essi intercorrono [29].

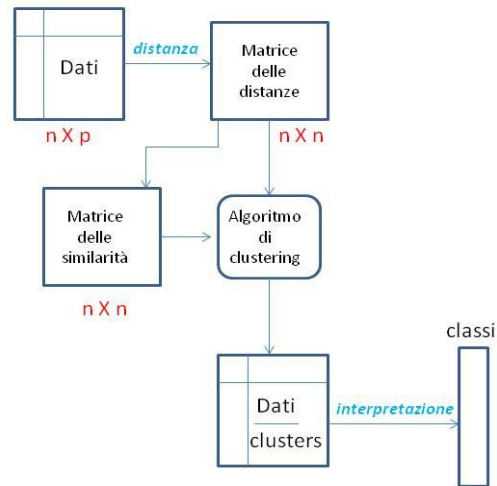
3.5. Analisi dei Cluster

L'analisi dei Cluster o ricerca di gruppi (CA), è uno dei metodi di classificazione che ricerca nei dati strutture non casuali. Questo obiettivo è comunemente perseguito associando il concetto di struttura non casuale a quello di raggruppamento e ricercando la presenza di raggruppamenti nello spazio dei dati, in contrapposizione all'ipotesi completa di omogeneità dei dati. I metodi di cluster analysis forniscono risposte sulla presenza di raggruppamenti (clusters) utilizzando il concetto di similarità [29]. Nella cluster analysis le classi non sono note a priori: al contrario la ricerca dell'esistenza di gruppi è continua. Possiamo dividere in due grandi categorie questi metodi: metodi gerarchici e metodi non-gerarchici. I primi includono metodi come *single linkage*, *average linkage*, *complete linkage*, mentre gli altri includono metodi le cui strategie sono molto più differenziate tra loro quali i *k-means*. I cluster che ciascun metodo individua sono caratterizzati dalla loro posizione nello spazio da un centroide, definito come il vettore delle medie delle variabili calcolate per gli oggetti assegnati al cluster, o da un centrotipo, definito come l'oggetto più rappresentativo tra gli oggetti assegnati al cluster. Ciascun cluster è caratterizzato dalla propria dimensionalità (numero degli oggetti che contiene), dalla compattezza degli oggetti attorno al centroide (definita dalla deviazione standard rispetto al centroide), dalla propria forma nello spazio dimensionale. Nella figura qui sotto, sono mostrati tre clusters di diverso tipo.



Il primo A è un cluster sferico e compatto; il secondo B è un cluster compatto dalla forma allungata ed il terzo C è un cluster costituito da alcuni oggetti sparsi.

I metodi di cluster analysis in generale si sviluppano secondo una procedura che può essere brevemente schematizzata in pochi passaggi: dopo aver selezionato un tipo di distanza, viene calcolata la matrice delle distanze e da questa la matrice di similarità; applicando l’algoritmo di clustering si ottiene la partizione finale degli oggetti. La possibile interpretazione di classi porta alla identificazione di classi (vedi figura).



I metodi gerarchici a loro volta si suddividono in due grandi categorie: i metodi gerarchici divisionali e i metodi gerarchici agglomerativi [Vandeginste 1998, 1].

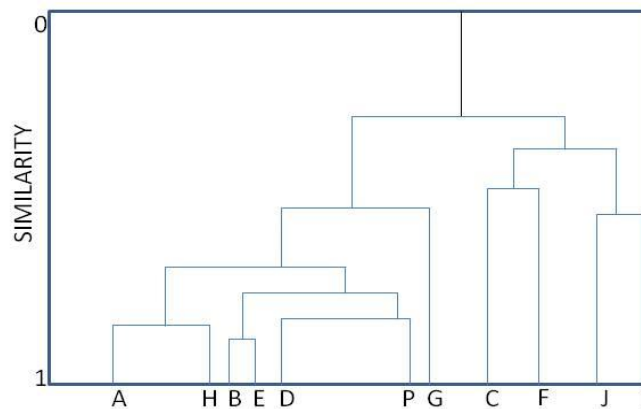
Il primo gruppo di metodi è poco usato e si basa su strategie che partono da un insieme che comprende tutti i dati iniziali e separano via via i campioni che differiscono maggiormente dagli altri. I metodi gerarchici agglomerativi al contrario, sono i più utilizzati e partono da un numero di cluster pari al numero di oggetti, procedendo alla loro fusione in clusters di dimensione sempre maggiore. Sono noti con i seguenti nomi a ognuno dei quali corrisponde un algoritmo che permette il calcolo della matrice di similarità:

1. *Weighted Average Linkage*
2. *Unweighted Average Linkage*
3. *Single Linkage*
4. *Complete Linkage*

5. *Centroid Linkage*
6. *Median Linkage*
7. *Ward method*

Calcolando la similarità tra i nuovi cluster e i restanti si ottiene un grafico chiamato **dendrogramma** che permette un'analisi visiva altamente informativa della gerarchia delle similarità tra gli oggetti considerati [31]. Un esempio grafico del risultato ottenuto applicando un metodo di clustering gerarchico agglomerativo è riportato in figura. Sull'asse delle ascisse sono riportati i diversi campioni rappresentati dalle lettere dell'alfabeto, ciascuno descritto da un certo numero di variabili.

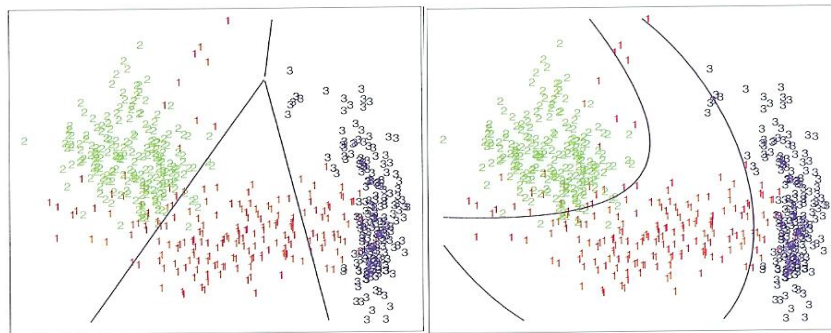
Sull'asse delle ordinate sono riportati i valori di similarità. Esaminando il grafico dal basso verso l'alto, si possono osservare le coppie di campioni che sono più simili tra di loro: i campioni B ed E sono i più simili tra di loro perché si uniscono per primi, successivamente si uniscono A con H e D con P e così via.



3.6. Metodi di classificazione multivariata

Analisi Discriminante Lineare e Quadratica

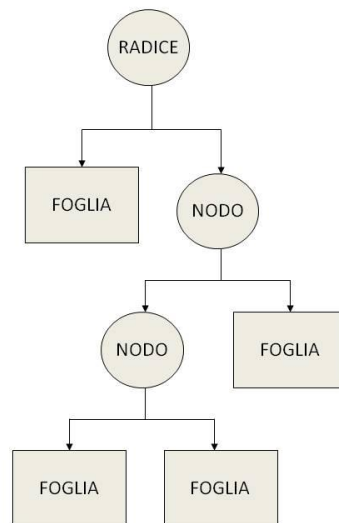
L'analisi discriminante lineare (LDA) introdotta da Fisher nel 1936 [32] presuppone che le variabili siano state misurate su osservazioni appartenenti a due o più gruppi. Come l'analisi discriminante quadratica (QDA), entrambi sono metodi parametrici che richiedono che i dati siano normalmente distribuiti e ogni classe abbia una propria matrice di covarianza. Lo scopo dell'analisi è quello di trovare una o più combinazioni lineari o quadratiche dei parametri che permettano di discriminare in maniera ottimale tra i vari gruppi: LDA classifica i campioni definendo un "confine" lineare ottimale tra le classi rappresentato da una retta; QDA ha una funzione di tipo quadratico rappresentata da una curva.



Tuttavia questi metodi hanno qualche svantaggio. Entrambi i metodi infatti possono essere utilizzati se e solo se le classi sono ampiamente omogenee. LDA non può essere usato se il numero di variabili n è maggiore del numero dei campioni m ; QDA invece, richiede un numero di variabili minore del numero di oggetti nella classe più piccola (m_{smallest}) [30]. Questo problema può essere risolto riducendo il numero delle variabili con la PCA prima dell'LDA e QDA oppure usando la stepwise regression, funzione capace di selezionare le variabili più importanti rispetto al vettore risposta y . In questo modo entrambi i metodi sono in grado di indicare le variabili responsabili della classificazione.

Il metodo CART

Il metodo CART (Classification and Regression Tree) è il metodo di classificazione ad albero più noto introdotto nel 1984 da Breiman [33, 28]. Questo metodo si basa sulla costruzione di una sequenza di partizioni binarie dei dati (*binary split*) in grado di creare un albero decisionale come regola di classificazione. Gli alberi decisionali sono costituiti da una radice (nodo superiore), ove tutti gli oggetti sono insieme; da nodi (punti intermedi dell'albero), ove gli oggetti sono provvisoriamente collocati durante la procedura di classificazione e da foglie (i nodi terminali dell'albero), a ciascuna delle quali è associata una classe e ove gli oggetti sono collocati al termine della sequenza decisionale (vedi figura).



Secondo questo metodo, ciascun nodo k dell'albero è caratterizzato da una singola variabile $j(k)$ e da un valore soglia per questa variabile $t_j(k)$. Per ogni i -mo oggetto, questo valore consente di rispondere alla domanda :

$$x_{ij(k)} \leq t_j(k) ?$$

Se la risposta è sì, il campione viene classificato nel ramo di sinistra dell'albero; in caso contrario, viene classificato nel ramo di destra. Partendo dalla radice, ogni oggetto viene classificato percorrendo sequenzialmente l'albero decisionale e confrontando i valori soglia con i corrispondenti valori che le variabili assumono per l'oggetto considerato. Per ogni nodo, viene selezionata la variabile che fornisce la

migliore separazione dei dati , utilizzando tre principali criteri di “divisione” (*splitting criteria*) quali Gini, Twoing, Deviance. La struttura ottimale dell’albero viene determinata mediante una procedura di validazione. L’albero decisionale CART ha quindi il vantaggio di essere molto semplice da costruire e non necessita una precedente scalatura dei dati.

4. NMR e Tecniche Computazionali di Simulazione

4.1. Principi generali

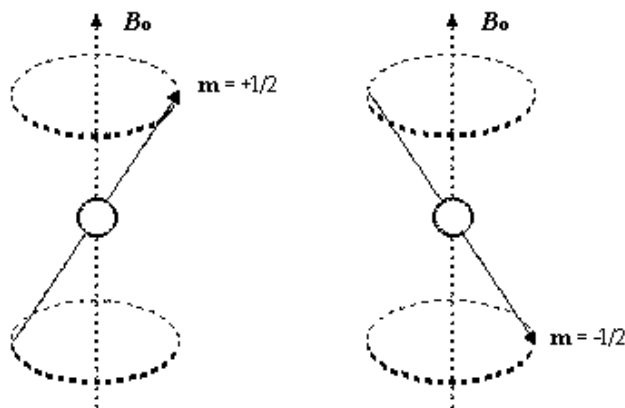
La Spettroscopia di Risonanza Magnetica Nucleare (NMR) è una tecnica analitica molto potente che permette di ottenere dettagliate informazioni sulla struttura molecolare dei composti in esame. La spettroscopia NMR misura l'assorbimento di radiazione elettromagnetica in molecole immerse in un forte campo magnetico. Questo assorbimento avviene ad opera dei nuclei di particolari atomi (tipicamente ^1H o ^{13}C) [34]. Quindi con l’NMR si esaminano direttamente i nuclei atomici e non gli elettroni: ogni informazione sull’intorno chimico viene dedotta osservando il comportamento dei nuclei atomici. Sono osservabili all’ NMR solo i nuclei che hanno un momento magnetico nucleare di spin e che quindi si comportano come l’ago di una bussola che si può orientare in un campo magnetico applicato. Il momento magnetico nucleare di spin \mathbf{m} è dato dalla relazione:

$$\mathbf{m} = g\hbar/2p$$

dove g è il rapporto giromagnetico, I è il numero quantico di spin nucleare, h è la costante di Plank. Lo spin nucleare viene prodotto dalle particelle che costituiscono il nucleo, protoni e neutroni. Queste si comportano come se fossero in rotazione (spin) attorno al loro asse e hanno spin $1/2$. In molti atomi (come nel ^{12}C) gli spin sono tutti appaiati, uno in opposizione all’altro e quindi si annullano reciprocamente e il nucleo atomico ha uno spin risultante I uguale a zero. In alcuni atomi, come nel caso di ^1H e ^{13}C , il nucleo possiede uno spin risultante I diverso da zero. Le regole per determinare lo spin nucleare si possono così riassumere:

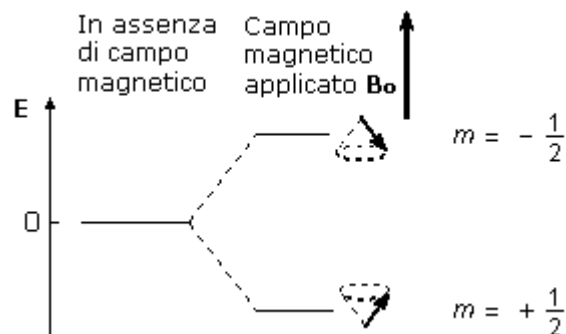
- 1) Se i protoni e i neutroni sono entrambi pari, allora il nucleo ha spin zero.
- 2) Se i protoni e i neutroni sono gli uni pari e gli altri dispari, allora il nucleo ha spin semi intero ($1/2, 3/2, 5/2, \dots$).
- 3) Se i protoni e i neutroni sono entrambi dispari, allora il nucleo ha spin intero ($1, 2, \dots$).

Rientrano nel primo caso ^{12}C e ^{16}O che hanno $I = 0$ e quindi non hanno momento magnetico di spin e non sono osservabili all'NMR. Rientrano nel secondo caso ^1H , ^{19}F , ^{13}C , ^{31}P e ^{15}N che hanno $I = 1/2$ e quindi sono osservabili all'NMR. Rientrano nel terzo caso ^2H e ^{14}N che hanno $I = 1$ e quindi anche questi sono osservabili all'NMR. Quando un nucleo dotato di spin viene immerso in un campo magnetico, il nucleo, come l'ago di una bussola, è sottoposto ad una coppia di forze che lo fanno ruotare per allinearlo con il campo magnetico esterno. Le possibili orientazioni che il nucleo può assumere in un campo magnetico sono governate dal numero quantico m che può assumere i valori da $-I$ a $+I$ (con incrementi di una unità) e quindi può assumere $2I+1$ valori diversi. Nel caso di un nucleo con spin $1/2$ (come ^1H) ci sono 2 possibili orientazioni: una con $m = -1/2$ e una con $m = +1/2$.



Come si vede nella figura precedente, il momento magnetico nucleare m non è perfettamente allineato col campo, ma continua ad oscillare attorno al campo magnetico applicato B_0 compiendo un moto di precessione simile a quello di una trottola. I due stati possibili del nucleo non hanno la stessa energia, ma lo stato con $m = +1/2$

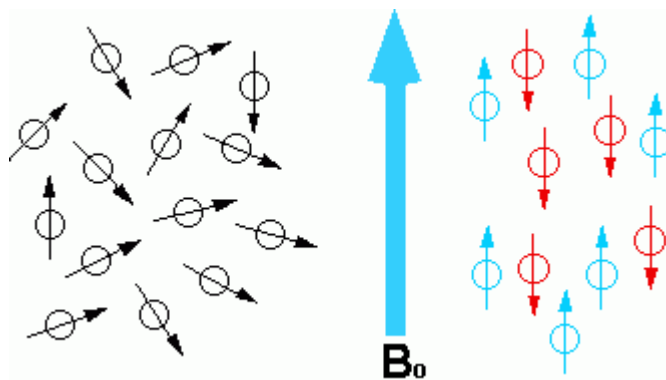
(allineato con il campo) si trova ad energia leggermente più bassa di quello con $m = -1/2$ (opposto al campo)



Il moto di precessione dei momenti magnetici nucleari avviene con una frequenza proporzionale alla differenza di energia tra i due livelli detta frequenza di Larmor data dalla formula:

$$\mathbf{n} = g \mathbf{B}_0 / 2\pi \text{ (in Hertz)}$$

dove g è il rapporto giromagnetico che dipende dal nucleo in esame. All'aumentare del campo applicato B_0 , aumenta la frequenza di Larmor e quindi la differenza di energia tra i livelli. Poiché la differenza di energia tra i due livelli è piccolissima, dell'ordine di $9 \cdot 10^{-6} \text{ kcal mol}^{-1}$, la popolazione di nuclei nei due stati è praticamente identica con una piccolissima prevalenza per lo stato a bassa energia allineato con il campo. A temperatura ambiente, in un campo magnetico di 1.41 Tesla (60 MHz) abbiamo un nucleo in più nello stato di bassa energia ogni duecentomila nuclei, in un campo di 7.05 Tesla (300MHz) abbiamo 5 nuclei in più nello stato di bassa energia ogni duecentomila nuclei.



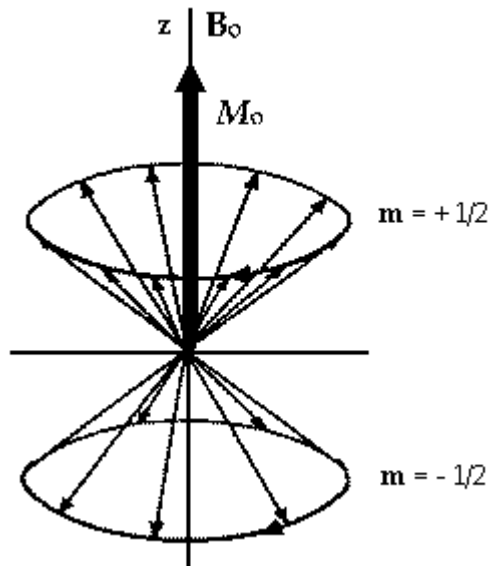
Eppure sono proprio questi pochi nuclei in eccesso allineati con il campo magnetico B_0 quelli che permettono di generare il segnale NMR. Si noti che con un campo magnetico più intenso (7.05 T) c'è una maggiore differenza di energia tra i livelli e quindi una maggiore frequenza di Larmor (300 MHz) e un maggior numero di nuclei in più allineati con il campo (5 su duecentomila) e quindi una maggiore sensibilità nell'analisi NMR [34].

Assorbimento di energia

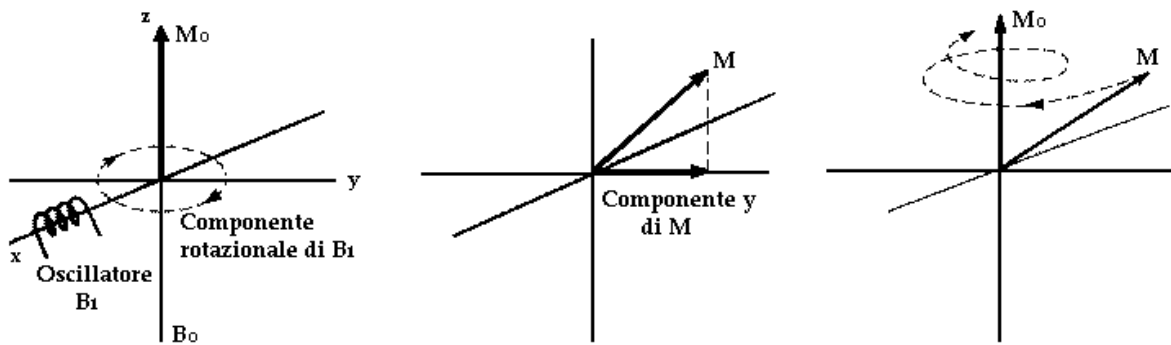
Se il campione viene irradiato con una radiazione elettromagnetica di frequenza uguale alla frequenza di Larmor, ci sarà una interazione della componente magnetica della radiazione con i momenti magnetici nucleari (anche questi oscillanti alla frequenza di Larmor). L'energia della radiazione potrà così essere trasferita ai nuclei. Ogni assorbimento di radiazione comporta un cambiamento di orientazione dello spin nucleare che ruoterà da allineato con il campo ad opposto al campo. Quando si verifica questa transizione di spin, si dice che i nuclei sono in risonanza con la radiazione applicata, da qui il nome di Risonanza Magnetica Nucleare, NMR. Anche nello stato di massima eccitazione, la popolazione dei nuclei nei due livelli permessi sarà circa uguale, data la piccolissima differenza di energia tra i livelli. Lo stato eccitato sarà solo leggermente più popolato di quello fondamentale con un rapporto di popolazioni invertito, cioè, in un campo di 7.05 Tesla (300MHz) ci potranno essere 5 nuclei in più nello stato di alta energia ogni duecentomila nuclei.

Produzione del segnale NMR

Il segnale NMR negli strumenti più moderni viene generato con il metodo ad impulso. Con questa tecnica tutti i nuclei di una specie vengono eccitati contemporaneamente da un impulso di radiofrequenza che contiene tutto l'intervallo di frequenze necessario. Per capire come l'impulso di radiofrequenza interagisce con i nuclei in esame dobbiamo introdurre un vettore chiamato Magnetizzazione Macroscopica M_0 , che è la risultante di tutti i momenti magnetici nucleari. Dato che esiste un leggero eccesso di nuclei che si trovano allineati col campo magnetico B_0 , M_0 sarà piccolo e allineato col campo. Indichiamo questa direzione come asse z.



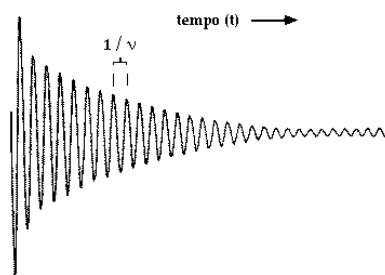
Se ora il campione viene irradiato lungo l'asse x con un impulso di radiofrequenza che contiene anche la frequenza ν di Larmor dei nuclei in esame (ad es 1H), i nuclei assorbiranno energia e subiranno una transizione di spin. A livello macroscopico si osserva che il vettore Magnetizzazione Macroscopica M_0 ruota allontanandosi dall'asse z per avvicinarsi al piano xy iniziando un moto di precessione attorno all'asse z.



A questo punto l'impulso di radiofrequenza cessa ed entra in funzione un circuito ricevente che ha lo scopo di misurare l'oscillazione della componente y del vettore M. E' un po' come se, dopo aver gridato di fronte a una parete rocciosa, restassimo in silenzio per ascoltare l'eco. L'energia assorbita dai nuclei viene lentamente ceduta agli atomi vicini a causa di fenomeni detti di rilassamento e il vettore M, compiendo delle spirali di precessione attorno all'asse z, si riporta al valore iniziale M_0 , posizione per la quale la componente M vale zero.

La durata $t(p)$ dell'impulso di radiofrequenza deve essere determinata con precisione per produrre un forte segnale NMR. Se $t(p)$ è tale da piegare M_0 di 90° , allora si produrrà il vettore M_y massimo. In genere $t(p)$ è di alcuni microsecondi. Il segnale raccolto è un segnale oscillante con frequenza ν , la frequenza di Larmor del nucleo in esame, che si smorza nel tempo e che viene detto FID (Free Induction Decay), libero decadimento dell'induzione.

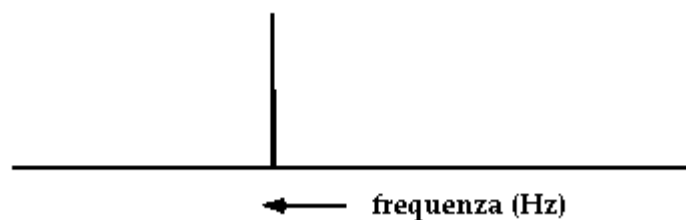
In figura è illustrato il FID del CH_3I per l'atomo di idrogeno.



Dato che nella molecola CH_3I gli idrogeni sono equivalenti, avranno tutti la stessa frequenza di risonanza. Nel grafico questa frequenza è facilmente individuabile misurando la lunghezza d'onda (la distanza tra due creste successive) e calcolandone il reciproco, secondo la formula:

$$\nu = 1/\lambda$$

dove ν è la frequenza e λ è la lunghezza d'onda. Si ottiene così il seguente grafico in funzione delle frequenze, chiamato spettro NMR, che mostra la frequenza assorbita dagli atomi di idrogeno nella molecola CH_3I :



Se il campione contiene nuclei con differenti frequenze di risonanza, questi vengono tutti eccitati contemporaneamente dall'impulso di radiofrequenza e quindi il segnale raccolto sarà una curva complessa, chiamata interferogramma, data dalla combinazione di più FID, uno per ogni frequenza assorbita dai nuclei. Per poter risalire alle singole frequenze che combinandosi tra loro hanno generato il tracciato complesso, è necessario applicare una procedura matematica detta Trasformata di Fourier che permette di passare dal grafico in funzione del tempo, il FID, al grafico in funzione delle frequenze, lo spettro NMR.

Spostamento Chimico

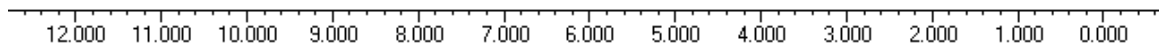
Dato che il dipolo magnetico μ di un certo nucleo è una costante, si potrebbe prevedere che tutti i nuclei di un certo tipo debbano risuonare esattamente alla stessa frequenza in un determinato campo magnetico applicato. Se le cose stessero così, la spettroscopia NMR sarebbe del tutto inutile perchè nello spettro avremmo un unico segnale.

Per fortuna, nelle molecole reali, questo fatto accade molto raramente. Infatti il segnale NMR di un certo nucleo risulta spostato nello spettro a frequenze più alte o più basse a seconda del suo intorno chimico. Questo accade perché gli elettroni sono particelle cariche e come tali, quando vengono immersi in un campo magnetico B_0 , si muovono e generano un piccolo campo magnetico opposto a quello molto più forte applicato. Questo campo magnetico secondario scherma il nucleo dalla piena forza del campo applicato.

Tanto più grande è la densità elettronica, tanto maggiore sarà questa schermatura, così i nuclei che si trovano in un intorno ricco di elettroni, sentiranno un campo magnetico più basso e quindi subiranno la transizione ad una frequenza applicata più bassa rispetto ai nuclei situati in intorni poveri di elettroni. Lo spostamento risultante nel segnale NMR per un dato nucleo è detto spostamento chimico e, in generale, protoni o carboni adiacenti ad atomi elettronegativi risultano deschermati quindi sentono un campo magnetico applicato più intenso e subiscono transizione a frequenze maggiori e, nello spettro, si trovano ad uno spostamento chimico più alto (vedi figura).

H deschermati
sentono un campo magnetico più forte
subiscono transizione a frequenze maggiori
hanno spostamenti chimici maggiori

H più schermati
sentono un campo magnetico più debole
subiscono transizione a frequenze minori
hanno spostamenti chimici minori

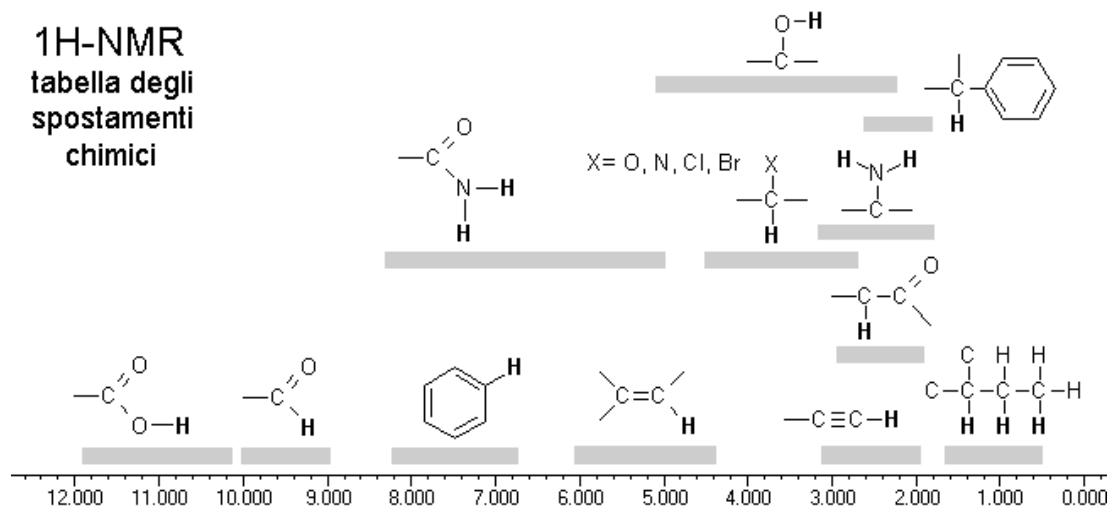


Lo spostamento chimico assoluto risulta tanto maggiore quanto più grande è il campo applicato, quindi strumenti diversi fornirebbero spostamenti chimici assoluti diversi in rapporto alla intensità del loro campo magnetico. Per questo si preferisce utilizzare lo spostamento chimico relativo δ definito dall'equazione:

$$\delta = \frac{\text{spostamento osservato rispetto al } (\text{CH}_3)_4\text{Si in Hz}}{\text{frequenza dell'oscillatore in MHz}} \text{ ppm}$$

In questo modo lo spostamento chimico δ , misurato in ppm (Hz/MHz), è indipendente dal campo applicato. Per convenzione, sia per $^1\text{H-NMR}$ che per $^{13}\text{C-NMR}$, lo zero della scala δ è attribuito all'assorbimento all'NMR del tetrametilsilano $(\text{CH}_3)_4\text{Si}$ nel quale i carboni e i protoni sono più fortemente schermati di quanto si osserva nella maggior parte delle altre molecole organiche. Per $^1\text{H-NMR}$, la scala δ generalmente si estende da 0 a 12 ppm; mentre per $^{13}\text{C-NMR}$ è molto più grande e copre l'intervallo 0-220 ppm. Nella scala δ del $^1\text{H-NMR}$, i semplici protoni idrocarburici tendono ad assorbire nella regione $\delta = 0.5-1.5$, i protoni sul carbonio adiacente al carbonile sono spostati a $\delta = 2$, gli atomi elettronegativi (ossigeno o alogeni) spostano i protoni a $\delta = 3-4$, i protoni vinilici sono spostati a $\delta = 5-6$, i protoni aromatici a $\delta = 7-8$, i protoni aldeidici a $\delta = 9-10$, ed i protoni degli acidi carbossilici sono i più spostati e cadono nell'intervallo 10-12 ppm.

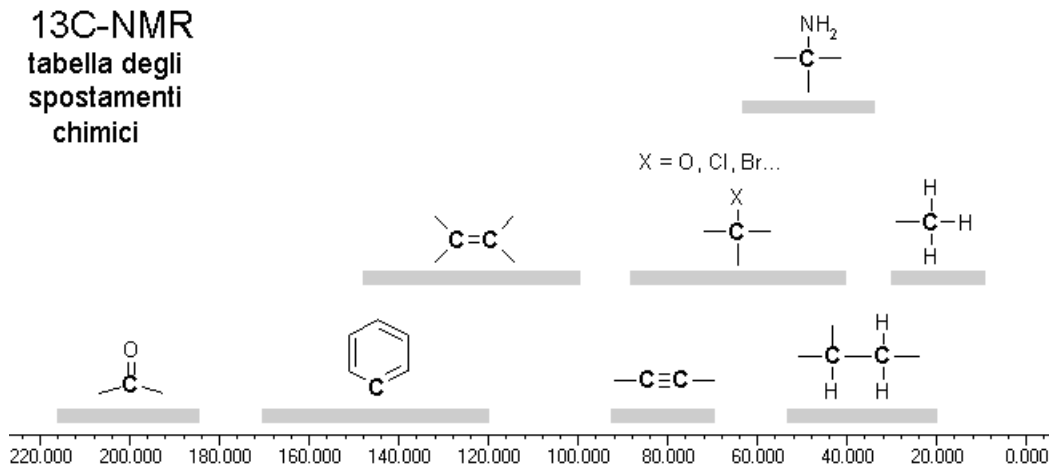
¹H-NMR
tabella degli
spostamenti
chimici



Spostamento Chimico d, ppm

Nella scala d del ¹³C-NMR, i semplici carboni dei metili CH₃ tendono ad assorbire nella regione d = 15-30, i semplici carboni dei metileni CH₂ sono spostati a d = 20-50, gli atomi elettronegativi (ossigeno o alogeni) spostano i carboni legati a d = 40-80, i carboni degli alchini sono spostati a d = 70-90, i carboni degli alcheni a d = 100-150, i carboni aromatici a d = 120-170, infine i carboni più fortemente spostati sono quelli dei carbonili, con valori di d = 180-220.

¹³C-NMR
tabella degli
spostamenti
chimici



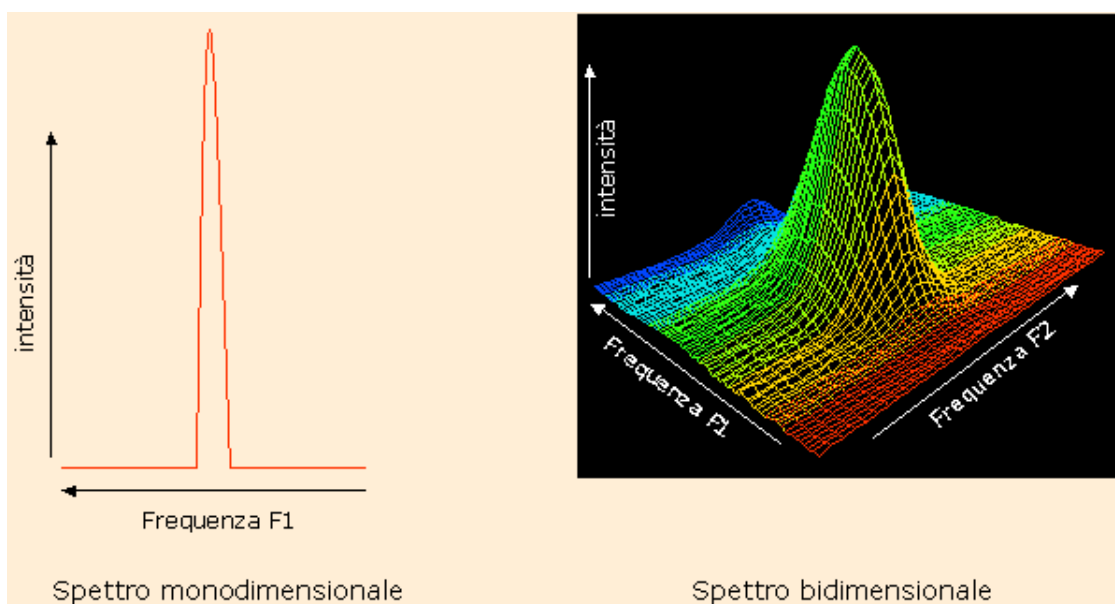
Spostamento Chimico d, ppm

Un'ulteriore caratteristica del ¹H-NMR è il fatto che l'intensità dell'assorbanza è proporzionale al numero di protoni che generano il segnale; ovvero, l'area sotto un certo

picco (l'integrale) è direttamente proporzionale al numero di quel tipo di protoni nella molecola. Le integrazioni sono date come i più semplici numeri interi che si ottengono dal rapporto tra le aree dei picchi, quindi l'acido acetico, CH_3COOH avrà due picchi nel $^1\text{H-NMR}$, uno a $d = 2$ con area = 3, ed un secondo a $d = 12$ con area = 1. L'acetato di metile, $\text{CH}_3\text{COOCH}_3$ avrà anch'esso due picchi nel $^1\text{H-NMR}$, uno a $d = 2$ con area = 1, ed un secondo a $d = 4$ con area = 1 (i due picchi hanno la stessa area relativa, quindi ognuno rappresenta tre idrogeni).

4.2. NMR bidimensionale

Lo spettro NMR bidimensionale 2D è in realtà un grafico a tre dimensioni (3-D); in tutti gli esperimenti NMR la dimensione omessa è sempre l'intensità, espressa in unità di misura arbitrarie. Le due dimensioni a cui si fa riferimento in un esperimento NMR-2D sono entrambe assi delle frequenze. Un esperimento di questo tipo richiede due trasformate di Fourier ortogonali tra di loro su due assi dei tempi indipendenti, per arrivare a due assi delle frequenze perpendicolari [35]. Negli esperimenti monodimensionali a singolo fattore vi è un unico fattore tempo che influenza lo spettro e cioè il tempo di acquisizione t_2 . Uno spettro 2D implica l'acquisizione simultanea di due funzioni temporali: due scale dei tempi indipendenti che, dopo una doppia trasformata di Fourier da luogo ad una matrice quadrata di punti che esprime valori di frequenza su entrambe gli assi [34].



4.3. Dinamica Molecolare

La simulazione di dinamica molecolare (MD) è usata per studiare una larga varietà di sistemi in vari stati (cristalli, soluzioni acquose, fasi gassose..etc). Essa permette di conoscere i minimi di energia potenziale che corrispondono a varie conformazioni molecolari mediante la seguente equazione di moto:

$$F_i(t) = m_i a_i(t)$$

dove F_i è la forza che agisce sull'atomo i-esimo

m_i è la massa dell'atomo i-esimo

$a_i(t)$ è l'accelerazione dell'atomo i-esimo

Questa equazione classica del moto è in grado di determinare le coordinate e la velocità dopo un certo tempo, conosciute le coordinate e la velocità iniziale. Esistono vari metodi di dinamica molecolare veloci e che richiedono poca memoria nel computer come Verlet, ABM4, Runge-Kutta 4. L'algoritmo di Verlet è il prototipo di tutti gli algoritmi di integrazione ed è tuttora usato. E' efficiente perché richiede un solo calcolo di forze (la parte più costosa dal punto di vista computazionale) ad ogni passo. E' numericamente stabile ed ha una proprietà assai importante per una buona MD: conserva la simmetria per inversione temporale delle equazioni del moto. In altri termini: se si inverte la velocità e la direzione del tempo ad un certo istante, la traiettoria ritraccia all'inverso quella seguita fino ad allora all'inverso.

Potenziali Empirici e meccanica molecolare

Lo studio teorico dei sistemi biologici permette di ricavare le relazioni tra struttura, funzione e dinamica a livello atomico. Molti dei problemi di interesse coinvolgono molti atomi per i quali non è ancora possibile adottare una trattazione quantistica. Tuttavia, i problemi diventano molto più trattabili utilizzando dei potenziali empirici, che sono molto meno dispendiosi (computazionalmente), anche se ciò porta ad una serie di approssimazioni e limitazioni. Le funzioni di energia potenziale (detti anche force fields) utilizzate portano ad un buon compromesso tra l'accuratezza fisica e l'efficienza computazionale. In effetti, tali funzioni sono in genere calibrate su dati

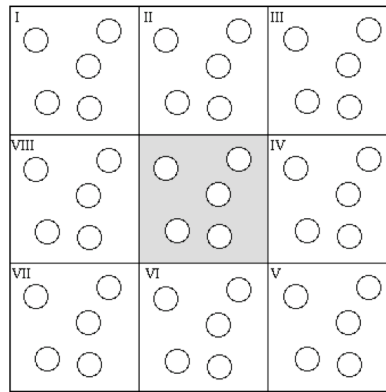
sperimentali e calcoli quantistici attraverso cui vengono ricavati i parametri dai quali tali funzioni dipendono.

Tra gli insiemi di funzioni di energia potenziale più usati, citiamo quelli usati dai più noti pacchetti di software per simulazioni MD: AMBER, CHARMM, GROMOS, più la variante OPLS alle funzioni AMBER. Una seria limitazione di tali approcci è la quasi impossibilità di descrivere drastici cambiamenti nella struttura elettronica del sistema: eventi come la formazione o la rottura di un legame non sono permessi, o più esattamente, sono male descritti. Una migliore descrizione richiederebbero approcci misti in cui parte delle interazioni sono trattate quantisticamente. Se tuttavia siamo interessati a seguire la dinamica di molecole biologiche in assenza di reazioni chimiche significative, tale limitazione non costituisce un problema.

Inclusione del Solvente in una Simulazione MD

L'acqua ha un ruolo fondamentale su struttura, dinamica e termodinamica delle molecole biologiche. Uno dei più importanti effetti del solvente è quello di schermo per le interazioni elettrostatiche. L'inclusione del solvente nelle simulazioni può essere fatta in vari modi. Il modo implicito più semplice è quello di includere una costante elettrostatica nel termine elettrostatico dell'energia potenziale, usando se necessario una costante dielettrica effettiva dipendente dalla distanza. Il modo esplicito consiste invece nell'includere molecole d'acqua nel sistema. Le molecole forniscono direttamente l'effetto di schermo (o almeno la parte legata al riorientamento delle stesse).

In questa trattazione più dettagliata bisogna utilizzare convenienti condizioni al contorno, sia per prevenire la diffusione delle molecole d'acqua in zone troppo lontane dalla proteina, che per permettere la simulazione e il calcolo delle proprietà macroscopiche utilizzando un numero limitato di molecole di solvente. Le condizioni al contorno periodiche permettono di fare la simulazione con un numero relativamente piccolo di molecole di solvente e in modo che le particelle risentano di forze analoghe a quelle che risentirebbero nel bulk della soluzione.



Il quadrato centrale è circondato dagli otto quadrati vicini e le coordinate delle particelle immagine in tali quadrati sono ottenute mediante semplici traslazioni delle posizioni delle corrispondenti particelle nel quadrato principale. Le forze nel quadrato principale sono calcolate utilizzando tutte le particelle (anche quelle dei quadrati secondari). Il raggio di cut-off è in genere scelto in modo che una particella del quadrato primario non veda la sua immagine nei quadrati secondari.

Ci sono casi in cui non è utile considerare le condizioni al contorno periodiche, e in cui, anche sfruttando la notevole potenza dei computer, le simulazioni vengono fatte incorporando un grande numero di molecole di solvente. In questi casi, la proteina viene circondata da una sfera d'acqua e ottenuta utilizzando un potenziale al bordo che confina le molecole d'acqua in tale volume. In altri casi, l'acqua viene confinata solo nella parte interessante della simulazione.

Set-up e running di una Simulazione MD

Il risultato di una simulazione MD è una serie temporale di conformazioni corrispondenti alle traiettorie o cammini seguiti da ogni atomo. Le varie tappe di una simulazione possono essere così schematizzate:

- **Inizializzazione.** Per cominciare si sceglie la configurazione iniziale del sistema $t = 0$. Nella maggior parte dei casi, per la simulazione di biomolecole si parte dalla struttura (cristallografica o NMR) ottenuta dal Brookhaven Protein Databank. La scelta della configurazione iniziale deve essere fatta con attenzione perchè essa può influenzare la qualità della simulazione.

Prima di partire con la vera e propria simulazione è spesso bene procedere ad una minimizzazione dell'energia della struttura. A questo punto vengono addizionate le molecole d'acqua: spesso le molecole d'acqua sono prese da una precedente simulazione fatta in un box di dimensioni opportune; la proteina viene inserita in tale box eliminando le molecole d'acqua ad essa sovrapposte. A questo punto si effettua una nuova minimizzazione dell'energia per permettere all'acqua ed alla proteina di assestarsi nella nuova situazione.

- **Riscaldamento.** Vengono assegnate le velocità iniziali (in modo da avere una bassa temperatura) e si avvia la simulazione. Nel caso di solvente 60 esplicito, prima si blocca la proteina e si lascia termalizzare le molecole d'acqua, poi si sblocca la proteina e si comincia l'evoluzione del sistema proteina + acqua insieme. Durante la fase di riscaldamento periodicamente le velocità vengono riscaldate in modo da aumentare la temperatura fino ad arrivare a quella prescritta.

- **Equilibratura.** Raggiunta la temperatura prescritta la simulazione continua in modo che possano essere monitorate quantità come pressione, temperatura ed energia. In questa fase, talvolta è necessario, per ovviare a innaturali fluttuazioni della temperatura, riscaldare le velocità opportunamente.

La fase continua fino a quando l'equilibratura dei parametri suddetti è raggiunta.

- **Fase di Produzione.** Nella fase finale, dopo l'equilibratura, lasciando andare la simulazione, si possono le medie dei parametri termodinamici di interesse. Tale fase può corrispondere a varie centinaia di picosecondi a nanosecondi.

5. Elettroforesi capillare

5.1. Principi generali

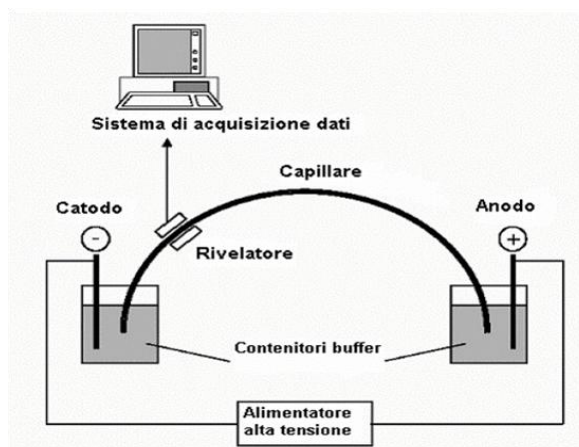
L'elettroforesi capillare (*Capillary Electrophoresis*, CE) è una tecnica separativa caratterizzata da consumi estremamente bassi di campione e di solventi, per cui fa parte a pieno titolo del settore della *green chemistry*. Il suo ridotto impatto ambientale è unito

a costi di esercizio relativamente bassi e ad alte prestazioni analitiche, specie in termini di risoluzione [36].

La migrazione degli analiti avviene all'interno di un capillare composto da silice fusa e ricoperto di poliimmide, di lunghezza variabile e di diametro interno 10-100 μm . Dentro il capillare si trova il mezzo in cui avviene la separazione, costituito da un sistema tampone detto *BackGround Electrolyte* (BGE) [36].

La separazione degli analiti è ottenuta applicando una differenza di potenziale tra due elettrodi posti alle estremità del capillare. Il campo elettrico così prodotto fa migrare gli analiti con differenti velocità, in base alle loro diverse caratteristiche chimico-fisiche, fino all'estremità del capillare dove è posto il rivelatore [37]. Il capillare, dato il suo elevato rapporto superficie/volume, permette una rapida dissipazione del calore aumentando così l'efficienza di separazione.

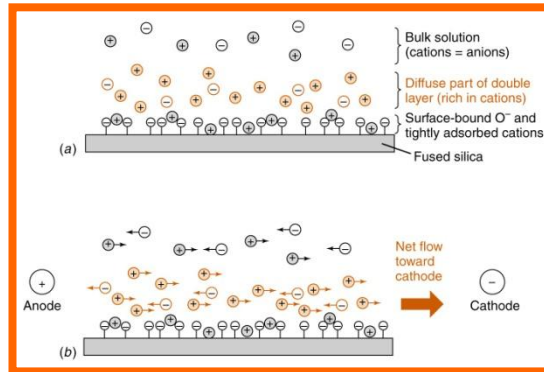
Il meccanismo dell'elettroforesi capillare si basa su due principi basilari: il flusso elettroosmotico e la mobilità elettroforetica.



Flusso elettroosmotico

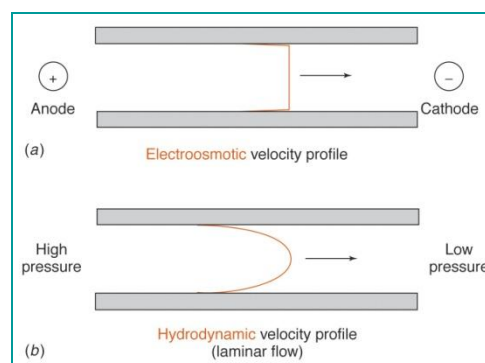
La parete interna del capillare è ricoperta di gruppi silanolici, che in condizioni di pH superiori a 3 si caricano negativamente. Con questi presupposti le cariche positive del BGE si dispongono in modo da formare un doppio strato elettrico sulla parete del capillare. Lo strato direttamente adiacente alla parete del capillare viene chiamato strato fisso o strato di Stern, è formato esclusivamente da cationi ed è contraddistinto da una scarsa mobilità dovuta all'adsorbimento delle cariche positive da parte dei gruppi silanolici caricati negativamente. Lo strato più vicino al centro del capillare, detto strato diffuso o strato di Gouy-Chapman, è ricco di ioni, in prevalenza cationi, che possiedono

una maggiore mobilità e si scambiano di continuo con gli ioni presenti nel resto della soluzione [36].



Con l'applicazione del campo elettrico alle estremità del capillare, i cationi dello strato diffuso tendono a migrare verso il polo negativo, ed essendo solvatati trasportano con sé le molecole d'acqua. All'interno del capillare viene così a crearsi un flusso netto di liquido che migra dall'anodo verso il catodo, definito flusso elettroosmotico (*ElectroOsmotic Flow*, EOF) [36].

L'elettroosmosi produce un flusso piatto, uniforme attraverso più del 99.9% della sezione del capillare, diversamente dal profilo parabolico che si ha in HPLC; in altre parole, la velocità del flusso diminuisce solo nell'immediata prossimità della parete. Questa particolarità apporta vantaggi nell'efficienza dell'analisi, in quanto la forza motrice del flusso è omogenea lungo il capillare rendendo la diffusione radiale trascurabile [38].



Qualsiasi effetto che diminuisce l'uniformità del flusso produce un allargamento delle bande di migrazione e diminuisce la risoluzione. Questa situazione negativa può essere riscontrata ad esempio nel caso dell'effetto Joule. Questo effetto fa sì che la corrente generata a causa dell'applicazione del campo elettrico ad una soluzione contenente ioni generi a sua volta calore. L'aumento di temperatura dipende dalla potenza generata (prodotto di voltaggio e corrente). Il gradiente di temperatura che si forma tra centro del capillare e pareti causa zone a diverse viscosità nel BGE e conseguente allargamento della zona di migrazione, perché i soluti al centro del capillare migrano più velocemente di quelli più vicini alle pareti a causa della minore viscosità. Inoltre, nell'elettrolita si possono creare moti convettivi con conseguente rimescolamento delle zone separate [39].

La velocità elettroosmotica può essere definita secondo l'equazione, dove μ_{eof} è la mobilità elettroosmotica ed E è il campo elettrico. μ_{eof} viene a sua volta descritta dalla formula $\mu_{EOF} = \varepsilon\zeta/4\pi\eta$, dove ε è la costante dielettrica del mezzo, η è la viscosità e ζ è il potenziale zeta [36]. Il potenziale ζ dipende dallo spessore del doppio strato δ , dalla costante dielettrica ε e dalla quantità di carica per unità di superficie e secondo l'espressione $\zeta = \frac{4\pi\delta e}{\varepsilon}$; a sua volta δ dipende dal numero di elettroni di valenza Z e dalla concentrazione dell'elettrolita C , come descritto dall'equazione:

$$\delta \cong \frac{1}{3 \cdot 10^7 |Z| \sqrt{C}}$$

In generale, quindi, con l'aumentare del pH del BGE aumenta la dissociazione dei gruppi silanolici e quindi EOF aumenta; con l'aumentare della concentrazione del BGE EOF diminuisce [39].

Mobilità elettroforetica

Quando si applica un campo elettrico E di intensità V/m ad una soluzione contenente specie ioniche con carica q (misurata in c), questo esercita sugli ioni una forza elettrica $F_E = qE$. Sulle specie cariche ha effetto anche la forza frizionale $F_F = fv_i$, diretta in senso opposto, dove f è il coefficiente d'attrito e v_i è la velocità della specie ionica. Il coefficiente d'attrito f è a sua volta dato da $f = 6\pi r_i \eta$, dove η è

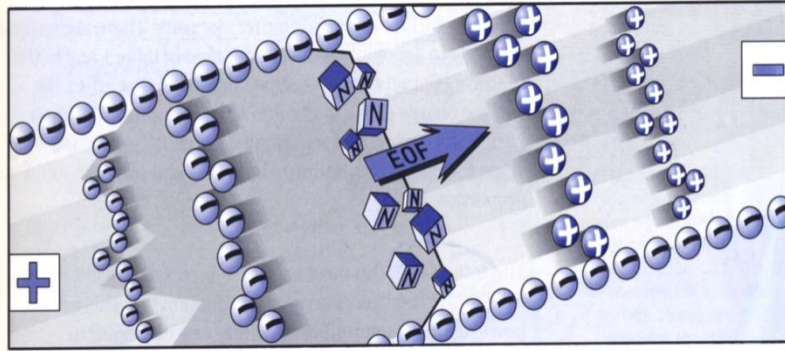
la viscosità del mezzo e r_i è il raggio dello ione. Le due forze, elettrica e frizionale, determinano la velocità elettroforetica della specie nel muoversi verso il compartimento elettrodico di segno opposto, pari a $v_i = \mu_i E$ [39]. Quando tali forze si eguagliano ($F_E = F_F$), la specie carica raggiunge la velocità stazionaria durante la corsa elettroforetica. Dall'uguaglianza $qE = -6\pi\eta r_i v_i$ si ricava quindi la velocità dello ione

$$v_i = \frac{qE}{6\pi\eta r_i}, \text{ da cui si definisce la mobilità elettroforetica come } \mu_i = \frac{q}{6\pi\eta r_i} \text{ [39].}$$

Quest'ultima formula evidenzia l'effetto che hanno la carica e la dimensione della specie ionica sulla mobilità elettroforetica. Specie di piccole dimensioni e fortemente cariche presentano un'elevata mobilità, al contrario di specie più grandi e poco cariche che hanno una bassa mobilità [40].

La mobilità finora descritta è la mobilità elettroforetica teorica, la cui espressione risulta valida solamente se si prendono in considerazione specie assimilabili ad una particella sferica, del tutto dissociata, con un valore di mobilità estrapolato a diluizione infinita. La mobilità che si ricava sperimentalmente è invece la mobilità elettroforetica effettiva (μ_E), dipendente da altri parametri aggiuntivi tra cui la solvatazione, la temperatura, la carica effettiva, la composizione, la forza ionica e il pH del BGE [39].

La mobilità apparente μ_A di una specie è la somma della mobilità elettroforetica effettiva dello ione e della mobilità elettroosmotica, come descritto dall'equazione $\mu_A = \mu_{EOF} + \mu_E$. I cationi si muovono nella stessa direzione del flusso elettroosmotico, verso il catodo, per cui la mobilità effettiva e quella del flusso elettroosmotico hanno lo stesso segno. Diversamente, gli anioni hanno mobilità elettroforetica effettiva diretta verso l'anodo, ma vengono comunque spinti verso il catodo, anche se con velocità inferiore rispetto ai cationi, grazie alla mobilità elettroosmotica. Le molecole neutre, prive di una propria mobilità elettroforetica perché prive di carica, migrano verso il rivelatore trasportate dal flusso elettroosmotico e giungendo tutte alla stessa velocità. Da ciò si deduce che per separare diverse specie neutre occorre utilizzare particolari sistemi operativi della CE [36].



Iniezione del campione

L'iniezione del campione nel capillare può avvenire attraverso due modalità: iniezione idrodinamica o iniezione elettrocinetica.

La prima impiega una differenza di pressione tra le due estremità del capillare di durata ed entità fissate sperimentalmente. In tal modo viene introdotto un piccolo volume di campione all'interno del capillare, dell'ordine dei nL. Il volume di iniezione può essere calcolato secondo l'equazione di Poiseuille:

$$V_c = \frac{\Delta P \pi d^4 t}{128 \eta L_t}$$

dove ΔP è la differenza di pressione tra le estremità del capillare, d è il diametro interno del capillare, t è il tempo di iniezione, η è la viscosità del campione, L_t è la lunghezza totale del capillare [40]. L'iniezione idrodinamica risulta migliore per l'analisi quantitativa perché più riproducibile.

Il metodo elettrocinetico impiega un campo elettrico per introdurre il campione nel capillare. Applicando una differenza di potenziale alle estremità del capillare, per un periodo di tempo prestabilito, viene introdotta nel capillare stesso una parte del campione attraverso la migrazione degli ioni. Le moli immesse nel capillare vengono calcolate secondo la seguente formula:

$$\text{moli iniettate} = \mu_A \left(E \frac{k_s}{k_c} \right) t \pi r^2 C$$

dove r è il raggio del capillare, C è la concentrazione del campione e k_s/k_c è il rapporto delle conduttività del BGE e del campione [38]. Il metodo elettrocinetico

comporta una maggiore variabilità rispetto a quello idrodinamico perché la quantità effettiva di campione iniettata dipende dalla mobilità degli ioni presenti nel campione: ioni con più alta mobilità sono campionati in maggiore quantità. Per questo viene principalmente utilizzato per elettroforesi capillare a scopo qualitativo o per elettroforesi su gel.

Con lo scopo di aumentare la sensibilità della metodica analitica, è possibile adottare una tecnica di preconcentrazione on-line del campione, il cosiddetto *sample stacking*, attraverso la quale il campione viene focalizzato in bande più strette [37,38]. Il funzionamento dello *stacking* si basa sulla conduttività, che è inversamente dipendente dal campo elettrico. La zona del campione iniettato è racchiusa tra due zone di BGE. Il campione ha una forza ionica minore rispetto al BGE, quindi anche la sua conduttività sarà inferiore. Di conseguenza, il campo elettrico a livello della zona del campione è molto superiore rispetto a quello al livello del BGE; questo porta ad un aumento della velocità elettroforetica degli analiti del campione e quindi ad un restringimento della banda di migrazione.

Sistemi di rivelazione

I più comuni sistemi di rivelazione usati nell'elettroforesi capillare sono i rivelatori UV e a fluorescenza. Per entrambi i metodi è necessario preparare il capillare creando una finestra di rivelazione. Tale finestra è ottenuta bruciando lo strato di poliimmide superficiale nella zona dove la luce emessa dal rivelatore incide sul capillare, bruciando o rimuovendo meccanicamente la parte interessata [39]. In entrambi i casi, prodotta la finestra di rivelazione, si procede con la rimozione dei residui con un pezzetto di carta imbevuto di acetone. Il metodo di rivelazione UV è quasi universale, mentre quello a fluorescenza è sensibile ad analiti fluorescenti per propria natura o a derivati fluorescenti [39].

La tecnica della rivelazione indiretta si basa sulla scelta di un BGE con proprietà chimico-fisiche a cui il rivelatore risponde con un segnale di fondo elevato. Quando le specie chimiche da analizzare raggiungono il rivelatore, abbassano la concentrazione del BGE e in questo modo viene prodotto un picco negativo. Qualsiasi ione che non possieda proprietà chimico-fisiche sfruttabili dai rivelatori a disposizione può essere rivelato con questo metodo dopo aver selezionato correttamente i componenti del BGE.

La rivelazione indiretta viene ad esempio usata con rivelatori UV per ioni inorganici, carboidrati ed amminoacidi non derivatizzati [39].

I metodi di rivelazione elettrochimici presentano elevata sensibilità e consistono in sistemi di rivelazione attraverso conducibilità e metodi amperometrici. Nella rivelazione attraverso conducibilità si misura la variazione di resistenza tra due elettrodi di platino posti l'uno di fronte all'altro all'interno del capillare [37]. Nei metodi amperometrici l'analita è ossidato o ridotto alla superficie di un elettrodo a cui è applicato un potenziale costante; viene così generata una corrente proporzionale alla concentrazione dell'analita. La zona di rivelazione deve essere isolata dalla zona di separazione, principalmente perché la corrente generata durante l'analisi è di parecchi ordini di grandezza superiore a quella misurata col rivelatore [37].

Analisi quantitativa

In CE i soluti migrano con velocità diverse, a differenza delle tecniche cromatografiche in cui tutti i soluti si muovono alla stessa velocità quando si trovano nella fase mobile. Quindi, se gli analiti presentano tempi diversi di permanenza nella regione di rivelazione, questo influenza l'area del picco: soluti a bassa mobilità rimangono nella finestra di rivelazione per un tempo più lungo di quelli a più alta mobilità, e hanno un'area del picco aumentata. Questo fenomeno viene corretto dividendo l'area del picco per il tempo di migrazione (area corretta A/t_m) [39].

I fattori principali che influenzano l'area del picco sono le variazioni di temperatura, l'adsorbimento del campione sulle pareti del capillare, variazioni nell'iniezione ed altri ancora, che possono essere strumentali e non. L'uso di uno standard interno risulta fondamentale per aumentare la riproducibilità dell'area corretta [39].

5.2. Modalità operative

In elettroforesi capillare esistono diverse modalità operative che ampliano i possibili campi di applicazione di questa tecnica.

Nell'elettroforesi capillare zonale (*Capillary Zone Electrophoresis*) all'interno del capillare si trova una semplice soluzione tampone. La separazione si basa

esclusivamente sulle differenze di mobilità elettroforetica tra gli analiti e quindi sulle loro diverse caratteristiche di carica e dimensione. Tale tecnica non è in grado di separare le specie neutre dato che queste non hanno carica e quindi non hanno mobilità propria, migrando insieme al flusso elettroosmotico [36, 37].

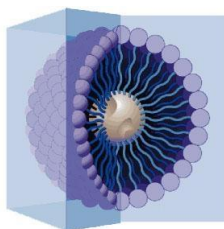
Per separare soluzioni campione contenenti specie neutre occorre utilizzare modalità operative che utilizzano fasi pseudostazionarie (PSP), che hanno una propria mobilità, o fasi stazionarie. In questo caso il meccanismo di separazione degli analiti carichi si basa sia sulla loro diversa mobilità elettroforetica sia sulla loro diversa ripartizione tra queste fasi e l'elettrolita acquoso. Gli analiti neutri sono invece separati solo in base alla loro diversa ripartizione [36, 37].

5.2.1. MEKC

L'elettroforesi capillare micellare (*Micellar ElectroKinetic Chromatography*, MEKC) è una tecnica in grado di separare sia molecole neutre che cariche. In questa tecnica vengono aggiunti al BGE tensioattivi a concentrazione superiore alla loro concentrazione micellare critica (CMC). Si formano quindi micelle che vanno a costituire la fase pseudostazionaria nel tampone di corsa [37,39].

I tensioattivi sono composti anfipatici, formati da una catena idrocarburica apolare e una porzione polare o ionica. Sono classificati in base alle loro caratteristiche di carica (anionici, cationici, zwitterionici, non ionici) o secondo il tipo di catena (idrocarburici, sali biliari, fluorocarburici) [36]. Le caratteristiche molecolari dei tensioattivi fanno sì che quando questi vengono immersi in acqua tendono a formare spontaneamente le micelle, aggregati molecolari in cui le teste idrofile sono rivolte verso l'esterno e le code idrofobe verso l'interno.

La *driving force* alla base della formazione delle micelle è la favorevole



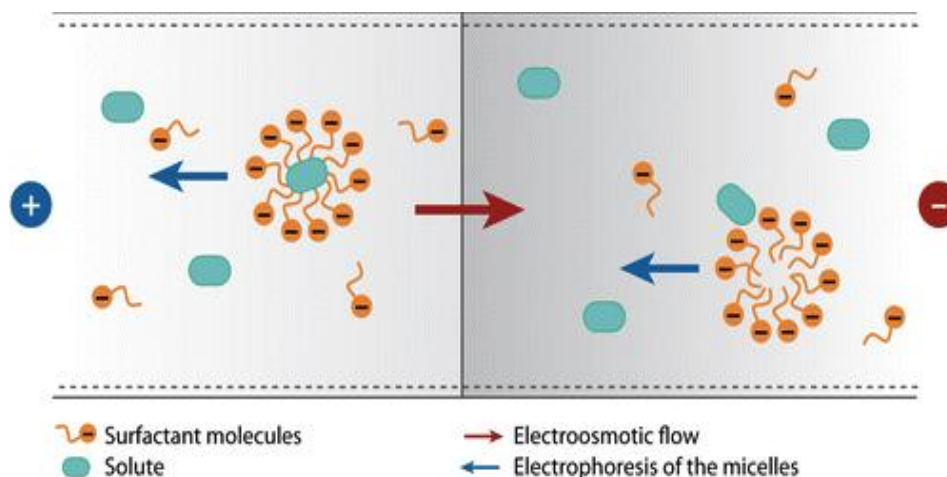
Struttura di una micella.

variazione di energia che guida l'allontanamento del tampone acquoso dalle catene idrofobiche del tensioattivo, promuovendo la formazione di un nucleo centrale e idrofobico ricoperto dalle teste polari o ioniche. La repulsione elettrostatica tra le teste nelle micelle ioniche e l'ingombro sterico delle teste idrate nel caso di quelle non ioniche si oppongono a questo guadagno di energia libera [36].

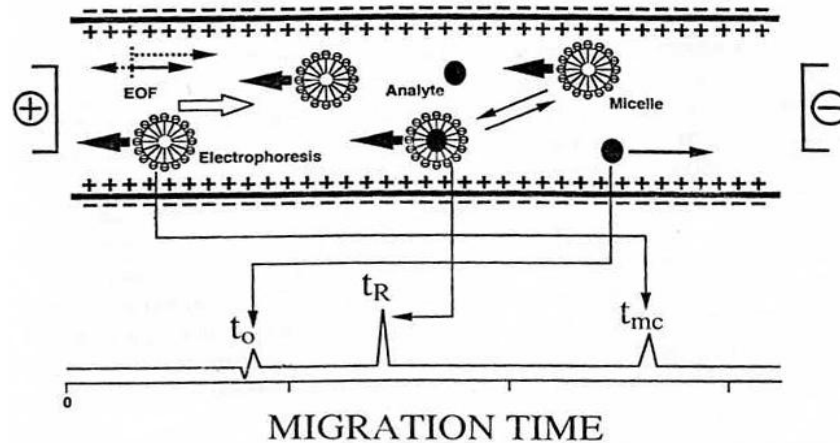
In MEKC i tensioattivi vengono comunemente usati in concentrazioni da 1 a 10 volte superiori alla CMC, dando così origine a micelle di forma sferica. Se la concentrazione viene ulteriormente aumentata o vengono aggiunti elettroliti o additivi organici, le micelle assumono forma ellissoidale, *rod-like* o altre forme in cui le teste polari risultano essere associate più vicine tra loro. Le micelle *rod-like* sono formate da una parte centrale cilindrica con due estremità sferiche e sono caratterizzate da larga dispersione dei numeri di aggregazione [36]. Le concentrazioni più usate si trovano nell'intervallo tra 20 e 100 mM; a concentrazioni superiori si ha un aumento di corrente che può portare ad un peggioramento della linea di base e ad ulteriori problemi per effetto Joule.

I tensioattivi non ionici e zwitterionici, avendo carica netta pari a zero, non possono essere usati da soli per la separazione dei composti neutri perché la loro mobilità è identica a quella del flusso elettroosmotico [36].

Per quanto riguarda il meccanismo di separazione, le molecole neutre sono in equilibrio dinamico tra la soluzione libera e l'interno della micella. Se viene applicato un campo elettrico, le specie neutre si separano in base alla loro diversa distribuzione tra la soluzione elettrolitica e le micelle.



La separazione degli analiti avviene all'interno di una finestra di migrazione definita dalla differenza tra la velocità di migrazione della soluzione di elettrolita, identica a quella del flusso elettroosmotico, e l'effettiva velocità di migrazione delle micelle. In altre parole, i soluti che risiedono esclusivamente nella fase mobile (soluzione dell'elettrolita) vengono eluiti dalla colonna al tempo t_0 , corrispondente al tempo di migrazione di un soluto neutro che non interagisce con la fase pseudostazionaria. I soluti che risiedono esclusivamente nella fase pseudostazionaria vengono eluiti al tempo t_{mc} , corrispondente al tempo di migrazione delle micelle. I soluti che si distribuiscono tra le due fasi vengono eluiti ad un tempo di migrazione t_R caratteristico per ogni soluto ($t_0 < t_R < t_{mc}$) [36]. L'eluizione dei soluti neutri dipende solamente dalla ripartizione nelle due fasi mentre per i soluti carichi incide pure la differenza delle mobilità elettroforetiche. La separazione degli analiti in esame dipende quindi dalla scelta di un adeguato tensioattivo che deve essere aggiunto al tampone ad opportuna concentrazione.



Rappresentazione schematica di un sistema MEKC contenente un tensioattivo anionico.

La teoria generale della MEKC rappresenta la confluenza tra i principi cromatografici e quelli elettroforetici.

La mobilità elettroforetica apparente di un analita in MEKC dipende da tre fattori:

1. la mobilità del flusso elettroosmotico;

2. la frazione di analita nella soluzione dell'elettrolita e la sua mobilità elettroforetica;

3. la frazione di analita nella fase pseudostazionaria e la mobilità elettroforetica delle micelle (assumendo che la mobilità del complesso analita-micella sia la stessa di quella delle micelle).

Il fattore di capacità di un soluto k' è definito come il rapporto tra le moli di analita nella fase pseudostazionaria e quelle nella fase acquosa, e viene calcolato misurando i tempi di migrazione secondo l'equazione:

$$k' = \frac{t_0 (t_R - t_0)}{t_0 (t_R / t_{mc} - t_0)}$$

La sola differenza tra il fattore di capacità in MEKC e in cromatografia convenzionale (dove $k' = \frac{t_R - t_0}{t_0}$) è il fattore $(t_R / t_{mc} - t_0)$ al denominatore dell'equazione, che tiene conto del movimento della fase pseudostazionaria. Quando t_{mc} tende ad infinito, si ha una vera e propria fase stazionaria e l'equazione assume la forma convenzionale [36]. Questa formula è valida solo se si ha la stessa mobilità elettroosmotica nelle modalità MEKC e CZE [51,52].

5.2.2. MEEKC

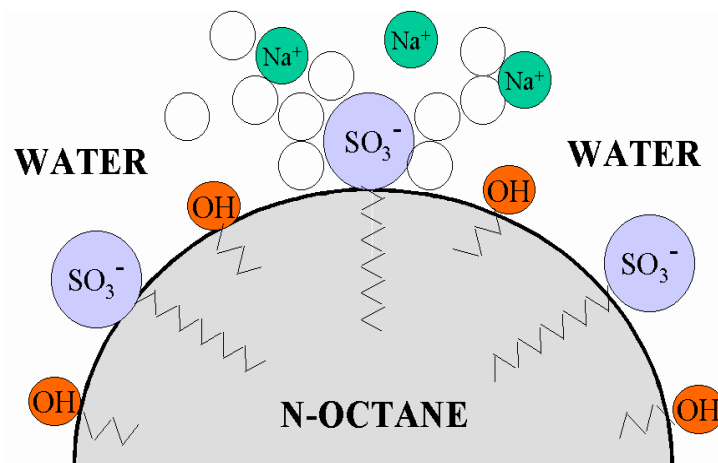
L'elettroforesi capillare a base di microemulsioni (*MicroEmulsion ElectroKinetic Chromatography*, MEEKC) è una modalità operativa che permette di ottenere separazioni molto efficienti di analiti sia carichi che neutri con ampio intervallo di idrosolubilità. Il BGE consiste in una microemulsione [41,42].

La microemulsione è una dispersione continua di gocce di un liquido immiscibile, solitamente olio in acqua, che può essere considerata come una modifica di un sistema micellare in cui nelle micelle è solubilizzato un solvente organico lipofilo [42]. Rispetto alle emulsioni le microemulsioni sono caratterizzate da minori dimensioni delle particelle disperse (<10nm), per cui sono sistemi trasparenti. Si tratta di sistemi stabilizzati dalla simultanea presenza di tensioattivi e di cotensioattivi (solventi organici più idrofili, ad esempio alcool a catena corta). I tensioattivi si interpongono tra la fase acquosa e quella oleosa, rendendo minima la tensione interfacciale tra i due liquidi immiscibili. L'aggiunta di cotensioattivi abbassa ulteriormente la tensione interfacciale,

rendendola quasi nulla. Occorre tuttavia scegliere la giusta proporzione di fase acquosa, fase oleosa, tensioattivo e cotensioattivo per ottenere una microemulsione stabile, altrimenti si possono verificare problemi di rottura della microemulsione stessa con separazione delle fasi immiscibili [43].

La presenza di tensioattivi ionici permette di separare soluti neutri grazie alla loro diversa ripartizione tra la fase oleosa e la fase acquosa costituita dal tampone. I soluti penetrano più facilmente la superficie di una goccia di olio che di una micella, più rigida, e questo permette l'applicazione della MEEKC ad un campo più ampio di soluti con diverse caratteristiche [44].

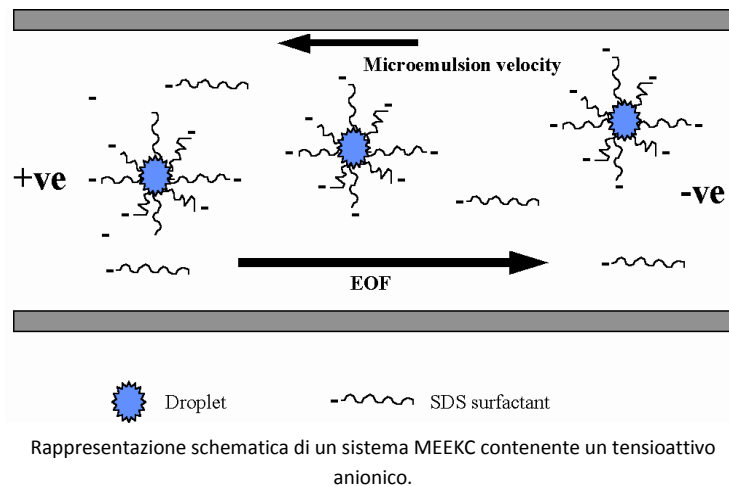
Nella MEEKC il tensioattivo più utilizzato è il sodio dodecilsolfato (SDS). Quest'ultimo è costituito da una catena idrocarburica a 12 atomi di carbonio, che rappresenta la parte idrofoba, e una testa idrofila costituita da un gruppo solfato carico negativamente. Il tensioattivo si dispone tutto attorno alla goccia di olio (ad esempio, *n*-ottano), le catene alchiliche penetrano nella goccia di olio, mentre i gruppi idrofili rimangono all'esterno della goccia in contatto con la fase acquosa. Le cariche negative dei gruppi solfato creano una forza repulsiva, che impedisce l'eccessivo raggruppamento delle gocce e previene la formazione di emulsioni. Il cotensioattivo, solitamente un alcool a catena corta (ad esempio, *n*-butanolo), collega la fase acquosa con la fase oleosa disponendo il proprio gruppo idrofilo sulla superficie della goccia, a contatto con la fase acquosa [43].



Schema rappresentante la disposizione dei componenti di una microemulsione.

Le gocce della microemulsione possono assumere diverse forme ma la maggioranza presenta forma sferica. Il sistema è definito ordinato se è caratterizzato da bassa polidispersione, ossia se le gocce hanno forme simili tra loro, normalmente sferiche [44].

Anche per la MEEKC è possibile misurare la finestra di migrazione dei soluti, attraverso l'utilizzo di *markers* di EOF e delle gocce della microemulsione. In generale, le microemulsioni con tensioattivi anionici migrano in senso contrario al flusso elettrosmotico provocando l'aumento del tempo di migrazione dei soluti che si distribuiscono nelle micelle. Soluti neutri idrosolubili come il metanolo si distribuiscono nella fase acquosa e vengono trascinati rapidamente da EOF verso il rivelatore al tempo t_0 , definito come il tempo di migrazione di un soluto neutro che non interagisce con la fase pseudostazionaria. Diversamente, soluti molto lipofili come il dodecilbenzene si ripartiscono all'interno delle gocce cariche e presentano tempo di migrazione pari a t_{ME} , che corrisponde al tempo necessario ad una goccia di olio per arrivare al rivelatore. Il tempo di migrazione t_R , caratteristico per ogni specie neutra, dipende dalla diversa distribuzione nelle due fasi ($t_0 < t_R < t_{ME}$).



Il fattore di capacità k' di un soluto in un sistema MEEKC indica il comportamento migrazionale degli analiti e definisce il rapporto tra le moli totali di analita nella PSP rispetto a quelle nella fase acquosa. Viene calcolato da una formula analoga a quella per la MEKC:

$$k' = \frac{t_R - t_0}{t_0 - t_R/t_{ME}}$$

Un fattore di capacità più elevato indica una migrazione più lenta e una maggiore ripartizione dell'analita nelle gocce di microemulsione [43]. Se il fattore di capacità k' è pari a 1 il soluto si ripartisce equamente tra le gocce di olio e la fase acquosa. Il fattore di capacità di un composto neutro e il suo tempo di migrazione possono essere direttamente correlati al grado di idrofilia (o lipofilia), permettendo l'uso della MEEKC nella valutazione della solubilità dei composti [45,46].

Sviluppo di un metodo MEEKC

La complessità della composizione delle microemulsioni e del processo di separazione MEEKC permette di effettuare una fine modulazione delle condizioni operative per ottenere separazioni in campioni particolarmente complessi.

In generale, per il controllo della selettività in MEEKC i tensioattivi sono molto più importanti rispetto al core lipofilo della goccia. L'adeguato tensioattivo permette di ottenere una buona separazione perché influenza la carica e la dimensione delle gocce d'olio ed il livello di accoppiamento ionico con i soluti carichi. In generale, aumentando la lunghezza della catena idrocarburica del tensioattivo si stabilizza la microemulsione riducendo la polidispersione [43]. Il tensioattivo più utilizzato è SDS perché dà un'ampia finestra di separazione [43]. Un aumento di concentrazione dei tensioattivi causa un aumento del tempo di migrazione dei soluti neutri, a causa di un accrescimento della densità delle cariche sulle gocce d'olio, e quindi un aumento di forza ionica con conseguente diminuzione di EOF [47]. D'altra parte, elevate concentrazioni di tensioattivi creano comunque microemulsioni più stabili, riducendo la tensione superficiale. Tipicamente si usano concentrazioni di SDS nell'intorno di 110 mM (3% w/w), anche se in casi particolari si è arrivati al 6.5%. A concentrazione pari allo 1.6% è stata riportata instabilità [48] e all'1.4% scarsa ripetibilità [49]. A concentrazioni di SDS uguali o inferiori al 2% è stata inoltre osservata disintegrazione delle microemulsioni dopo poche ore [48].

Il pH della soluzione tampone influisce sensibilmente sulla selettività della separazione, poiché ha effetto sia sul grado di ionizzazione dei soluti che sull'intensità

di EOF. Il range di pH 7-9 usualmente scelto in MEEKC produce un EOF di velocità abbastanza elevata. In tale intervallo di pH i soluti ionizzabili si presentano in forma dissociata. In generale infatti i composti basici rimangono protonati fino a valori di pH 11-12, mentre i composti acidi (pK_a 3-6) sono ionizzati. Per annullare la ionizzazione dei soluti si può agire usando valori estremi di pH [41,50]. Nel caso di analiti acidi, utilizzando valori molto bassi di pH (intorno a pH 1.5) viene eliminato EOF e quindi per far eluire le gocce di olio si usa una polarità inversa [41,46]. Diversamente, per annullare la ionizzazione dei composti basici si usano tamponi a pH 12 [46].

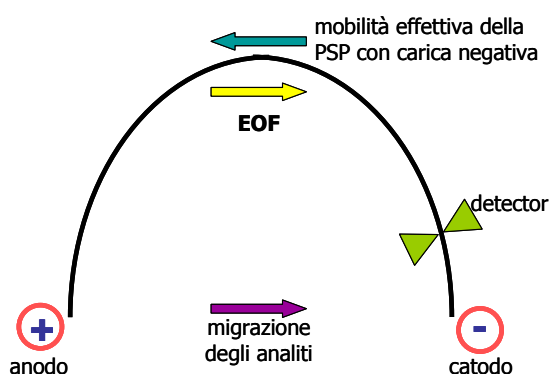
Come fase oleosa vengono usati solventi organici altamente lipofili, tra cui i più comuni sono *n*-eptano e *n*-ottano [41]. Gli alcani con catena formata da un numero dispari di atomi di carbonio come l'eptano sono meno tossici [52], ma l'ottano dà microemulsioni maggiormente ripetibili [53]. Le interazioni dell'analita con la fase organica lipofila giocano un ruolo limitato ai fini della selettività. Cambiando la concentrazione della fase organica nel range 0.5-2% si ottengono piccole modifiche nella migrazione.

Il cotensioattivo è il parametro primario da considerare per una rapida ottimizzazione dato che è il fattore più importante per controllare e cambiare la selettività. Il cotensioattivo più usato è *n*-butanolo. Si osservano variazioni significative di selettività sostituendolo ad esempio con *n*-propanolo, acetonitrile, 2-etossietanolo e tetraidrofurano [53,54]. Aumentando la concentrazione del cotensioattivo si ha un incremento delle dimensioni delle gocce [55], una più bassa densità di carica, e di solito una riduzione dei tempi di migrazione per i composti neutri [43], mentre i fattori di capacità non sono alterati [47]. Variando la concentrazione di cotensioattivo viene influenzata anche la viscosità e quindi il flusso elettroosmotico.

In MEEKC la temperatura deve essere tenuta costantemente sotto controllo in quanto influenza la solubilità del soluto, che è correlata al fattore di capacità [43]. La selettività di una miscela test di analiti neutri non viene influenzata dalla temperatura. Tuttavia, quando la soluzione da analizzare comprende soluti sia carichi che neutri si osservano variazioni di selettività, dato che la temperatura ha effetti diversi su di loro. [43]. La mobilità elettroforetica di una specie ionizzata è infatti influenzata dalla temperatura nell'ordine del 2% per °C [43].

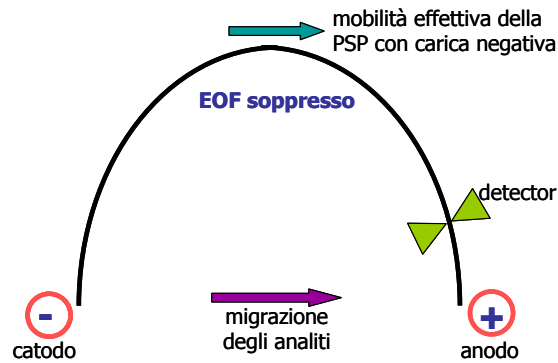
Polarità normale

Nella polarità normale l'anodo, corrispondente al polo positivo, è posizionato in entrata al capillare (*inlet*) e il catodo, corrispondente al polo negativo, è posizionato in uscita (*outlet*). Utilizzando una fase pseudostazionaria (PSP) negativa, questa presenta una propria mobilità effettiva negativa (verso l'anodo). Per trascinare tutti gli analiti e la PSP negativa verso l'uscita dal capillare è quindi necessario utilizzare un BGE alcalino in modo da ottenere un flusso elettroosmotico sufficientemente elevato. In queste condizioni la carica degli analiti risulta nulla o negativa.



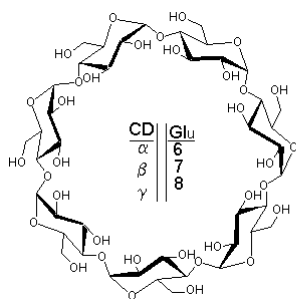
Polarità inversa

Nella polarità inversa il catodo è posizionato all'estremità *inlet* del capillare e l'anodo è posizionato all'estremità *outlet*. La mobilità effettiva di una PSP con carica negativa è diretta verso il polo positivo e quindi verso l'*outlet*. Per far sì che tutti gli analiti migrino in direzione dell'anodo occorre annullare il flusso elettroosmotico. Questa condizione viene ottenuta usando un tampone a basso pH ed alta forza ionica. La PSP carica negativamente (le micelle o le gocce d'olio della microemulsione) migra verso l'anodo ed è capace di trascinare con sé gli analiti, che in queste condizioni sono neutri o carichi positivamente.

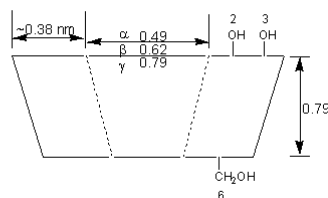


5.3. Utilizzo di Ciclodestrine

In elettroforesi capillare l'uso di un BGE contenente una ciclodestrina o una miscela di ciclodestrine è molto frequente.



Struttura delle ciclodestrine native (α , β , γ).



Le ciclodestrine (CD) sono oligosaccaridi ciclici costituiti da catene di 6, 7 o 8 molecole di D-(+)-glucosio con legami α -(1-4), definite rispettivamente α -, β - e γ -ciclodestrine. Non possiedono gruppi riducenti [56] e sono caratterizzate da una forma

troncoconica, l'interno della quale è altamente lipofilo mentre la superficie esterna ha caratteristiche idrofile (fig.) [57]. Ogni singolo monomero possiede cinque atomi di carbonio asimmetrici e questo conferisce a tutto il polisaccaride proprietà ottiche, che possono essere modulate grazie alla derivatizzazione sui carboni in posizione 2, 3 e 6 [57]. La possibilità di derivatizzazione dei composti nativi ha portato ad ottenere un gran numero di ciclodestrine che possono presentare carica netta nulla, positiva o negativa [57]. I derivati alchilici e idrossialchilici mostrano migliore solubilità delle ciclodestrine native e sono relativamente economici. Tra le ciclodestrine ionizzabili le più frequenti sono i derivati anionici come solfati, solfoalchilici, carbossialchilici, fosfati. Tuttavia vengono usati anche derivati cationici come le ciclodestrine amino

sostituite [57]. Le ciclodestrine neutre sono principalmente utilizzate per separare soluti con una carica netta. La formazione del complesso soluto/ciclodestrina porta ad una variazione della mobilità elettroforetica del soluto complessato rispetto al soluto libero e ciò permette di separare gli analiti sulla base della loro diversa affinità per le ciclodestrine. La mobilità media apparente di un soluto (μ_{app}) può essere correlata alla mobilità intrinseca del soluto (μ_{Λ}) e del complesso di inclusione soluto-ciclodestrina ($\mu_{\Lambda CD}$), come $\mu_{app} = \frac{\mu_{\Lambda} + \mu_{\Lambda CD} K [CD]^-}{1 + K [CD]^-}$ [56, 58]. Quindi, la qualità della risoluzione dipende della concentrazione della ciclodestrina nell'elettrolita e la selettività dipende dal fatto che si formi un complesso più o meno stabile con la ciclodestrina [56, 58].

In campo farmaceutico le ciclodestrine vengono utilizzate con successo per separare e quantificare principi attivi in matrici biologiche e per determinare i principi attivi e le loro impurezze all'interno delle forme farmaceutiche. Questi oligosaccaridi sono infatti capaci di differenziare gli isomeri posizionali di alcuni composti aromatici, principalmente strutture in cui sono presenti anelli benzenici o naftaleni [8,56].

Altri tipi di applicazione trovano le ciclodestrine cariche negativamente, che si comportano come una specie di micelle che migrano con una velocità inferiore rispetto al resto del tampone allargando così la finestra di separazione. Concettualmente queste costituiscono una fase pseudostazionaria simile a quella che si ottiene nella MEKC. In questo caso le prestazioni sono particolarmente sensibili sia alla concentrazione delle ciclodestrine che alle variazioni di pH, che modifica la velocità di EOF e quindi la dimensione della finestra di separazione [56]. La presenza di una carica sulla ciclodestrina permette la separazione di miscele contenenti sia composti neutri che carichi. Il principale svantaggio che limita l'applicazione delle ciclodestrine anioniche è il fatto che usandone elevate concentrazioni si sviluppano elevati livelli di corrente e occorrono tempi di analisi molto lunghi [56].

L'utilizzo delle ciclodestrine risulta particolarmente utile per le analisi ambientali di ricerca di contaminanti tossici, spesso presenti in miscele complesse di composti molto simili in basse concentrazioni. Molti dei contaminanti ambientali sono caratterizzati da strutture neutre e insolubili in acqua, come i policloro bifenili, le

policloro diossine e gli idrocarburi policiclici aromatici. Particolare attenzione è stata rivolta verso questa ultima classe di contaminanti, principalmente per la loro cancerogenicità, quindi è stata sviluppata una serie di metodi per la loro determinazione [59-62]. Tra questi metodi quello che dà le migliori prestazioni [62] consiste nell'utilizzare due β -ciclodestrine, una anionica ed una neutra, così che i soluti sono separati in base alla loro differente affinità per la ciclodestrina neutra, che migra insieme al flusso elettroosmotico, e quella per la ciclodestrina carica, che risulta rallentata dall'attrazione verso l'anodo. Le coppie di ciclodestrine principalmente utilizzate sono due: solfobutil- β -CD/metil- β -CD e carbossimetil- β -CD/idrossipropil- β -CD. L'utilizzo della prima coppia porta a migliori risultati per quanto riguarda la risoluzione, l'efficienza e la velocità di analisi [56].

Attualmente le ciclodestrine costituiscono il gruppo più importante di selettori chirali per elettroforesi capillare [56]. Questo principalmente per il loro alto potere risolutivo e per la variabilità e disponibilità di ciclodestrine presenti sul mercato. Il meccanismo più ampiamente accettato per l'enantiorisoluzione con una ciclodestrina consiste nell'inclusione dell'analita chirale nella cavità e nello stabilirsi di un'interazione secondaria con i gruppi ossidrilici sul bordo della ciclodestrina [56]. Queste interazioni differiscono tra gli enantiomeri e sono responsabili delle differenze nelle costanti di inclusione K_1 e K_2 . L'inclusione del soluto nella cavità è una condizione necessaria ma non sufficiente per l'enantiorisoluzione, per cui la dimensione e la forma del selettore e del selettando devono essere attentamente controllate. L'enantioseparazione è possibile anche se i due enantiomeri hanno la stessa costante di legame con il selettore, se i due complessi diastereomerici hanno diversa mobilità [56, 63]

I risultati migliori di enantiorisoluzione vengono ottenuti con ciclodestrine modificate. Sostituendo gruppi ossidrilici con gruppi alchilici si aumenta la solubilità e la flessibilità della ciclodestrina; questi due fattori facilitano l'adattamento del soluto ed aumentano la stabilità del complesso risultante come pure le capacità di enantiorisoluzione [64].

In generale le ciclodestrine cariche offrono un potere risolutivo maggiore rispetto a quelle neutre grazie alle possibili interazioni elettrostatiche con enantiomeri di segno

opposto. Tenendo sotto particolare controllo il flusso elettroosmotico è possibile anche separare sostanze cariche di segno uguale a quello del selettore. Data la natura principalmente basica delle molecole con attività biologica lo sviluppo di nuove ciclodestrine e il loro utilizzo è sbilanciato a favore di quelle anioniche. Le ciclodestrine cationiche vengono principalmente utilizzate per la separazione di molecole che presentano gruppi acidi, e il loro utilizzo richiede il rivestimento del capillare con poliacrilamide per evitare che la carica della ciclodestrine si leghi in maniera stabile ai gruppi silanologici del capillare [57].

5.4. Convalida

La convalida di una procedura analitica viene effettuata per assicurare che questa risponda allo scopo per cui è stata sviluppata, mostrando adatte prestazioni e fornendo dati affidabili. Il processo di convalida può essere suddiviso in diverse fasi: identificazione di appropriati parametri, disegno degli esperimenti per la loro valutazione, definizione dei limiti di accettabilità per ogni parametro e trasferimento dei risultati nel procedimento [65-68].

In campo farmaceutico la convalida risponde a precise normative e costituisce parte fondamentale del documento di registrazione per un nuovo farmaco. Inoltre, dato che gli aspetti analitici sono fortemente integrati in tutte le fasi dello sviluppo e del controllo di qualità di un principio attivo, l'uso di procedure analitiche convalidate permette di evitare perdite di tempo e denaro dovute a decisioni sbagliate perché basate su risultati non affidabili. Si fondano infatti su risultati analitici aspetti importanti della vita di un farmaco, come ad esempio la determinazione della sua emivita da studi di stabilità, la necessità di prove tossicologiche nel caso che le impurezze eccedano determinati livelli, il rilascio di lotti o il loro blocco in seguito a risultati fuori specifica.

Recentemente per armonizzare i requisiti per la registrazione dei medicinali sono state elaborate direttive comuni in Europa, Giappone e Stati Uniti, in modo da semplificare l'iter legislativo senza compromettere qualità, efficacia e sicurezza. L'ICH (International Conference on Harmonization) ha emesso specifiche linee guida tra cui quelle che definiscono i vari parametri di convalida delle procedure analitiche [65].

Queste forniscono la base per il calcolo e l'interpretazione dei diversi parametri, ma rimane responsabilità dell'analista il disegno di un opportuno protocollo sperimentale e la definizione di criteri di accettabilità a cui seguano valutazioni appropriate dei risultati [67].

I parametri indicati dalla ICH sono specificità, robustezza, limite di rilevabilità e determinazione (LOD e LOQ), linearità e range, esattezza e precisione, criteri di idoneità del sistema. I parametri da valutare nei casi particolari dipendono dall'applicazione: ad esempio, nel dosaggio di un principio attivo non è necessario valutare LOD e LOQ, mentre nel caso delle impurezze è necessario valutare LOQ o LOD nel caso che, rispettivamente, si voglia effettuare un dosaggio quantitativo o un test limite. In questo ultimo caso il solo altro parametro da testare oltre al LOD è la specificità.

La convalida delle procedure analitiche viene efficacemente supportata da tecniche di disegno sperimentale [69]. Se viene effettuata nelle prime fasi di sviluppo del metodo può influenzare aspetti della procedura analitica come numero di repliche o modo di calibrazione e aiuta ad eliminare possibili fonti di errore [69], per cui è preferibile effettuare la convalida in modo progressivo iniziando dai dati di precisione e robustezza.

Il processo di convalida è inoltre in continua evoluzione e vale per tutta la vita della procedura analitica. Infatti assicura che la procedura rimanga efficiente ed affidabile nel tempo, attraverso l'adozione di misure per controllare e mantenere lo stato convalidato a lungo termine nelle applicazioni di routine. Può essere necessaria una riconvalida della procedura (parziale o totale) nel caso vi siano variazioni nella sintesi del farmaco, nella composizione del prodotto finito o nel metodo analitico, o nel caso si vogliano sfruttare nuove tecnologie o automatizzare una procedura. La convalida costituisce la base per un continuo monitoraggio dei risultati attraverso *system suitability tests* e *control charts*, facendo così parte essenziale del concetto globale di assicurazione di qualità.

6. Quality by Design

6.1. Principi generali

Con Quality by Design (QbD) si intende un modello di qualità fondamentale in campo farmaceutico, da tenere in considerazione sia durante lo sviluppo dei prodotti che dei processi. Il concetto di QbD è stato recentemente introdotto dalla US Food and Drug Administration (US FDA) [70- 72] e i principi e gli strumenti che ne stanno alla base sono stati descritti nelle linee guida emesse dalla International Conference on Harmonization (ICH) [73-75]. Nell'ambito del QbD vengono utilizzati in modo estensivo ed integrato strumenti come valutazione dei rischi e strategie di tipo chemiometrico al fine implementare la qualità dei prodotti e dei processi.

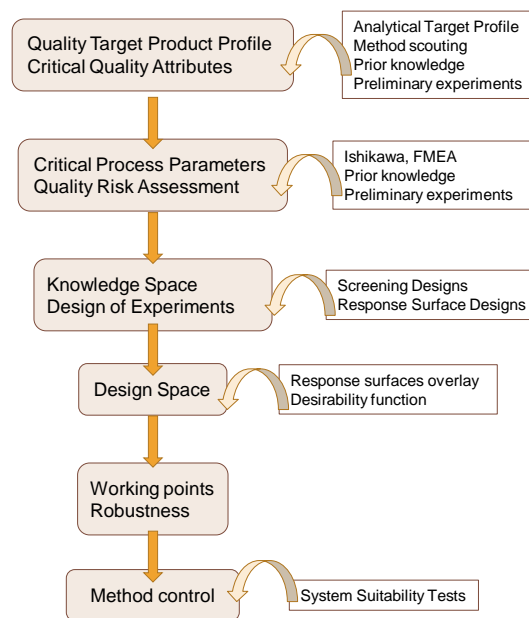
A livello di industria farmaceutica, l'utilizzo di strategie multivariate è in continua diffusione, sia nel settore analitico che in quello tecnologico. Tale approccio permette infatti di progettare un sistema di qualità ed assicurare quindi un prodotto finale di qualità (farmaco, nutraceutico) [76]. A livello di organi di controllo, la linea guida PAT (Process Analytical Technology) [71], emessa dalla US FDA, raccomanda l'uso di strategie multivariate come strumenti utili per la comprensione dei processi, il continuo miglioramento e lo sviluppo di strategie che minimizzano i rischi. Il QbD è in particolare definito come un approccio sistematico per lo sviluppo di sistemi che prevede la definizione di obiettivi ed enfatizza la comprensione del prodotto e del processo mediante la definizione di modelli matematici e la gestione dei rischi [73]. L'International Conference on Harmonization (ICH) descrive i principi e gli strumenti per implementare il "Quality by Design" [73-75]. Uno di questi strumenti è appunto il Disegno degli Esperimenti (DoE), che permette di comprendere come le variabili di processo e/o di formulazione influenzino il prodotto finale attraverso la definizione del cosiddetto *design space*. Il *design space* (DS) è definito come una regione dello spazio sperimentale dove la combinazione dei valori assunti dalle variabili assicura qualità nelle prestazioni del processo e la qualità stessa può essere descritta per mezzo di un modello matematico [73].

Le strategie chemiometriche permettono di stabilire relazioni tra differenti misure di un sistema chimico e lo stato del sistema attraverso l'applicazione di metodi matematici o statistici [12]. I modelli matematici che vengono trovati con l'impiego di strategie

multivariate permettono una comprensione del processo e contribuiscono a costruire e a mantenere la qualità del prodotto finale migliorando globalmente l'efficienza del processo [77]. In generale, l'uso di un DoE rispetto ad una strategia univariata ha il vantaggio di richiedere un minor numero di esperimenti, e di fornire informazioni circa le interazioni tra fattori mediante la stima di modelli matematici che correlano le variabili in esame alla risposta. La selezione di un'appropriata strategia sperimentale è quindi fondamentale per ottenere un'informazione corretta ed esauriente del problema in esame con il minimo costo. Il DoE è inoltre fondamentale negli studi di preformulazione e formulazione farmaceutica, dove svolge un ruolo fondamentale e sempre maggiormente richiesto dagli organi di controllo per identificare le variabili di formulazione e/o di processo attive sulle risposte selezionate e permettere quindi un razionale ed efficiente sviluppo della forma farmaceutica con i desiderati requisiti di velocità di rilascio, stabilità ed efficacia terapeutica [7,78].

Applicando i principi del QbD la qualità è progettata direttamente all'interno del processo fin dall'inizio, in contrapposizione quindi al classico approccio Quality by Testing (QbT), in cui si verifica la qualità del prodotto controllandolo alla fine del processo di lavorazione. Il QbD era stato originariamente concepito per lo sviluppo del prodotto farmaceutico ed è stato principalmente applicato a piccole molecole nel caso di problemi di formulazione [79-83] e di sintesi del principio attivo (*Active Pharmaceutical Ingredient*, API). Il QbD ha però ultimamente richiesto sempre maggiore attenzione nel caso dello sviluppo di metodi analitici separativi, in quanto questi vengono utilizzati nel controllo di qualità dei principi attivi e dei prodotti farmaceutici allo scopo di garantire la qualità del prodotto stesso e quindi la sicurezza dei pazienti [84]. I metodi analitici possono così essere considerati come una parte integrante del concetto del QbD e devono soddisfare i requisiti per cui sono stati sviluppati, in modo analogo ai requisiti da soddisfare nel caso di una formulazione farmaceutica [84]. Molti termini utilizzati nel QbD sono familiari ai chimici analitici, incluso il disegno sperimentale, che viene già ampiamente usato nello sviluppo di metodi cromatografici e di elettroforesi capillare [85, 86]. Recentemente l'applicazione dei principi del QbD nello sviluppo dei metodi analitici è stata descritta in diversi articoli scientifici, oggetto di una *review* di Vogt et al. [87]. La maggioranza dei metodi pubblicati utilizza la cromatografia mentre si ritrova un solo lavoro è dedicato all'uso

del QbD nell'elettroforesi capillare. Inoltre, non tutti i lavori pubblicati implementano ogni aspetto del QbD e i diversi aspetti possono essere stati interpretati diversamente dai vari autori. La figura sotto riportata descrive una guida generale del QbD per lo sviluppo di metodi analitici separativi [97].



6.2. Schema generale applicativo del QbD

Definizione del profilo target della qualità del prodotto (QTPP) ed attributi critici della qualità (CQA)

Il profilo del target analitico (*Analytical Target Profile, ATP*) è una serie di criteri che definiscono cosa verrà misurato (per esempio, il livello di una impurezza specificata) ed i criteri di prestazione che devono essere raggiunti attraverso la misura (per esempio, livello di accuratezza, precisione e range), senza tuttavia specificare il metodo stesso [87]. Sulla base dell'ATP, in una fase preliminare di investigazione vengono valutati differenti metodi e/o tecniche analitiche in modo da avvicinarsi all'obiettivo del metodo (*method scouting*). In generale, nei metodi separativi lo scopo di questa fase è il

raggiungimento della massima selettività con adeguata efficienza insieme ad una buona riproducibilità e ripetibilità delle misure. Nei metodi HPLC questa ricerca riguarda la selezione di condizioni di base, per esempio tipi di colonna (dimensioni e composizione), pH e modificatori organici [87]. Nei metodi CE la ricerca riguarda l'applicazione di diverse modalità operative (e quindi la selezione al loro interno) sulla base della composizione del BGE, il quale può consistere in un semplice soluzione acquosa o contenere una fase pseudostazionaria costituita principalmente da micelle o microemulsioni. Dopo questi esperimenti preliminari, il flusso lavoro del QbD può iniziare, per prima cosa definendo il profilo target della qualità del prodotto (*Quality Target Product Profile*, QTPP) e gli attributi critici della qualità (*Critical Quality Attributes*, CQAs). Il QTPP viene definito come un "riassunto delle caratteristiche di qualità di un prodotto farmaceutico che idealmente dovrà essere ottenuto per assicurare la qualità desiderata, tenendo conto della sicurezza e dell'efficacia del prodotto farmaceutico stesso" [65]. L'applicazione di questa definizione ai metodi analitici separativi implica che per prima cosa dovrebbero essere ben definiti gli obiettivi della separazione. Anche se gli obiettivi specifici possono variare, in generale il target dell'ottimizzazione è la massima selettività ottenuta in tempi minimi, il tutto entro un'area di lavoro ben conosciuta e robusta [89]. Per un API questo implica la separazione dalle impurezze insieme al soddisfacimento di specifici requisiti normativi, ad esempio riguardanti i dati di convalida così come definiti dalla ICH [65].

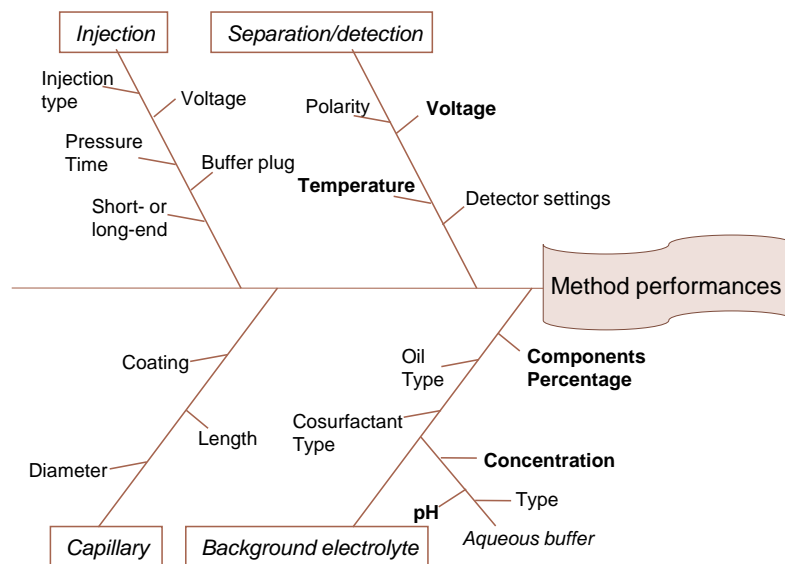
I CQA sono definiti invece come "proprietà o caratteristiche fisiche, chimiche, biologiche o microbiologiche che dovrebbero essere comprese entro un limite, range o distribuzione appropriata per assicurare la qualità del prodotto desiderata" [65]. Nei metodi analitici, i CQA sono variabili chiave di risposta, cosicché il cromatogramma o l'elettroferogramma possano essere relazionati alla rappresentazione matematica della qualità. I CQA possono essere classificati in due principali categorie: "deve avere" e "possibilmente dovrebbe avere". Per esempio, la risoluzione di una coppia di picchi adiacenti non deve avere un valore minore di 1.5 e preferenzialmente dovrebbe avere un valore superiore a 1.8. Alcuni esempi di CQA sono le risoluzioni critiche R_s , il tempo di analisi, l'efficienza e, in contrasto con l'approccio tradizionale QbT, la robustezza. Recentemente è stato anche introdotto il criterio di separazione S [91, 92], definito come la differenza tra i tempi di ritenzione misurati all'inizio del secondo picco e alla fine del

primo picco della coppia critica. Gli stessi autori hanno preferito S a R_s per definire la qualità della separazione, dato che il modellamento di R_s può portare ad una predittività carente, quando la selettività (ordine) dei picchi cambia drasticamente. Inoltre, anche se S e R_s sono altamente correlati, il computo di S è risultato più semplice e con una minore incertezza associata.

Identificazione dei parametri critici del processo (CPP) attraverso la valutazione dei rischi della qualità (QRA)

I parametri critici del processo (*Critical Process Parameters*, CPPs) sono stati definiti come “parametri le cui variazioni hanno un impatto sui CQA e quindi devono essere monitorati o controllati per assicurare che il processo produca la qualità desiderata” [73]. Questa definizione si adatta perfettamente anche ai metodi analitici. Lionberger et al. hanno proposto una classificazione dei parametri in tre categorie: non classificati, critici o non critici. La criticità di un parametro non classificato è indeterminata o sconosciuta, mentre un parametro viene definito critico quando un cambiamento dello stesso può portare ad un prodotto che non soddisfa i QTPP. Nella fase di sviluppo si dovrebbero quindi condurre studi che portino a convertire i parametri non classificati in critici o non critici, altrimenti potrebbe essere necessario vincolarli a valori fissati o mantenerli in intervalli ristretti dato che potrebbero essere critici [94]. La lista iniziale dei parametri che possono influenzare i CQA può essere piuttosto vasta, ma può essere ridotta effettuando una valutazione della loro importanza per mezzo della valutazione dei rischi della qualità (*Quality Risk Assessment*, QRA). QRA è un processo scientifico che porta all’identificazione dei parametri critici del processo e quindi all’eliminazione dei rischi, con la conseguenza di un’alta probabilità che il metodo analitico soddisfi i QTPP in tutte le condizioni d’uso [94]. Un grande numero di parametri può essere quindi eliminato in modo sicuro attraverso gli strumenti del QRA, ad esempio grazie all’analisi degli effetti delle modalità di fallimento (*Failure Mode Effects Analysis*, FMEA) [74] e ai diagrammi di Ishikawa [74, 95]. Nella FMEA le variabili sono classificate sulla base della possibilità di fallimento (probabilità), dell’influenza sui risultati analitici (severità) e sulla difficoltà di rilevazione (rilevabilità), confluendo poi in un numero di priorità di rischio (*Risk Priority Number*, RPN). I fattori con un valore di RPN sopra il livello limite devono essere esaminati

approfonditamente, mentre i fattori con un basso valore di RPN possono essere eliminati da studi successivi [87]. I diagrammi di Ishikawa classificano i rischi in categorie differenti: errori associati alla strumentazione, ai materiali, ai metodi, alle misure, al clima del laboratorio, ai fattori umani. Recentemente è stato presentato un diagramma di Ishikawa applicato ad un metodo HPLC per il controllo delle impurezze, mettendo in evidenza i diversi tipi di fattori [94]. È stata quindi eseguita un'analisi CNX per decidere quali fattori dovessero essere controllati (control, C), quali potessero essere fonti di rumore (noise, N), e su quali dovessero essere condotti esperimenti per determinare range accettabili [94]. In generale è stato possibile mettere in evidenza che per la grande maggioranza delle separazioni HPLC le condizioni critiche sono il gradiente, la temperatura, il pH della fase acquosa, i modificatori organici e la fase stazionaria [96]. Nella figura sotto riportata è invece rappresentato un generale diagramma di Ishikawa utile per lo sviluppo di un metodo di cromatografia elettrocinetica con microemulsione (MEEKC) [97]. Sono evidenziati in grassetto i CPP che vengono tipicamente valutati attraverso DoE; gli altri parametri sono di solito fissati attraverso esperimenti preliminari o sulla base di conoscenze pregresse.



Ricerca dello spazio di conoscenza (KS) mediante disegno degli esperimenti (DoE)

Dopo la fase QRA, possono essere effettuati studi di screening attraverso l'analisi degli effetti condotta con il DoE. In questo modo è possibile valutare l'effetto dei CPP di più alto livello nell'analisi QRA e ridurre ulteriormente il numero dei CPP

stessi selezionando i fattori che hanno maggiore effetto [98]. Gli strumenti del QRA, la conoscenza a priori ed esperimenti preliminari sono quindi usati per definire lo spazio di conoscenza (*Knowledge Space*, KS), che è la regione sperimentale investigata dal DoE [66,98]. Potrebbero essere necessari ulteriori esperimenti preliminari al fine di una definizione adeguata del dominio sperimentale [9], perché la scelta di un KS idoneo minimizza il rischio di non ottenere una buona separazione [91]. Lo scopo è quello di eseguire un numero minimo di esperimenti su un numero massimo di fattori, così i modelli relativi agli effetti principali sono stimati nella maggior parte dei casi mediante i disegni di Plackett-Burman e disegni frazionari a due livelli [85]. Per esempio, è stato utilizzato un disegno di Plackett-Burman per effettuare uno studio di screening sugli effetti del pH della fase mobile, della temperatura, della velocità di flusso, del volume di iniezione e della concentrazione di metanolo sulla risoluzione e sul fattore di scodamento. I diagrammi di Pareto che sono stati ottenuti hanno messo in evidenza che i fattori staticamente significativi erano il pH, la temperatura e il volume di iniezione, e questi sono stati quindi ulteriormente studiati [100]. La fase di screening viene anche tipicamente utilizzata per la valutazione di fattori discontinui quali colonne, scelta del solvente, scelta del tampone. Nella successiva metodologia dello studio della superficie di risposta (*Response Surface Methodology*, RSM) vengono ottimizzati i fattori continui quali temperatura e programmazione del gradiente [90]. La fase di screening attraverso il DoE può avere come risposta il numero dei picchi, mentre nella RSM si può avere un perfezionamento degli attributi di qualità correlati alla risoluzione di specifici componenti. Di conseguenza, l'automatizzazione nel *peak tracking* è recentemente diventato uno strumento molto utile nello screening [90]. Talvolta l'ottimizzazione può partire direttamente dalla RSM in quanto la fase di screening può essere eliminata grazie a esperimenti preliminari o alla conoscenza a priori [88, 89, 99].

Nello studio della superficie di risposta possono essere utilizzati diversi disegni per valutare gli effetti principali, quadratici e/o di interazione dei vari fattori, quali Box-Behnken, D-optimal, central composite, disegno fattoriale saturo, disegno fattoriale frazionario o disegno di Doehlert [85, 86]. Awotwe-Otoo et al. hanno applicato un disegno di Box-Behnken per studiare l'effetto del pH, della temperatura e del volume di iniezione sulla risoluzione e sul fattore di scodamento dei picchi [100]. Debrus et al. hanno usato un disegno fattoriale saturo per ottimizzare il metodo e simultaneamente

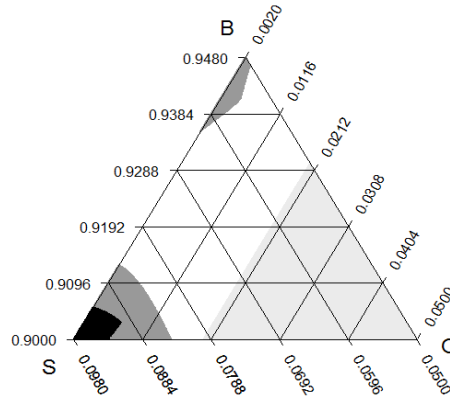
stimare la sua robustezza, considerando come fattori il gradiente, la temperatura e il pH. Gli stessi autori hanno anche suggerito che quando l'unico obiettivo è l'ottimizzazione del metodo si possono utilizzare disegni più semplici come il disegno fattoriale saturo o il central composite design [91]. Il disegno fattoriale saturo è stato applicato anche da Molnar et al. per valutare gli effetti del gradiente, della temperatura, del pH e della composizione ternaria [89]. È stato proposto anche un altro approccio di DoE [101, 102], in cui la prima fase di sviluppo del metodo include lo studio di diverse colonne analitiche insieme alle condizioni critiche, per esempio pH e solvente organico, mentre in una seconda fase vengono valutati i fattori strumentali rimanenti. La perdita di dati nella prima fase, casuata dal possibile cambiamento del pattern di eluizione dei picchi, può essere risolta usando il DoE accoppiato al calcolo automatizzato di risposte legate all'andamento generale dell'analisi, che possono essere di due tipi: basate sul conteggio dei picchi e basate sui parametri dei picchi stessi [101,102]. Infine, deve essere menzionato il disegno di miscele, definito come uno speciale RSM nel quale la risposta è funzione della diversa proporzione dei componenti [21]. In questo ambito, il modello speciale cubico ridotto di Scheffè con una matrice a tredici esperimenti è stato ampiamente utilizzato per l'ottimizzazione della composizione di una microemulsione nei metodi CE per l'analisi di prodotti farmaceutici, seguita da uno studio della superficie di risposta per ottimizzare i fattori di processo come concentrazione di ciclodestrina, temperatura, voltaggio e concentrazione della soluzione tampone [10,11, 103]. Questo tipo di modello e di matrice sono stati inoltre utilizzati per identificare il range in cui si ha la formazione di una microemulsione stabile, invece di usare i classici diagrammi di fase pseudoternaria [104]. Piepel et al. hanno per la prima volta in letteratura utilizzato una matrice mista di variabili miscela-processo (Mixture Process Variables Approach, MPV) nello sviluppo di un metodo MEEKC, per valutare simultaneamente gli effetti della composizione della microemulsione (fase oleosa, fase acquosa, miscela tensioattivo/cotensioattivo) e delle variabili di processo (differenza di potenziale, pH, concentrazione di tampone) sull'efficienza dei picchi [88]. Questo nuovo approccio ha permesso di identificare significativi effetti di interazione tra le variabili.

Definizione dello spazio del disegno

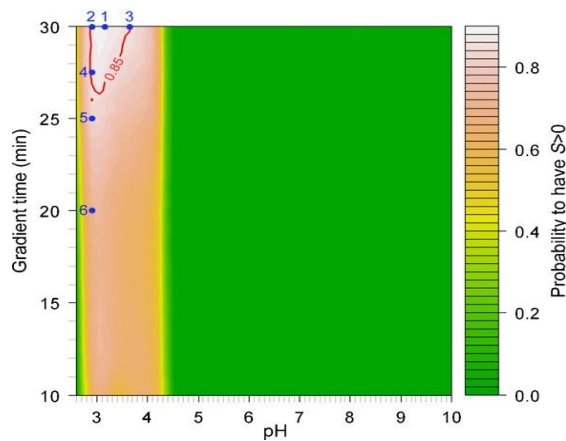
Lo spazio del disegno (*Design Space*, DS) per i metodi analitici è stato definito come una regione multidimensionale che include ogni combinazione dei valori assunti dalle variabili che è stata dimostrata assicurare la qualità dei dati prodotti dal metodo [87, 94]. Il DS è delimitato dai cosiddetti confini di fallimento (*Edges of Failure*, EoF), fuori dai quali le prestazioni del metodo non sono accettabili [66]. Di conseguenza il metodo analitico dovrebbe essere sviluppato e convalidato non solo in una particolare condizione fissata, ma in un intervallo di condizioni [96, 98]. Variazioni all'interno del DS non vengono intese come cambiamenti del metodo e non necessitano di alcuna ulteriore approvazione normativa [73]. È inoltre possibile sia “stabilire spazi del disegno indipendenti per una o più operazioni unitarie, sia stabilire un singolo spazio del disegno che si estenda su operazioni multiple” [73]. Questo ha portato Monks et al. a proporre che in cromatografia il DS possa essere considerato come due entità: lo spazio del disegno per la colonna (*Column Design Space*, CDS) e lo spazio del disegno dell'eluente (*Eluent Design Space*, EDS). Una volta quindi trovata una colonna robusta con una sua equivalente, la definizione del DS per tutti gli altri fattori è stata effettuata dagli stessi autori sulla colonna selezionata.

Il DS può essere rappresentato graficamente o può essere descritto matematicamente con l'uso di equazioni che descrivano le relazioni tra i parametri che portano a soddisfare i requisiti [73]. Quando sono coinvolte più risposte, gli approcci più frequentemente utilizzati sono la sovrapposizione delle risposte [66, 101] e la funzione di desiderabilità [105-107]. Nella sovrapposizione delle risposte, gli obiettivi delle diverse risposte sono rappresentati in un grafico, calcolato da un software. Negli studi di desiderabilità, la qualità totale è riassunta grazie all'utilizzo di un indice D di desiderabilità. Questo viene ottenuto assegnando valori desiderati e accettabili ai CQA e calcolando le desiderabilità parziali d_i . Queste possono variare tra 0 e 1, dove “1” corrisponde ad un valore altamente desiderabile e “0” a un valore non accettabile. I valori che aumentano da 0 a 1 esprimono un aumento nella desiderabilità del criterio. Tutte le desiderabilità parziali d_i vengono quindi riunite nella desiderabilità totale D , che deve essere massimizzata [106]. Piepel et al. hanno specificato un valore limite accettabile e uno desiderato per ciascuna risposta. Sulla base dei modelli MPV calcolati,

il DS è stato definito mediante i limiti accettabili, e all'interno del DS è stata definita la regione sub ottimale mediante i limiti ottimali. Sia il DS che la regione sub ottimale erano regioni pentadimensionali e non potevano essere illustrati nella loro totalità; la figura sotto riportata mostra il DS e la regione sub ottimale per una combinazione selezionata delle variabili di processo [88].



In diversi lavori è stato inoltre puntualizzato che quando si definisce il DS occorre prendere in considerazione l'incertezza del modello, valutando l'errore nei criteri; altrimenti non vengono fornite indicazioni su quanto bene e con quale frequenza il metodo soddisfi le specifiche [91,105 , 107, 108]. Così la confidenza nelle condizioni ottimali predette è stata valutata attraverso uno studio di Monte-Carlo per la valutazione dell'errore predittivo del modello [91, 107]. A titolo di esempio, Debrus et al. hanno valutato per ciascuna condizione operativa la probabilità stimata che i CQA soddisfacessero simultaneamente i requisiti, e i risultati sono stati mostrati come superfici di probabilità, come si può vedere nella seguente figura.



In questo caso il DS è stato determinato considerando la probabilità che il valore di S sia più alto rispetto a un valore selezionato ($\pi(S>0)$), definito come soglia di qualità. In questo modo la qualità nel DS non viene soltanto predetta, ma viene anche assicurata attraverso la stima del rischio di non rientrare nei limiti accettabili. Non vi sono comunque ancora documenti normativi che definiscano come stimare il livello di qualità richiesto [91].

Selezione delle condizioni operative

All'interno del DS esistono molteplici condizioni operative (*working points*) ottimali tra cui scegliere. La condizione operativa può essere selezionata scegliendo l'optimum con un alto indice di desiderabilità [100] o il punto in cui si ha la più alta probabilità che i CQA soddisfino le specifiche. È inoltre possibile selezionare le condizioni operative, siano esse nel DS o nella regione sub ottimale, sulla base di considerazioni pratiche, dato che l'unico requisito richiesto è infatti che questo punto si trovi all'interno del DS [88].

Una importante valutazione per questa scelta consiste nel tener conto della robustezza, quindi nella diretta integrazione di questo parametro di convalida nello sviluppo del metodo. In questo caso, più che trovare un singolo massimo di una funzione, l'obiettivo è localizzare un *plateau* di risposta. In alcuni casi le condizioni ottimali possono essere direttamente selezionate dai diagrammi CQA, i quali possono aiutare a rivelare aree di bassa o alta robustezza [90]. Per esempio, Monks et al. hanno selezionato condizioni di lavoro robuste osservando direttamente le superfici di risposta, secondo i CQA risoluzione critica, robustezza e tempo di analisi [96, 109]. Swartz et al. hanno suggerito che i metodi in prossimità degli EoF risultano avere prestazioni accettabili di media, quindi questi confini dovrebbero essere spostati all'interno del DS per soddisfare la robustezza e definire uno "spazio operativo di processo" [66]. Una simulazione Monte-Carlo è stata quindi utilizzata per fornire, per un dato punto candidato come condizione operativa, una misura quantitativa della variazione della risposta, definita come indice di capacità del processo (*Process Capability Index, C_p*). I valori di robustezza C_p sono stati modellati come ulteriore risposta e sono stati identificati i valori dei fattori che portavano simultaneamente a soddisfare i requisiti di prestazione del metodo e gli obiettivi di robustezza [66].

Robustezza

Fino a tempi recenti, il test di robustezza veniva eseguito durante la convalida del metodo e spesso portava a rilevare problemi tali che il metodo doveva essere ottimizzato di nuovo. Nel QbD, la robustezza può essere direttamente inclusa nelle qualità desiderate del metodo e valutata durante lo sviluppo del metodo mentre si selezionano il DS e le condizioni operative ottimali [90, 109]. Il DS stesso può essere considerato come una zona di robustezza teorica, poiché al suo interno la variazione delle condizioni del metodo non influenza in maniera significativa la qualità della separazione [91].

Sono stati utilizzati diversi approcci per stimare la robustezza del metodo. Come menzionato precedentemente, il DoE con la simulazione di Monte-Carlo è stato usato per integrare un modello quantitativo di robustezza all'interno dello sviluppo del metodo cromatografico [66]. Un'altra possibilità è quella di eseguire un DoE intorno alla condizione operativa ottimale, per esempio utilizzando una matrice di Plackett-Burman. Gavin et al. hanno scelto un disegno fattoriale frazionario a due livelli per valutare gli effetti della composizione della fase mobile e della temperatura su tre aspetti critici delle prestazioni del metodo (tempo di analisi, fattore di scodamento, retropressione) [84]. Monks et al. hanno visualizzato la robustezza del metodo come un corpo geometrico dentro lo spazio della risoluzione nel quale la risoluzione critica non cade sotto un determinato valore. Hanno poi utilizzato un disegno fattoriale saturo centrato sulle condizioni ottimali per valutare la robustezza dei punti candidati [96]. Gli stessi autori hanno determinato i valori di gradiente, temperatura e pH mediante lo studio visuale delle superfici di risposta, confermando le tolleranze con 12 nuovi esperimenti che hanno fornito buoni valori di risoluzione. I 12 cromatogrammi sono stati poi utilizzati per costruire un nuovo cubo più piccolo che dimostrava di produrre una risoluzione critica maggiore di 1.5 nell'intero range. Gli intervalli di tolleranza per i fattori misurati e calcolati (velocità di flusso, percentuale di eluente finale ed iniziale) sono stati poi valutati con un software senza il bisogno di ulteriori esperimenti. Un diverso software è stato usato per valutare l'equivalenza della colonna [89].

Molnar et al. hanno messo in evidenza che le posizioni delle coppie critiche di picchi potrebbero essere non costanti nei cromatogrammi, indicando la presenza di diversi

picchi critici in movimento. È stata quindi esaminata ciascuna coppia critica con il proprio diagramma di risoluzione e si è scoperto che ogni regione robusta è costituita da diverse sub regioni corrispondenti a differenti coppie di picchi critiche. Per questo motivo è stato ritenuto insufficiente e non pratico fissare il metodo in relazione ad una singola coppia critica perché piccoli cambiamenti nelle condizioni sperimentali potevano cambiare la coppia critica stessa. È stata invece suggerita come più ragionevole la definizione del DS per diverse coppie di picchi critici in modo tale che la loro risoluzione superasse un determinato valore [66].

Controllo del metodo

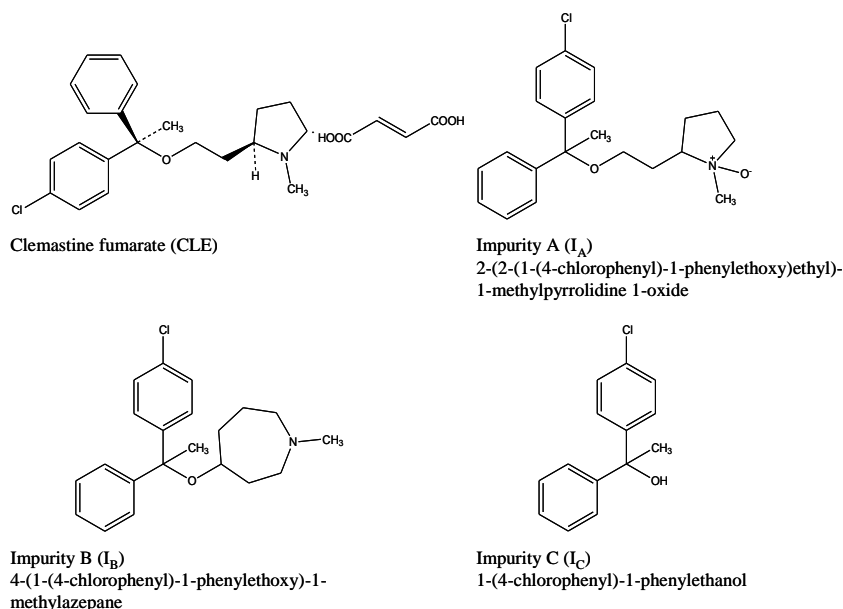
Dovrebbe infine essere progettata una strategia di controllo (*Control Strategy*, CS) per assicurare che un prodotto di qualità richiesta venga regolarmente prodotto [73]. La 75 definisce la CS come “una serie pianificata di controlli, derivati dalla comprensione di prodotti e processi, che assicura le prestazioni del processo e la qualità del prodotto” [75]. Con un adattamento analitico, la CS è stata definita come i controlli sui fattori di input che assicurano che il metodo soddisfi sia i tradizionali criteri di *system-suitability* che obiettivi più ampi legati alle prestazioni [94]. I test di *system-suitability* sono una parte standard dell’applicazione di routine e sono tipicamente stabiliti durante la convalida del metodo. Una strategia di controllo include normalmente appropriati *system suitability criteria* per gestire i rischi, per cui il QRA può essere d’aiuto nell’identificare una specifica strategia di controllo. In linea di principio, un appropriato *system suitability test* potrebbe essere il solo elemento di controllo necessario per assicurare le prestazioni del metodo selezionato [87], perché contribuisce ad identificare le modalità di fallimento e può prevenire la generazione di risultati erronei [84]. Un interessante esempio di CS nel QbD è stato elaborato da Gavin et al. [84]; tali autori, durante gli studi sulla robustezza di un metodo per il controllo di qualità di un principio attivo e relative impurezze, hanno rilevato che la risoluzione delle impurezze seguiva la stessa tendenza alla variazione di alcune condizioni del metodo quali temperatura e *n*-propanolo. Sono state così scelte una coppia di impurezze con basso tempo di ritenzione e su di esse sono stati eseguiti i test di *system suitability*, in quanto i due composti coinvolti erano facilmente reperibili. Questo caso è stato particolarmente fortunato; in generale i test di *system-suitability* devono essere

implementati con controlli addizionali per gestire i rischi in modo appropriato. In questo senso una strategia più comune potrebbe essere quella di definire dei range di accettabilità per tutti i valori di risoluzione critica. Ciò può essere ottenuto sulla base dei valori misurati durante gli studi di ripetibilità, in quanto gli studi di precisione forniscono informazioni importanti riguardo alla variabilità del sistema analitico [9, 10, 103].

7. Risultati e Discussione

7.1. Studi di Molecular Modeling e NMR per la valutazione dei meccanismi alla base delle separazioni in cromatografia elettrocinetica

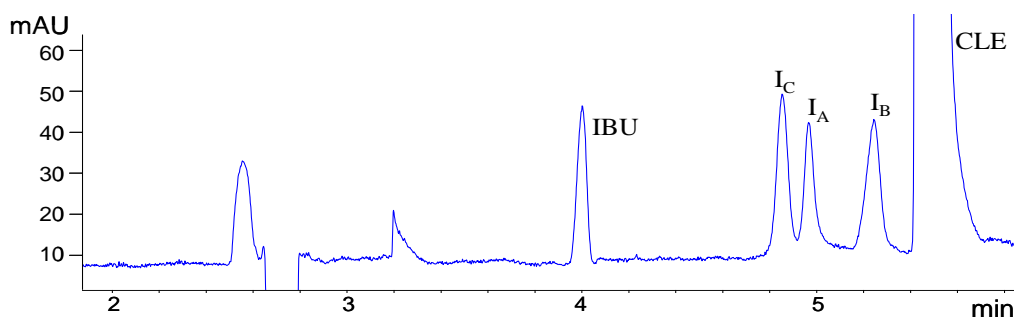
L'elettroforesi capillare non fornisce informazioni dirette per quanto riguarda i meccanismi di interazione intermolecolare tra analita e fase pseudostazionaria o tra analita e ciclodestrina. Per il controllo di qualità dell'antistaminico clemastina (CLE) per la prima volta è stato utilizzato nel nostro gruppo di ricerca un sistema costituito da una microemulsione e due ciclodestrine (metil- β -Cd (M β CD) e 2,6-dimetil- β -CD (DM β CD)) [103]. L'uso di un sistema così complesso è stato necessario in particolare per ottenere la separazione della clemastina dalla I_B, impurezza che presenta stesso peso molecolare e differisce solo per la posizione di due gruppi metilenici.



La miglior composizione della microemulsione è stata trovata con un disegno di miscele, mentre l'ottimizzazione delle variabili di processo concentrazione delle ciclodestrine e voltaggio è stata effettuata con uno studio della superficie di risposta. Le condizioni sperimentali ottimizzate sono state le seguenti: 10 mM tampone borato,

89.8%; *n*-eptano, 1.5%; SDS/*n*-butanolo, 8.7%, in rapporto 1:2; 18 mM M β CD, 38 mM DM β CD; 17 kV. Applicando queste condizioni la separazione completa degli analiti si ottiene in circa 5.5 minuti scegliendo come standard interno l'ibuprofene (IBU).

Anche in questo caso il metodo è stato convalidato secondo le linee guida ICH. Quindi, con l'obiettivo di valutare meglio i meccanismi alla base di questa separazione e di indagare la formazione effettiva di due altre fasi pseudostazionarie dovute alle ciclodestrine, è stato effettuato uno studio di modellistica molecolare e di NMR.



Molecular Modeling

Sono stati condotti studi di *molecular modeling* sia in solvente implicito che esplicito e sono state calcolate le energie di *docking* (somma di interazioni di Van Der Waals ed elettrostatiche) dei complessi di inclusione analita/CD ricercando per ogni complesso la conformazione più stabile. Ogni simulazione è iniziata dalla conformazione più probabile del complesso ospite-CD (sei per la Clemastina e impurezza B, due per SDS, vedi figura). Per le simulazioni in solvente implicito è stato evidenziato come per la DM β CD si abbiano dei valori di energia libera di *docking* statisticamente diversi tra I_B-DM β CD e gli altri due complessi (in tutti i casi $p < 0.05$). Mentre la differenza di energia tra il complesso CLE-DM β CD e SDS-DM β CD non è significativamente diversa ($p = 0.53$).

a)	Mean (Water)	SD (Water)	Mean (Solvent)	SD (Solvent)
<i>CLE-DMβCD</i>	-25.5	2.3	-17.8	2.2
<i>I_B-DMβCD</i>	-27.0	2.4	-19.1	2.3
<i>SDS-DMβCD</i>	-22.5	0.9	-17.6	0.8
<i>CLE-MβCD</i>	-22.2	1.2	-15.1	1.7
<i>I_B-MβCD</i>	-22.6	3.0	-15.0	1.6
<i>SDS-MβCD</i>	-20.4	1.0	-16.3	0.7

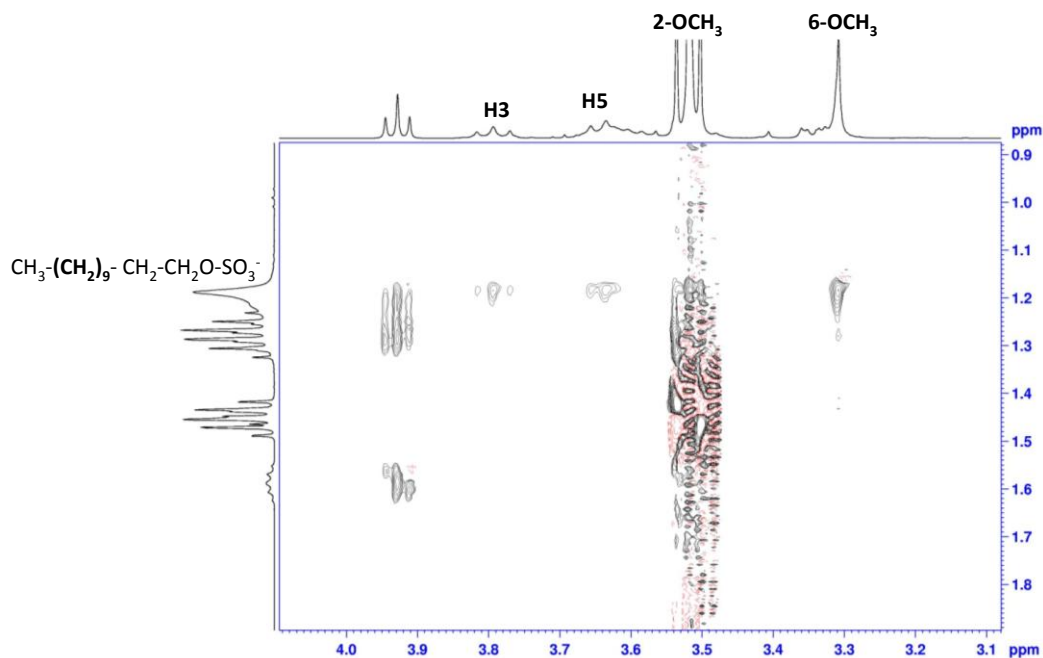
Anche in questo caso l'energia per il complesso con I_B è più favorita. I ΔG dei complessi degli analiti con la stessa ciclodestrina, pur essendo nel caso della DMβCD statisticamente diversi, risultano comunque molto vicini (1.4 Kcal/mol circa di differenza media). Questa esigua differenza non è sufficiente, da sola, a spiegare la separazione elettroforetica che si ottiene sperimentalmente in presenza contemporanea dell'SDS. Infatti è sperimentalmente dimostrato che in presenza della sola ciclodestrina, nel nostro caso DMβCD, non si ha alcuna separazione. Le simulazioni in solvente esplicito hanno fornito analoghi risultati.

Si è quindi ipotizzato che il ruolo del SDS sia quello di “amplificare” le differenze di stabilità tra i complessi CLE-DMβCD e I_B-DMβCD entrando in competizione con gli analiti per lo stesso substrato, la DMβCD, per la formazione del complesso.

NMR

Per confermare i risultati del molecular modeling sono stati effettuati spettri 2D-ROESY. In primo luogo sono stati eseguiti esperimenti NMR per stabilire l'esistenza di interazioni tra le ciclodestrine e i componenti della microemulsione (n-eptano, n-butanolo e SDS). Per entrambe le ciclodestrine investigate si è riscontrata la formazione di complessi solo con il tensioattivo sodio dodecilsolfato. La soluzione in D₂O di DMβCD e SDS ha mostrato una forte interazione tra i metileni della coda lipofila del sodio dodecilsolfato, i protoni interni (H3-H5) della ciclodestrina ed anche con entrambi i metossili in posizione 2-6. Questo comportamento ha suggerito che la coda lipofila del

tensioattivo si introduce all'interno della cavità della ciclo destrina lasciando la testa idrofila all'esterno. Vista l'asimmetria dell'SDS, si possono ottenere due tipi di complessi: uno con il metile terminale adiacente ai metossili in posizione 2 e l'altro con il metile adiacente ai metossili in posizione 6.



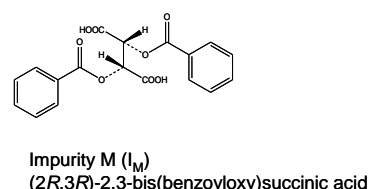
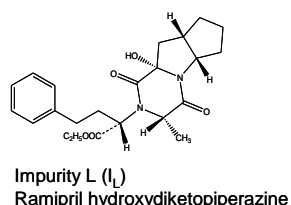
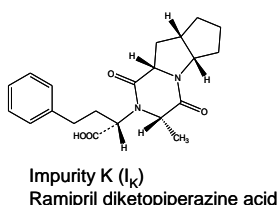
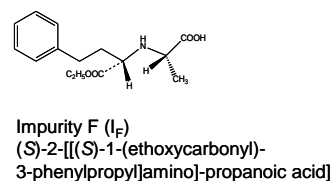
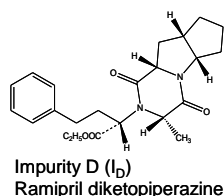
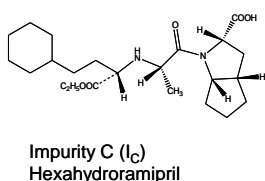
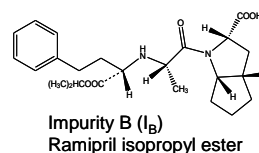
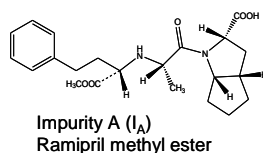
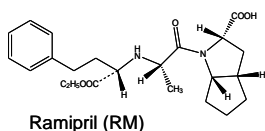
Spettro 2-D ROESY tra la microemulsione e 20mM DMβCD in D₂O

Dallo spettro appare evidente che entrambi i complessi sono possibili poiché il -CH₃ dell'SDS mostra un effetto NOE con entrambi i metossili. Gli esperimenti NMR condotti in presenza della microemulsione hanno inoltre mostrato che non ci sono interazioni tra DMβCD e l'eptano, solvente organico che costituisce il cuore delle gocce oleose, permettendo di confermare la formazione di seconde fasi pseudo stazionarie che si trovano nella fase acquosa della microemulsione. Contemporaneamente è stata evidenziata un'interazione tra DMβCD e SDS presente nella fase acquosa, confermando l'ipotesi di una distribuzione della ciclodestrina nel carrier al di fuori della fase oleosa.

7.2. Sviluppo di un metodo CE per l'analisi del Ramipril e delle sue impurezze

L'obiettivo di questo lavoro è stato lo sviluppo di un metodo CE per la determinazione di RM e delle sue otto impurezze in forma farmaceutica, che possa essere usato per analisi di routine nel controllo di qualità.

Le formule di struttura di RM e delle sue impurezze sono di seguito riportate.



Tra gli analiti considerati, RM, I_A, I_B, I_C, I_F hanno natura zwitterionica, possedendo un gruppo carbossilico ed un gruppo amminico secondario. I_D e I_L sono neutre e differiscono tra loro per un solo gruppo ossidrilico. I_K e I_M hanno caratteristiche acide, presentando rispettivamente uno e due gruppi carbossilici. Quindi, per effettuare l'analisi simultanea di questi composti occorre scegliere una adatta fase pseudostazionaria, sia per le diverse proprietà acido-base degli analiti stessi, sia per il fatto che gli analiti neutri sono privi di una propria mobilità elettroforetica e quindi non possono essere separati con una semplice CZE.

Inoltre, quattro analiti (RM, I_A, I_B e I_C) contengono un residuo simile alla prolina che causa la presenza di isomeri cis e trans a livello del legame amidico. Di conseguenza, per la determinazione di questi composti potrebbero essere teoricamente utilizzati due diversi *pattern* elettroforetici, relativi ad un diverso numero di

Damköhler: a) basso numero di Damköhler. In questo caso l'isomerizzazione è sufficientemente lenta da far sì che l'interferenza sia trascurabile, per cui i due conformeri sono separati con risoluzione alla linea di base; b) alto numero di Damköhler. In questo caso l'isomerizzazione è rapida rispetto alla scala temporale della migrazione, per cui gli isomeri cis e trans migrano come un singolo picco.

Questa situazione pone una significativa sfida analitica, perché tra le condizioni a) e b) si può verificare una condizione indesiderata, descritta da un numero di Damköhler intermedio, dove è presente una regione *plateau* tra i picchi con conseguente deterioramento dell'efficienza di separazione. Questa regione di interferenza rappresenta un problema soprattutto a livello di RM, perché la concentrazione test finale di questo composto deve essere circa cento volte più alta di quella delle impurezze correlate. Invece, per le impurezze di RM il *plateau* non è rilevabile alle concentrazioni test e quindi non rappresenta un reale problema.

Il numero di Damköhler dipende dalle diverse condizioni sperimentali, per cui la prima serie di esperimenti effettuati aveva lo scopo di stabilire i requisiti analitici minimi del metodo, quali tipo di modalità operativa e di fase pseudostazionaria, che potessero condurre alla situazione ottimale a) o b). Tra queste due situazioni, in linea di principio probabilmente la migliore utile a risolvere il problema analitico considerato corrisponde alla situazione b). Infatti, dal punto di vista pratico dalla situazione a) potrebbero derivare due problemi, concernenti:

- una ridotta sensibilità per le impurezze, dato che lo *splitting* del picco dell'analita nei picchi corrispondenti ai due isomeri causa la diminuzione della altezza, e quindi la diminuzione del rapporto segnale-rumore con la conseguenza di limiti di rilevazione e di determinazione più alti;

- una aumentata complessità della separazione, dovuta alla rilevazione di isomeri cis e trans separati e quindi ad un numero maggiore di picchi da risolvere.

Polarità normale

Nella prima serie di esperimenti la differenza di potenziale è stata applicata in polarità normale.

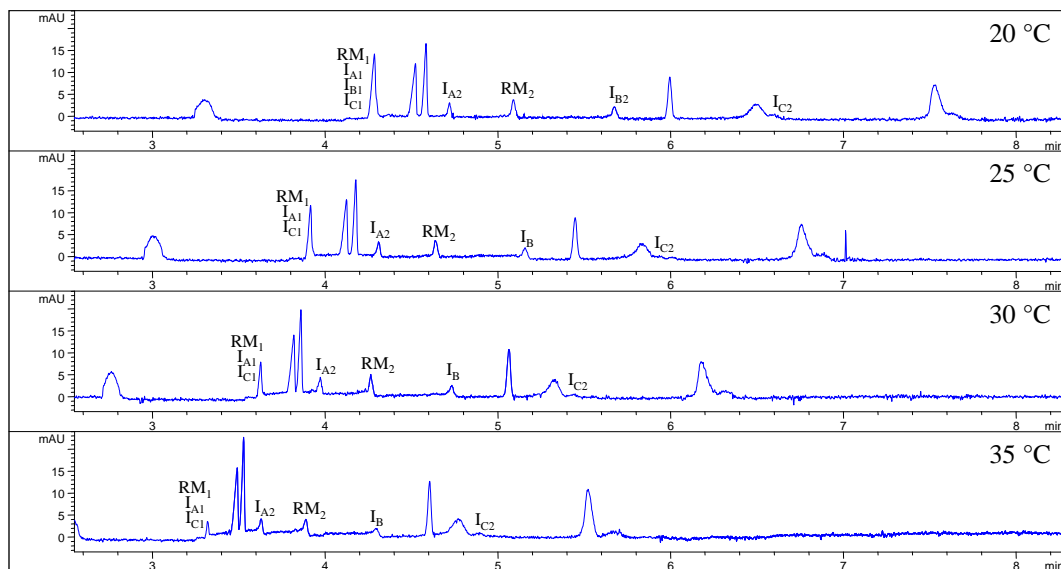
MEKC

I primi esperimenti MEKC hanno riguardato la scelta di un idoneo tampone che consentisse di ottenere la separazione di un totale di tredici picchi: le coppie di isomeri separati per RM, I_A, I_B e I_C e i picchi singoli per le rimanenti cinque impurezze.

È stato inizialmente valutato l'effetto del pH, parametro fondamentale nelle ottimizzazioni CE in quanto capace di modulare sia l'entità di EOF sia la dissociazione di analiti acidi e basi deboli. Sono stati esaminati BGE costituiti da borace 10 mM, SDS 30 mM con valori di pH compresi tra 8.5 e 10.0. La variazione di pH esercitava un effetto significativo sul numero dei picchi e la selettività più alta è stata ottenuta a pH 9.5.

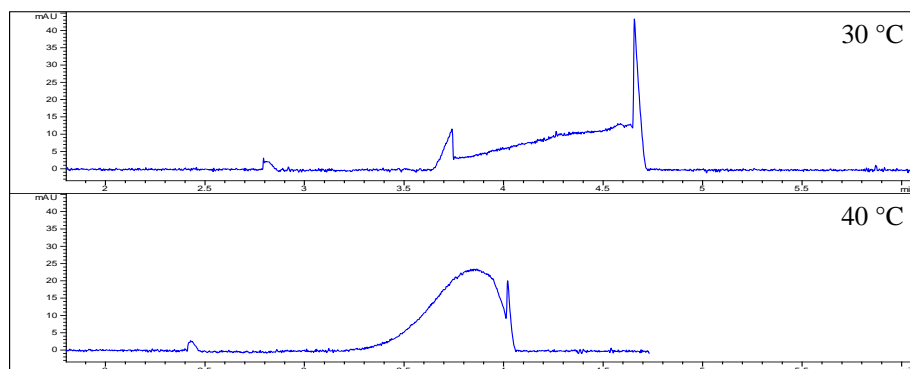
Sono stati quindi presi in considerazione altri sistemi tampone validi in un range di pH intorno a 9.5, inclusi il tampone borato, il tampone borato-fosfato e il tampone Britton-Robinson (questo ultimo costituito da una miscela di acido borico, acido acetico e acido fosforico), sempre mantenendo SDS ad una concentrazione pari a 30 mM. I migliori risultati sono stati ottenuti con tampone borato 20 mM, che ha dato la più alta selettività a valori di pH 9.50-9.75. L'uso di questo tampone ha anche portato ad un significativo miglioramento dell'efficienza degli ultimi picchi, quindi è stato scelto per costituire la base del BGE.

Per questo tipo di sistema non è stato intrapreso uno studio approfondito per arrivare ad un singolo picco per i due isomeri. Infatti, il fattore più importante che agisce sul numero di Damköhler è la temperatura; tuttavia, elevando la temperatura da 20 °C a 25 °C, solo gli isomeri relativi alla I_B sono stati portati a migrare come un picco singolo. Anche aumentando la temperatura fino 35 °C gli isomeri di RM e delle impurezze I_A e I_C continuavano a migrare separatamente; alzando ulteriormente la temperatura si osservava una grave e generale perdita di efficienza, non compatibile con la possibilità di ottenere una sufficiente sensibilità per le impurezze.



Elettroferogrammi di una soluzione degli analiti 0.04 mg mL^{-1} a diversi valori di temperatura. BGE, borato 20 mM pH 9.50 , SDS 30 mM . Lunghezza del capillare, 64.5 cm ; lunghezza d'onda di rivelazione, 210 nm ; differenza di potenziale, 30 kV

Per quanto riguarda RM, utilizzando un campione a concentrazione elevata (2 mg mL^{-1}) a $20\text{-}30 \text{ }^\circ\text{C}$ veniva chiaramente evidenziata la zona di *plateau* corrispondente all'interferenza tra migrazione e isomerizzazione. Aumentando la temperatura fino a $60 \text{ }^\circ\text{C}$ si osservava un graduale collasso dei due picchi che tuttavia non portava ad un picco sufficientemente simmetrico ed efficiente da consentirne l'uso per un'accurata analisi quantitativa.



Elettroferogramma di una soluzione di RM 2 mg mL^{-1} . BGE, borato 20 mM pH 9.50 , SDS 30 mM . Lunghezza del capillare, 64.5 cm ; lunghezza d'onda di rivelazione, 210 nm ; differenza di potenziale, 30 kV

MEEKC

Nell'ottica di completare lo studio condotto in polarità normale è stata valutata anche la possibilità di utilizzare come BGE una microemulsione. Come prova preliminare è stata scelta una microemulsione standard precedentemente impiegata come punto di partenza per ulteriori studi di ottimizzazione coinvolgenti analiti aventi diverse caratteristiche acido-base [7, 10]. Tale microemulsione era costituita dal 90.95% di fase acquosa (borace 10 mM pH 9.2), 1.05% di fase oleosa (*n*-eptano), 8.00% di tensioattivo/cotensioattivo (SDS/*n*-butanolo) in rapporto 1:2. In lavori precedenti è stato mostrato come questa microemulsione avesse alta versatilità ed alto potere risolutivo [7, 10], tuttavia in questo caso i risultati ottenuti erano lontani dalle prestazioni desiderate. Non è stato quindi intrapreso uno studio più approfondito per ottimizzare la composizione della microemulsione.

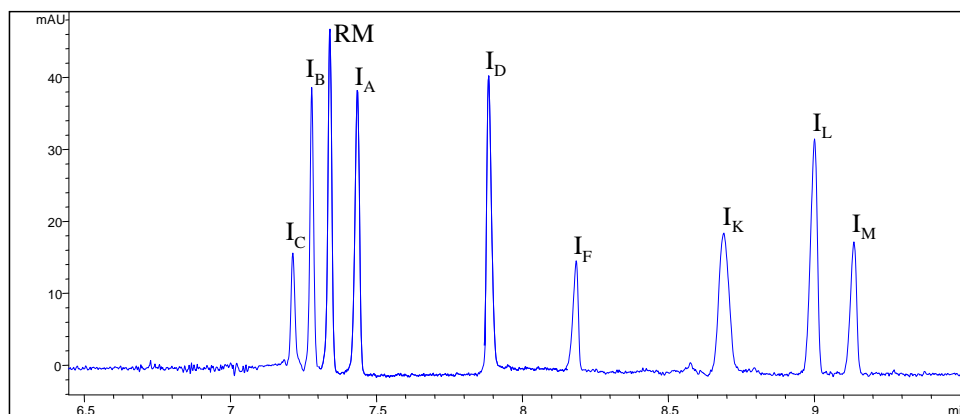
Polarità inversa

Nella successiva serie di esperimenti la differenza di potenziale è stata applicata in polarità inversa.

In questo caso, è stato selezionato come elettrolita un tampone fosfato 100 mM pH 2.5 a cui sono state aggiunte fasi pseudostazionarie a base di micelle e di microemulsioni. Le prime analisi hanno evidenziato che il cambiamento del pH del BGE rispetto alla polarità normale aveva un importante effetto sulla interconversione *cis-trans*, favorendo la situazione b). Infatti, lavorando a pH acido tutti i composti che davano luogo alla isomerizzazione *cis-trans* migravano come singolo picco.

Nel caso della MEKC, usando un sistema a base di SDS a concentrazione 50-100 mM si è osservata una selettività molto bassa per quanto riguarda la separazione dei picchi relativi a RM, I_A, I_B, I_C. La sovrapposizione di tali picchi era pressoché completa anche a valori di concentrazione degli analiti 0.04 mg mL⁻¹. Questo sistema non sembrava quindi promettente, ed è stato scartato senza approfondire la sua modulazione.

Nel caso della MEEKC, già dai primi elettroferogrammi sono stati ottenuti buoni risultati. Come prova preliminare è stata presa in considerazione la microemulsione usata valutando la MEEKC in polarità normale, con la sola differenza costituita dal tipo di fase acquosa, cioè un tampone fosfato 100 mM pH 2.5 invece di borace 10 mM pH 9.2.



Elettroferogramma di una soluzione degli analiti 0.04 mg mL^{-1} . BGE, microemulsione composta da 90.95% fase acquosa (fosfato 100 mM pH 2.5), 1.05% fase oleosa (*n*-eptano), 8.00% tensioattivo/cotensioattivo (SDS/*n*-butanolo) in rapporto 1:2. Lunghezza del capillare, 64.5 cm; lunghezza d'onda di rivelazione, 210 nm; differenza di potenziale, -25 kV; temperatura, $20 \text{ }^\circ\text{C}$

L'elettroferogramma ottenuto analizzando il campione degli analiti a concentrazione 0.04 mg mL^{-1} ha evidenziato una risoluzione alla linea di base di tutti i composti in 9 minuti circa. Tale sistema è stato quindi scelto come punto di partenza per effettuare l'ottimizzazione multivariata.

Ottimizzazione multivariata della modalità operativa MEEKC

Durante l'ottimizzazione multivariata il campione era costituito da una miscela test contenente 4 mg mL^{-1} di RM e 0.04 mg mL^{-1} delle impurezze (1% rispetto al componente principale), cioè valori di concentrazione utili per la successiva analisi del campione reale. Il campione conteneva anche lo standard interno acido flufenamico (FL), scelto per le sue caratteristiche di assorbimento e perché nell'elettroferogramma migrava prima di tutti gli altri picchi senza problemi di interferenza.

Nella prima fase dell'ottimizzazione è stata selezionata la composizione ottimale della microemulsione con l'ausilio del disegno di miscele. Nella seconda fase è stato effettuato uno studio della superficie di risposta per fattori indipendenti, in modo da individuare il valore ottimale della concentrazione di fosfato, della differenza di potenziale e della temperatura.

Disegno di miscela

La prima fase dell'ottimizzazione del metodo MEEKC in polarità inversa ha previsto l'uso del disegno di miscele per trovare la migliore composizione della microemulsione. In particolare, i componenti del sistema erano i seguenti: fase acquosa, tampone fosfato 100 mM pH 2.5; fase oleosa, *n*-eptano; tensioattivo, sodio dodecilsolfato; cotensioattivo, *n*-butanolo. Gli ultimi due sono stati usati in rapporto 1:2 e quindi considerati come un unico componente.

Lo scopo era di trovare le proporzioni tra i componenti della miscela (X_1 , fase acquosa (W); X_2 , fase oleosa (O); X_3 , tensioattivo/cotensioattivo, S/CoS) che permettessero la completa separazione degli analiti, considerando quindi come risposte le risoluzioni critiche tra i picchi. Ad ogni componente è stato imposto un limite superiore ed uno inferiore, riportati nella tabella seguente.

	Componente ^[a]	Limite inferiore	Limite superiore
X_1	W	88.00%	93.90%
X_2	O	0.10%	2.00%
X_3	S/CoS	6.00%	10.00%

^[a]La somma dei componenti è pari al 100%

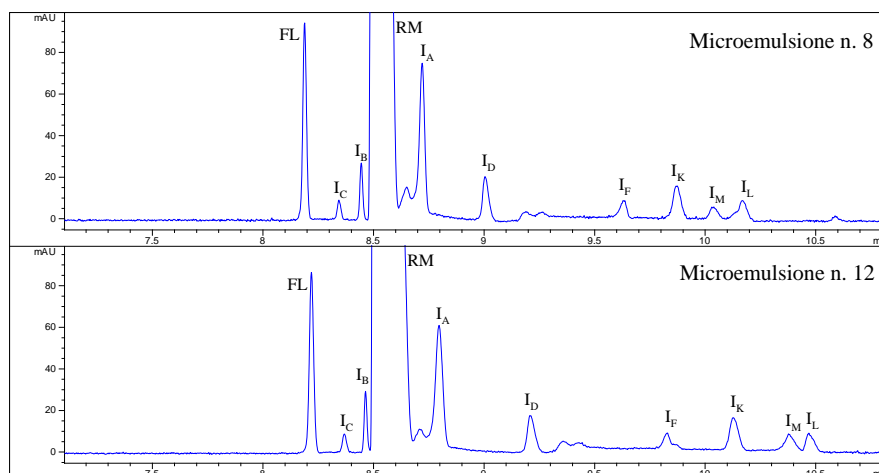
Per un obiettivo di ottimizzazione come questo, il modello ipotizzato dovrebbe essere predittivo, in modo da permettere la simulazione del comportamento delle proprietà studiate in tutto il dominio sperimentale. In questo caso è stato ipotizzato il seguente modello cubico ridotto [111].

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_1 x_2 x_3 + \varepsilon$$

dove y rappresenta la risposta, la variabile x_1 la fase acquosa, x_2 la fase oleosa, x_3 il componente tensioattivo/cotensioattivo. La matrice scelta per calcolare i 7 coefficienti del modello consisteva in un disegno di Scheffé simplex centroide a 13 esperimenti corrispondenti a 13 diverse miscele.

Purtroppo gli elettroferogrammi ottenuti non hanno permesso di trattare statisticamente le risposte considerate. Infatti, in diversi esperimenti si è verificata un'inversione dell'ordine di migrazione di coppie di analiti e/o una sovrapposizione dei relativi picchi. Di conseguenza non è stato possibile misurare in modo univoco la risoluzione tra i composti coinvolti, rendendo difficile tenere sotto controllo gli effetti delle diverse condizioni sperimentali. Comunque, dall'esame degli elettroferogrammi

ottenuti i migliori risultati sono stati ottenuti utilizzando come fasi pseudostazionarie la microemulsione n.8 e la microemulsione n.12.



Elettroferogrammi di una soluzione di RM 4 mg mL^{-1} e di impurezze 0.04 mg mL^{-1} . Lunghezza del capillare, 64.5 cm; lunghezza d'onda di rivelazione, 210 nm; differenza di potenziale, -25 kV; temperatura, $20 \text{ }^{\circ}\text{C}$.

In questi due casi è stata ottenuta una buona separazione degli analiti in circa 11 minuti, con lo stesso ordine di migrazione. Tra le due microemulsioni è stata poi scelta la n.8 per il proseguimento dello studio di ottimizzazione, in quanto dotata di maggiore stabilità, forse a causa della maggiore percentuale del componente *S/CoS* (10.00% rispetto a 9.00%). La composizione ottimale per il BGE era quindi costituita da: *W*, 88.95%; *O*, 1.05%; *S/CoS*, 10.00%.

Metodologia della superficie di risposta

Dopo aver stabilito la composizione della fase pseudostazionaria, lo studio è proseguito con la valutazione dei fattori indipendenti, in particolare degli effetti della concentrazione del tampone fosfato, della differenza di potenziale e della temperatura. Lo scopo era quello di ottenere la risoluzione alla linea di base di tutti gli analiti in un breve tempo di analisi.

È stata quindi applicata la metodologia della superficie di risposta, considerando come fattori concentrazione di fosfato (*Phosphate conc.*, mM), differenza di potenziale (*V*, kV) e temperatura (*T*, $^{\circ}\text{C}$), che sono stati studiati in domini sperimentali specifici.

Le risposte considerate sono state le risoluzioni critiche R_1 (I_B/RM) e R_2 (RM/I_A), da massimizzare, ed il tempo di analisi t , misurato come tempo di migrazione dell'ultimo picco I_L , da minimizzare.

Il disegno di Doehlert [12] ha permesso di ottenere le superfici di risposta stimando i coefficienti di un modello polinomiale di secondo ordine:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \varepsilon$$

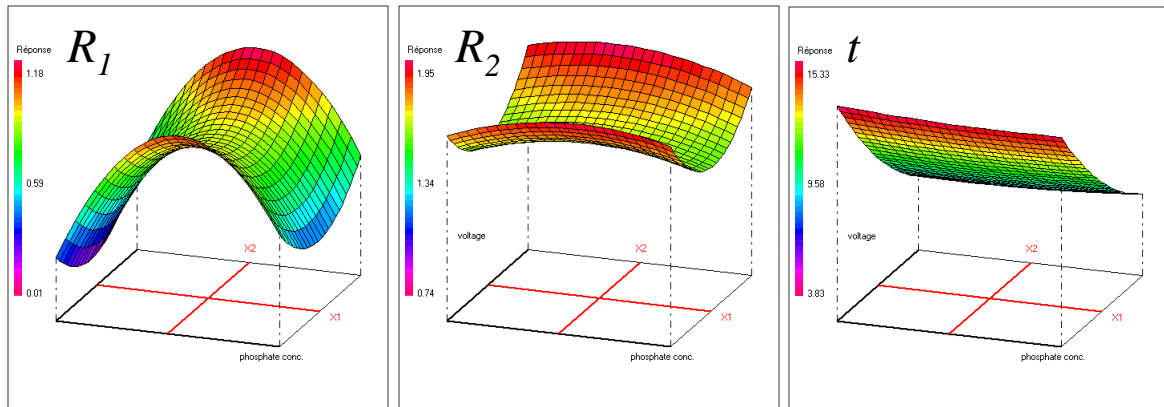
dove y rappresenta la risposta sperimentale, x_i i fattori indipendenti, β i coefficienti del modello, espressione degli effetti principali (lineari e quadratici) e di interazione, ed ε è l'errore sperimentale.

Il numero di esperimenti era pari a 15, inclusi 3 esperimenti al centro del dominio sperimentale. Le analisi sono state eseguite secondo quanto previsto dal piano sperimentale, riportato di seguito insieme alle risposte ottenute.

Exp. n.	Phosphate conc. (mM)	V (kV)	T (°C)	R_1	R_2	t (min)
1	100	25	18	0.49	1.58	11.13
2	70	25	18	0.01	1.48	9.90
3	93	30	18	0.96	1.76	7.66
4	78	20	18	0.83	1.86	14.70
5	93	20	18	0.84	1.81	14.00
6	78	30	18	0.99	1.89	8.03
7	93	27	20	0.86	1.80	8.69
8	78	23	16	0.92	1.80	12.69
9	93	23	16	0.93	1.77	12.08
10	85	28	16	0.92	1.76	9.45
11	78	27	20	0.91	1.81	9.17
12	85	22	20	0.96	1.67	12.33
13	85	25	18	0.87	1.61	10.73
14	85	25	18	0.92	1.64	10.08
15	85	25	18	0.83	1.66	10.27

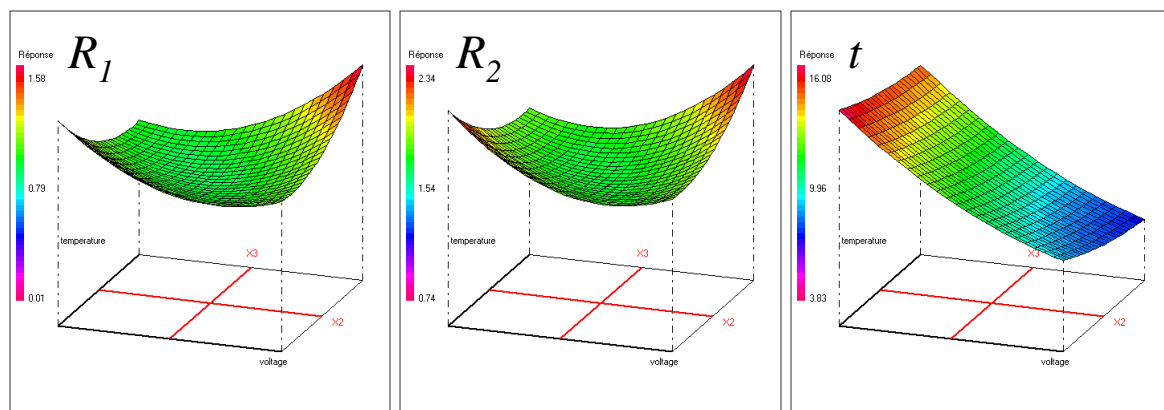
Le risposte ottenute sono state trattate statisticamente e l'analisi della varianza ha indicato che il modello di regressione calcolato era valido e significativo per ognuna delle risposte considerate.

I risultati ottenuti sono stati rappresentati graficamente tramite le superfici di risposta. Le seguenti superfici sono state ottenute riportando i valori di risoluzione R_1 e R_2 e i valori del tempo di analisi in funzione della concentrazione di fosfato e della differenza di potenziale.



Superfici di risposta per R_1 , R_2 , t in funzione di concentrazione di fosfato e differenza di potenziale. La temperatura è stata fissata al valore centrale del dominio sperimentale

Di seguito sono invece riportate le superfici di risposta ottenute in funzione della differenza di potenziale e della temperatura.

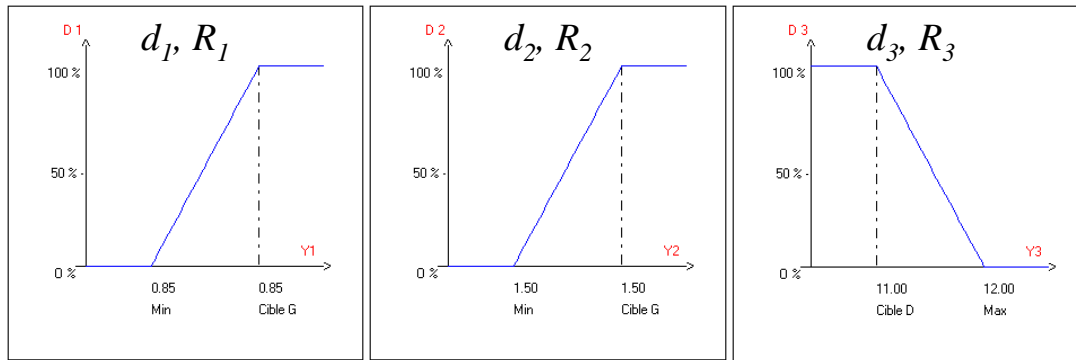


Superfici di risposta per R_1 , R_2 , t in funzione di differenza di potenziale e temperatura. La concentrazione di fosfato è stata fissata al valore centrale del dominio sperimentale

Per quanto riguarda la risposta R_1 è stato rilevato un evidente effetto quadratico del fattore concentrazione di fosfato. Nel caso della risposta R_2 si avevano effetti quadratici significativi sia della temperatura che della differenza di potenziale; tra questi due fattori è stata messa in evidenza anche una forte interazione positiva. Infine, i fattori influenti sul tempo di analisi sono risultati essere la temperatura e la differenza di potenziale, esercitando come previsto un effetto negativo su tale risposta.

Dall'analisi dei risultati è stata evidenziata la necessità di individuare i valori dei fattori che portassero al compromesso ottimale tra le diverse risposte, in modo da ottenere dal metodo le migliori prestazioni. Un problema di questo tipo può essere risolto attraverso la funzione di desiderabilità di Derringer (D) [12]. Tale funzione è una strategia decisionale multicriterio, misura della qualità totale del fenomeno, e costituisce un adatto strumento per confrontare risposte diverse e per selezionare l'optimum con proprietà più desiderabili. Questo tipo di approccio è fondato sulla definizione di una funzione di desiderabilità parziale d_i per ognuno dei criteri considerati [29]. Qualunque sia la funzione prescelta (lineare, a scalino, triangolare...) i valori della risposta -la desiderabilità- sono compresi tra zero (non accettabile) e uno (massima desiderabilità; in questo caso ulteriori miglioramenti non hanno importanza). In un secondo passaggio la qualità totale D è calcolata utilizzando la media geometrica delle singole funzioni d_i $D = (d_1 \cdot d_2 \cdot \dots \cdot d_m)^{1/m}$ [112, 30]. Inoltre, in dipendenza dell'importanza attribuita alla risposta, le funzioni individuali d_i possono essere pesate e di conseguenza la funzione di desiderabilità totale D assume la forma $D = \sqrt[m_1^{w_1} \cdot d_1^{w_1} \cdot d_2^{w_2} \cdot \dots \cdot d_m^{w_m}]{m_1^{w_1+w_2+\dots+w_m}}$. Viene poi applicato un algoritmo di calcolo alla funzione D in modo da determinare il set dei valori delle variabili che la massimizzano. Il valore di D è massimo dove la combinazione dei differenti criteri è globalmente ottimale [112, 30].

In questo caso è stato assegnato lo stesso peso a tutte le risposte considerate e le funzioni di desiderabilità parziale d_i e la funzione D sono state calcolate per mezzo del software NEMROD-W [113]. Definendo quindi le funzioni di desiderabilità parziali d_1 per R_1 , d_2 per R_2 , d_3 per t , la funzione D ha assunto la forma $D = \sqrt[3]{d_1 d_2 d_3}$. Le funzioni d_i avevano la forma di seguito raffigurata.



Il primo requisito da soddisfare era la separazione alla linea di base di tutti i picchi, e per R_1 e R_2 sono stati fissati dei valori minimi pari a 0.85 e 1.5. Il valore di risoluzione 1.5 è quello che generalmente definisce la separazione alla linea di base di due picchi di dimensioni simili. Nel caso di R_1 , tuttavia, un valore misurato pari a 0.85 era sufficiente per la separazione della relativa coppia di picchi a causa della differente ampiezza dei picchi stessi. Il secondo requisito è stato la velocità dell'analisi, quindi la funzione di desiderabilità parziale per la risposta tempo di analisi è stata definita accettando valori inferiori a 12 minuti e ritenendo ottimali valori inferiori a 11 minuti.

Dall'esame della rappresentazione bidimensionale della funzione di desiderabilità D è possibile osservare come la zona dove il valore di D era massimo corrispondeva a valori intermedi della concentrazione di fosfato e ad un intervallo medio-alto per la differenza di potenziale. Per la temperatura non sono stati invece evidenziati particolari vincoli, dato che tutti i valori compresi nel dominio, combinati con opportuni livelli di concentrazione di tampone e di differenza di potenziale, potevano corrispondere a $D = 1$. Per questo ultimo fattore si è quindi deciso di selezionare un valore basso, pari a 17 °C, in modo da incrementare l'efficienza dei picchi relativi alle impurezze senza che la strumentazione presentasse particolari problemi nel mantenimento della temperatura selezionata.

Attraverso la funzione di desiderabilità totale sono state quindi individuate le seguenti condizioni ottimizzate: concentrazione di fosfato, 90 mM; differenza di potenziale, -26 kV (polarità inversa); temperatura, 17 °C.

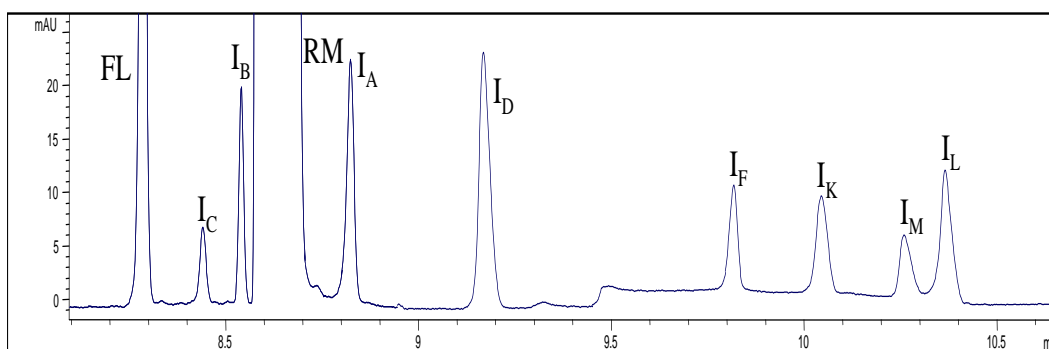
Applicando le condizioni ottimizzate si è ottenuta la completa separazione degli analiti in circa 10 minuti, con una buona efficienza dei picchi relativi alle impurezze.

Il modello è stato convalidato previsionalmente, ad un livello di probabilità del 95%, applicando la formula seguente:

$$\hat{y}_u \pm t_{\alpha/2, \nu} \left(\frac{1}{m} + d_u \right)^{1/2} s$$

dove \hat{y}_u è il valore predetto della risposta, m è il numero di repliche sperimentali, d_u è la funzione di varianza nel punto predetto e s è la deviazione standard calcolata dagli esperimenti al centro del dominio sperimentale [12]. I valori delle risposte misurate ricadevano all'interno dei rispettivi intervalli di confidenza, mostrando quindi un buon accordo tra predetto e misurato.

Nella figura seguente è riportato l'elettroferogramma ottenuto nelle condizioni ottimizzate. La corrente generata era circa $-80 \mu\text{A}$.



Elettroferogramma di una soluzione di RM 4 mg mL^{-1} e di impurezze 0.04 mg mL^{-1} nelle condizioni ottimizzate. BGE, microemulsione composta da 88.95% fase acquosa (fosfato 90 mM pH 2.5), 1.05% fase oleosa (*n*-eptano), 10.00% tensioattivo/cotensioattivo (SDS/*n*-butanolo) in rapporto 1:2. Lunghezza capillare, 64.5 cm; lunghezza d'onda di rivelazione, 210 nm; differenza di potenziale, -26 kV ; temperatura, $17 \text{ }^\circ\text{C}$.

Convalida

Il metodo MEEKC per la determinazione di RM e delle sue impurezze è stato convalidato nelle condizioni sperimentali ottimizzate (BGE: fosfato 90 mM pH 2.5 88.95%, *n*-eptano 1.05%, SDS/*n*-butanolo in rapporto 1:2 10.00%; differenza di potenziale: -26 kV ; temperatura: $17 \text{ }^\circ\text{C}$), seguendo le linee guida ICH [40] ed usando standard puri di RM e di impurezze. Salvo espressamente indicato, il campione utilizzato per eseguire gli esperimenti è stato preparato aggiungendo volumi appropriati

delle soluzioni standard in modo da ottenere una concentrazione pari a 4 mg mL^{-1} di RM e una concentrazione pari a 0.02 mg mL^{-1} delle impurezze (0.5% w/w, intorno al centro del loro range di linearità). Ai fini quantitativi è stato utilizzato il metodo dello standard interno, rappresentato dall'acido flufenamico a concentrazione pari a 0.1 mg mL^{-1} .

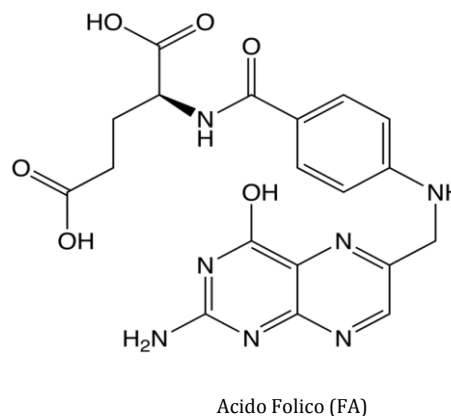
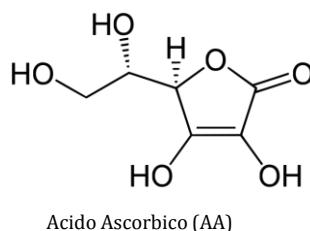
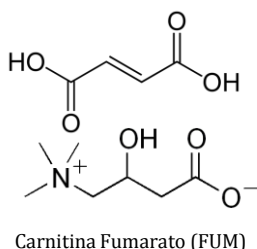
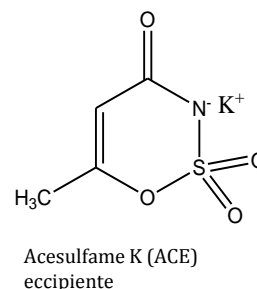
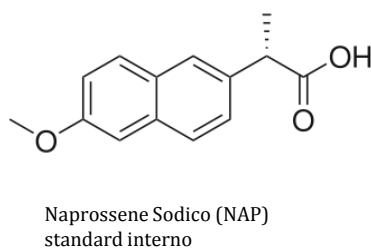
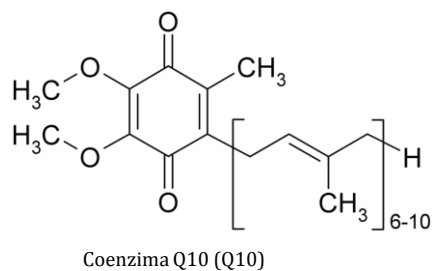
Applicazioni

Una volta convalidato, il metodo MEEKC è stato applicato per effettuare l'analisi di un campione reale costituito da compresse di Triatec[®], contenenti 10 mg di RM. Sono state effettuate quattro analisi ($n=4$, $\alpha/2=0.025$), ottenendo risultati in accordo con la composizione dichiarata: *recovery*, $98.7 \pm 1.5\%$; RSD, 1.0%). Non è stata rilevata alcuna impurezza di RM.

7.3. Sviluppo di un metodo MEEKC per l'analisi di un integratore a base di CoQ10 mediante l'utilizzo di una matrice mista di variabili miscela-processo (MPV)

L'obiettivo di questo lavoro è stato lo sviluppo di un metodo CE rapido e selettivo utile per l'analisi dell'integratore alimentare Proxeed NF[®], prodotto dalla Sigma-Tau. Tale integratore, distribuito sotto forma di granulato contenuto in bustine, è basato sul coenzima Q10 (Q10), associato ad altri composti attivi.

I componenti dell'integratore che mostravano un picco rilevabile alla concentrazione test e alla lunghezza d'onda di rilevazione (215 nm) includevano, oltre al Q10, l'acido ascorbico (AA), l'acido folico (FA), il fumarato (FUM) e l'eccipiente acesulfame K (ACE). Considerando anche lo standard interno naprossene sodico (NAP), la separazione coinvolgeva un totale di 6 composti, le cui formule di struttura sono di seguito riportate.



Tra i composti considerati, il Q10 è un chinone liposolubile con caratteristiche neutre, mentre AA e FA hanno natura acida. Quindi, per effettuare l'analisi simultanea di questi composti occorre scegliere una adatta fase pseudostazionaria (PSP), sia per le diverse proprietà acido-base degli analiti stessi, sia per il fatto che gli analiti neutri sono privi di una propria mobilità elettroforetica e quindi non possono essere separati con una semplice CZE.

Scelta della modalità operativa

Sono stati condotti esperimenti preliminari per la scelta della modalità operativa più idonea all'analisi, sia in termini di polarità applicata che di tipo di PSP (micelle o gocce oleose di una microemulsione).

Polarità normale

Nella prima serie di esperimenti la differenza di potenziale è stata applicata in polarità normale.

Nella prima fase dello studio è stata presa in considerazione una PSP a base di micelle formate da SDS, usando come tampone una soluzione di borace 10 mM. In queste condizioni, le principali problematiche analitiche riguardavano particolari aspetti

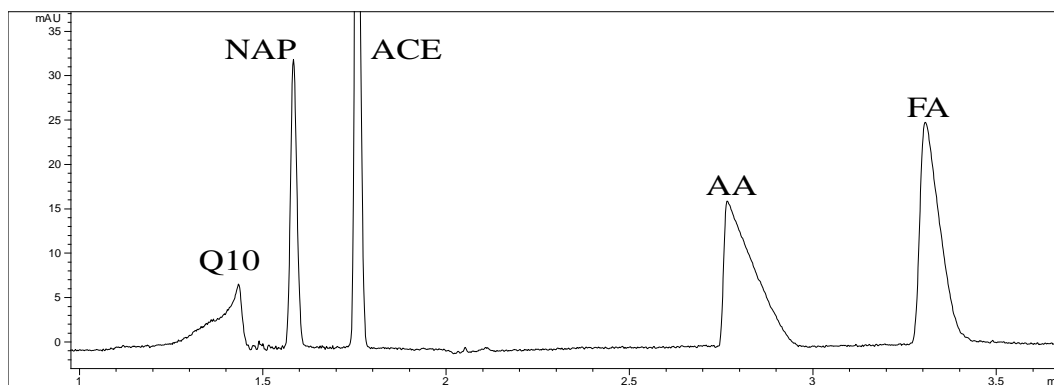
di selettività ed efficienza. Per quanto riguarda la selettività, all'inizio dell'elettroferogramma era presente una zona estremamente critica per la separazione di AA e FA, che avevano tempi di migrazione molto simili. Il principale problema era comunque costituito dall'efficienza del picco relativo al Q10, così bassa da rendere impossibile la sua rilevazione anche a diverse concentrazioni di tensioattivo.

È stata poi valutata anche la possibilità di utilizzare come BGE una microemulsione. Come prova preliminare è stata utilizzata una microemulsione standard già impiegata per la determinazione del Q10 nel plasma tramite MEEKC [114]. Tale microemulsione era costituita dal 91.14% di fase acquosa (tampone fosfato 50 mM pH 10.00), 0.81% di fase oleosa (*n*-ottano), 8.05% di tensioattivo/cotensioattivo (SDS/*n*-butanolo) in rapporto 1:4.5. In tali condizioni si evidenziava un tempo di analisi piuttosto lungo (circa 15 minuti), con una efficienza molto bassa del picco relativo al Q10.

Polarità inversa

Nella successiva serie di esperimenti la differenza di potenziale è stata applicata in polarità inversa. Anche con questo tipo di polarità la MEKC non ha portato a buoni risultati a causa della scarsissima efficienza del picco relativo al Q10, per cui questo sistema è stato scartato senza approfondire la sua modulazione.

Nel caso della MEEKC, già dai primi elettroferogrammi sono stati invece ottenuti risultati promettenti. Come prova preliminare è stata presa in considerazione la microemulsione usata per valutare la MEEKC in polarità normale, con la sola differenza costituita dal tipo di fase acquosa, cioè un tampone acetato 100 mM pH 5.00 al posto del tampone fosfato 50 mM pH 10.00. Di seguito viene riportato l'elettroferogramma ottenuto in queste condizioni e lavorando in *short end injection*, ovvero iniettando il campione dalla parte del rivelatore, in modo da ottenere bassi tempi di analisi.



Elettroferogramma di una soluzione degli analiti alle concentrazioni test. BGE, microemulsione composta da 91.14% fase acquosa (acetato 100 mM pH 5.00), 0.81% fase oleosa (*n*-ottano), 8.05% tensioattivo/cotensioattivo (SDS/*n*-butanolo) in rapporto 1:4.5. Lunghezza totale del capillare, 48.5 cm; lunghezza effettiva, 8.5 cm (*short end injection*). Lunghezza d'onda di rivelazione, 215 nm; differenza di potenziale, -17kV; temperatura, 20 °C.

In queste condizioni è stata ottenuta una risoluzione alla linea di base degli analiti, pur con scarsa efficienza dei picchi. Questo sistema è stato giudicato promettente ed è stato quindi scelto come punto di partenza per effettuare l'ottimizzazione multivariata.

Ottimizzazione multivariata della modalità operativa MEEKC

L'obiettivo della fase di ottimizzazione era di individuare la proporzione dei componenti della microemulsione (MC) e i valori delle variabili di processo (PV) che permettessero di ottenere una buona efficienza dei picchi in modo da consentire una accurata analisi quantitativa. Come risposte sono stati quindi scelti i valori di efficienza dei picchi relativi a Q10 (Eff.Q10), acido ascorbico (Eff.AA) e acido folico (Eff.FA).

L'ottimizzazione multivariata del sistema MEEKC si è articolata in diverse fasi:

- 1) scelta del modello e del disegno sperimentale MPV per massimizzare l'efficienza dei picchi;
- 2) interpretazione degli effetti degli MC e delle PV sulle risposte attraverso l'analisi delle superfici di isorisposta *contour plots* e dei *perturbation plots*;
- 3) individuazione dello spazio del disegno e al suo interno di una regione ottimale entro la quale selezionare le condizioni ottimizzate per MC e PV.

Modello e disegno sperimentale MPV

Il dominio sperimentale è stato scelto sulla base di prove preliminari, imponendo limiti inferiori e superiori ai componenti di miscela e alle variabili di processo secondo quanto mostrato nella tabella seguente, dove sono riportati i valori originali e codificati per i diversi fattori in studio.

Componente di miscela	Variabili originali				Variabili codificate		
	Unità	Notazione	Limite inferiore	Limite superiore	Notazione	Limite inferiore	Limite superiore
Fase acquosa	Proporzione	B	0.900	0.948	x_1	0	1
Tensioattivo/cotensioattivo	Proporzione	S	0.050	0.098	x_2	0	1
Olio	Proporzione	O	0.002	0.020	x_3	0	0.375
Variabile di processo							
Voltaggio applicato	kV	V	22	27	z_1	-1	1
Concentrazione del tampone	mM	BC	95	105	z_2	-1	1
pH del tampone	pH	pH	4.5	5.5	z_3	-1	1

È stato quindi postulato un modello MPV per poi generare un disegno sperimentale MPV mediante l'utilizzo del criterio di qualità *I*-optimal. Per i componenti di miscela sono stati considerati i termini relativi ad un modello cubico speciale, mentre per le variabili di processo sono stati studiati gli effetti principali, di interazione e quadratici [7, 10, 11, 103].

Incrociando questi due modelli si otteneva il modello misto MPV a 70 termini di seguito riportato.

$$\begin{aligned}
 y_{SC-Q} = & g_1^0 x_1 + g_2^0 x_2 + g_3^0 x_3 + g_{12}^0 x_1 x_2 + g_{13}^0 x_1 x_3 + g_{23}^0 x_2 x_3 + g_{123}^0 x_1 x_2 x_3 \\
 & + (g_1^1 x_1 + g_2^1 x_2 + g_3^1 x_3 + g_{12}^1 x_1 x_2 + g_{13}^1 x_1 x_3 + g_{23}^1 x_2 x_3 + g_{123}^1 x_1 x_2 x_3) z_1 \\
 & + (g_1^2 x_1 + g_2^2 x_2 + g_3^2 x_3 + g_{12}^2 x_1 x_2 + g_{13}^2 x_1 x_3 + g_{23}^2 x_2 x_3 + g_{123}^2 x_1 x_2 x_3) z_2 \\
 & + (g_1^3 x_1 + g_2^3 x_2 + g_3^3 x_3 + g_{12}^3 x_1 x_2 + g_{13}^3 x_1 x_3 + g_{23}^3 x_2 x_3 + g_{123}^3 x_1 x_2 x_3) z_3 \\
 & + (g_1^{12} x_1 + g_2^{12} x_2 + g_3^{12} x_3 + g_{12}^{12} x_1 x_2 + g_{13}^{12} x_1 x_3 + g_{23}^{12} x_2 x_3 + g_{123}^{12} x_1 x_2 x_3) z_1 z_2 \\
 & + (g_1^{13} x_1 + g_2^{13} x_2 + g_3^{13} x_3 + g_{12}^{13} x_1 x_2 + g_{13}^{13} x_1 x_3 + g_{23}^{13} x_2 x_3 + g_{123}^{13} x_1 x_2 x_3) z_1 z_3 \\
 & + (g_1^{23} x_1 + g_2^{23} x_2 + g_3^{23} x_3 + g_{12}^{23} x_1 x_2 + g_{13}^{23} x_1 x_3 + g_{23}^{23} x_2 x_3 + g_{123}^{23} x_1 x_2 x_3) z_2 z_3 \\
 & + (g_1^{11} x_1 + g_2^{11} x_2 + g_3^{11} x_3 + g_{12}^{11} x_1 x_2 + g_{13}^{11} x_1 x_3 + g_{23}^{11} x_2 x_3 + g_{123}^{11} x_1 x_2 x_3) z_1^2 \\
 & + (g_1^{22} x_1 + g_2^{22} x_2 + g_3^{22} x_3 + g_{12}^{22} x_1 x_2 + g_{13}^{22} x_1 x_3 + g_{23}^{22} x_2 x_3 + g_{123}^{22} x_1 x_2 x_3) z_2^2 \\
 & + (g_1^{33} x_1 + g_2^{33} x_2 + g_3^{33} x_3 + g_{12}^{33} x_1 x_2 + g_{13}^{33} x_1 x_3 + g_{23}^{33} x_2 x_3 + g_{123}^{33} x_1 x_2 x_3) z_3^2 + \varepsilon
 \end{aligned}$$

La stima di questo modello richiedeva un numero di esperimenti troppo elevato ai fini del nostro studio. Inoltre, tale modello includeva interazioni di ordine superiore tra

MC e PV che sulla base di esperienze pregresse sarebbero presumibilmente risultate non significative.

È stato quindi selezionato un modello meno complesso, comprendente 46 termini, come base per generare il disegno sperimentale MPV:

$$\begin{aligned}
y = f(x, z) = & g_1^0 x_1 + g_2^0 x_2 + g_3^0 x_3 + g_{12}^0 x_1 x_2 + g_{13}^0 x_1 x_3 + g_{23}^0 x_2 x_3 + g_{123}^0 x_1 x_2 x_3 \\
& + (g_1^1 x_1 + g_2^1 x_2 + g_3^1 x_3 + g_{12}^1 x_1 x_2 + g_{13}^1 x_1 x_3 + g_{23}^1 x_2 x_3 + g_{123}^1 x_1 x_2 x_3) z_1 \\
& + (g_1^2 x_1 + g_2^2 x_2 + g_3^2 x_3 + g_{12}^2 x_1 x_2 + g_{13}^2 x_1 x_3 + g_{23}^2 x_2 x_3 + g_{123}^2 x_1 x_2 x_3) z_2 \\
& + (g_1^3 x_1 + g_2^3 x_2 + g_3^3 x_3 + g_{12}^3 x_1 x_2 + g_{13}^3 x_1 x_3 + g_{23}^3 x_2 x_3 + g_{123}^3 x_1 x_2 x_3) z_3 \\
& + (g_1^{12} x_1 + g_2^{12} x_2 + g_3^{12} x_3) z_1 z_2 + (g_1^{13} x_1 + g_2^{13} x_2 + g_3^{13} x_3) z_1 z_3 \\
& + (g_1^{23} x_1 + g_2^{23} x_2 + g_3^{23} x_3) z_2 z_3 + (g_1^{11} x_1 + g_2^{11} x_2 + g_3^{11} x_3) z_1^2 \\
& + (g_1^{22} x_1 + g_2^{22} x_2 + g_3^{22} x_3) z_2^2 + (g_1^{33} x_1 + g_2^{33} x_2 + g_3^{33} x_3) z_3^2 + \varepsilon
\end{aligned}$$

dove y rappresenta la risposta considerata (efficienza del picco relativo a Q10, AA e FA); x_i indica la trasformazione in L-pseudocomponenti dell' i -esimo componente di miscela; z_j indica la codificazione [-1, 1] della j -esima variabile di processo. I coefficienti del modello g vengono calcolati sulla base dei dati sperimentali. Il pedice del coefficiente g indica gli MC coinvolti in tale termine del modello, mentre l'apice indica le PV coinvolte.

I termini del modello riportati nella prima riga, aventi zero come apice, rappresentano gli effetti lineari, quadratici e cubici relativi agli MC quando tutte le PV si trovano al loro valore codificato pari a zero, corrispondente al centro del dominio sperimentale. I termini nelle restanti linee rappresentano gli effetti delle PV sulle proprietà degli MC. Così, un modello a 46 termini permette di conoscere 1) gli effetti lineari di tutte e tre le PV sulle proprietà lineari, quadratiche e cubiche speciali degli MC e 2) le proprietà lineari degli MC influenzati dagli effetti di interazione e quadratici delle tre PV.

Il modello a 46 termini è stato selezionato allo scopo di generare un disegno sperimentale MPV capace di approssimare le relazioni esistenti tra le variabili risposta (Eff.Q10, Eff.AA, Eff.FA), gli MC e le PV.

Per stimare i coefficienti del modello MPV, quantificare l'errore sperimentale e valutare la validità del modello per ogni risposta, è stato pianificato un disegno

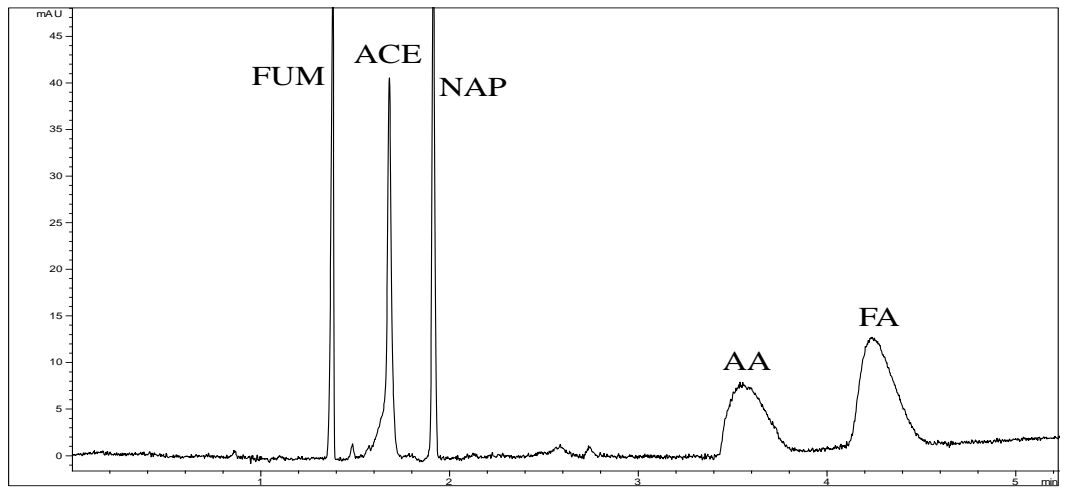
sperimentale a 62 esperimenti. Tale numero deriva dalla somma dei 46 termini del modello MPV, di 10 esperimenti per valutare la validità del modello, e di 6 esperimenti replicati.

Il piano sperimentale MPV a 62 esperimenti è stato generato mediante il software Design-Expert [Stat Ease] usando un algoritmo basato sul criterio di qualità *I-optimal*. Tale algoritmo è in grado di selezionare i cosiddetti disegni ottimali sia sulla base di calcoli puntuali basati sui punti del disegno e quindi a partire da punti candidati, sia sulla base di calcoli basati sulle coordinate che non richiedono di specificare punti candidati [20]. In questo modo è possibile ottenere la stima dei coefficienti del modello con il minor numero di esperimenti e la miglior qualità dell'informazione. In questo studio il miglior disegno *I-ottimale* è risultato essere un disegno generato attraverso l'algoritmo basato sulle coordinate.

Per quanto riguarda la misura delle risposte, sono stati riscontrati problemi significativi nella misura dell'efficienza del picco relativo al Q10 e in grado minore anche per i picchi relativi ad AA e FA.

In particolare, si è verificata una grande variabilità nel valore misurato di Eff.Q10 nelle diverse condizioni sperimentali, che portava talvolta ad un tale allargamento del picco da far sì che questo non fosse rilevabile. In tali casi, ad Eff.Q10 è stato assegnato un valore pari a 10. Inoltre, a causa dell'incertezza della misura è risultato difficile misurare valori di Eff.Q10 minori di 2000, per cui in tali casi è stato assegnato uno stesso valore pari a 1000.

La figura sotto riportata mostra l'elettroferogramma ottenuto relativamente al punto (#15) del disegno sperimentale MPV. Si può notare una scarsa efficienza del metodo unita tuttavia ad una buona selettività e ad un breve tempo di analisi.



Elettroferogramma relativo al punto (#15) del disegno sperimentale MPV. Il picco relativo a Q10 non è rilevabile a causa della sua scarsa efficienza

Sulla base degli esperimenti replicati, sono state ottenute le stime dell'errore sperimentale e quindi della deviazione standard delle risposte Eff.Q10, Eff.AA e Eff.FA, pari rispettivamente a 1885, 373 e 3025. Occorre a tale proposito notare che tutte le coppie di repliche di Eff.Q10 tranne una avevano un valore assegnato identico pari a 10 o a 1000. Di conseguenza, la stima dell'errore sperimentale così calcolato, pari a 1885, potrebbe essere non rappresentativa dell'incertezza reale per valori superiori di Eff.Q10.

Per sviluppare i modelli MPV per le tre risposte considerate, i componenti di miscela sono stati codificati mediante le trasformazioni in L-pseudocomponenti [21]:

$$x_1 = (B - 0.9)/(1 - 0.9 - 0.05 - 0.002) = (B - 0.9)/(0.048)$$

$$x_2 = (S - 0.05)/0.048$$

$$x_3 = (O - 0.002)/0.048$$

Le variabili di processo sono state codificate negli intervalli $-1 \leq z_i \leq 1$, usando le seguenti trasformazioni:

$$z_1 = (V - 24.5)/2.5$$

$$z_2 = (BC - 100)/5$$

$$z_3 = (\text{pH} - 5)/0.5$$

La codificazione degli MC in L-pseudocomponenti riduce la collinearità che deriva dai piccoli intervalli dei componenti [21]. La codificazione $-1 \leq z_i \leq 1$ delle PV fornisce una più semplice interpretazione del modello MPV. La porzione di un modello MPV che contiene solo i termini relativi agli MC rappresenta l'effetto degli MC quando i valori codificati delle PV sono pari a zero, cioè i valori delle PV sono fissati al centro

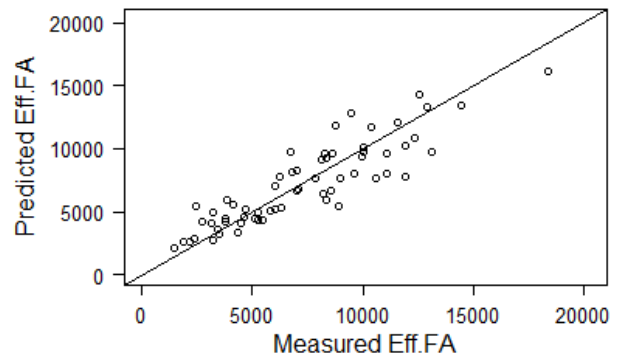
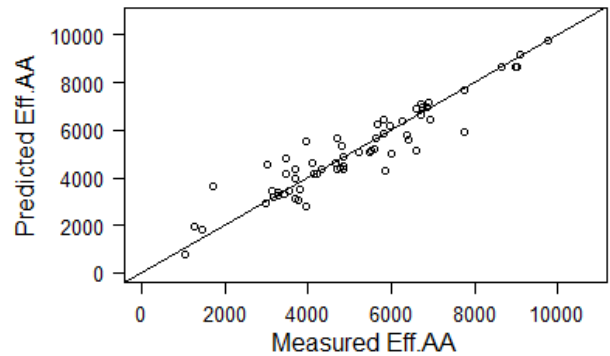
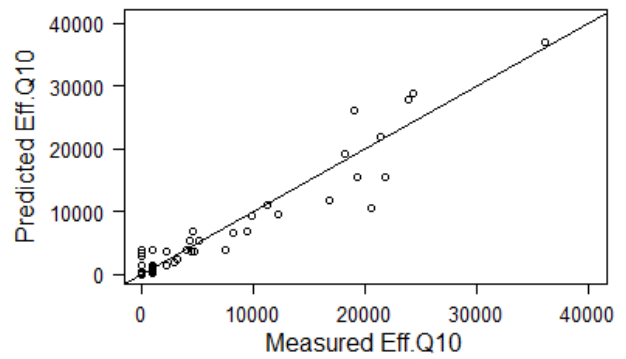
del dominio sperimentale. I termini che contengono sia MC che PV rappresentano gli effetti del cambiamento delle PV dai loro valori centrali sulle proprietà di miscela degli MC.

Per ogni risposta, il modello a 46 termini con le variabili codificate è stato calcolato sulla base dei dati ottenuti effettuando gli esperimenti previsti dal disegno sperimentale MPV a 62 punti usando il metodo dei minimi quadrati ordinari (*Ordinary Least Squares*, OLS), modulo presente in Design-Expert . Tuttavia, poiché molti termini del modello sono risultati essere statisticamente non significativi, è stato utilizzato un metodo statistico di selezione delle variabili (*backward elimination*) per sviluppare forme ridotte del modello originario a 46 termini. Usando questo approccio sono stati ottenuti modelli accettabili per le risposte Eff.AA e $(\text{Eff.FA})^{0.5}$. La trasformazione radice quadrata di Eff.FA è stata selezionata in base all'analisi dei dati eseguita con trasformazione Box-Cox , resa possibile da un modulo presente in Design-Expert. Questo approccio nell'ottenimento di modelli MPV non ha portato a risultati accettabili per Eff.Q10. La ragione più probabile di questo è la difficoltà nel determinare i valori di efficienza per Q10 nel caso di picchi non rilevabili o di scarsa efficienza, per cui è stata necessaria l'assegnazione di valori arbitrari di efficienza rispettivamente pari a 10 o a 1000 (in 36 casi su 62 punti). Tale raggruppamento e assegnazione di valori per due sottoinsiemi di dati per Eff.Q10 può causare problemi nell'approccio di regressione OLS. Quindi per il modello relativo all'Eff.Q10 è stato utilizzato un diverso approccio, detto modello generalizzato lineare (*Generalized Linear Model*, GLIM) . Si tratta di una generalizzazione flessibile di OLS che permette di trattare variabili risposta che non seguono una distribuzione normale. Come parte del processo di selezione del modello sono stati esaminati diversi *link* e diverse distribuzioni di errore, individuando nel *link* logaritmico e nella distribuzione dell'errore di Poisson il modello che forniva i migliori risultati.

I modelli ridotti MPV ottenuti attraverso *backward elimination* a partire dal modello MPV a 46 termini sono stati studiati approfonditamente.

Nella figura seguente sono mostrati i grafici che riportano i valori predetti utilizzando i modelli della tabella precedente rispetto ai valori misurati per i punti del disegno sperimentale.

I valori di R^2 sono pari a 0.913, 0.879 e 0.804 rispettivamente per $\ln(\text{Eff.Q10})$, Eff.AA e $(\text{Eff.FA})^{0.5}$, quindi non così alti quanto desiderato (>0.95). Tuttavia, ciò era previsto data la difficoltà riscontrata nella misura dell'efficienza del picco. I valori di R^2 sono tuttavia abbastanza alti da spiegare una parte sostanziale della variazione nei valori di efficienza dei picchi attraverso i modelli MPV. Quindi, i modelli possono essere usati (tenendo comunque presenti le incertezze) per interpretare gli effetti degli MC e delle PV sulle risposte, sulla base dei termini statisticamente significativi dei modelli stessi. Un termine è considerato statisticamente significativo se la confidenza che il suo coefficiente è diverso da zero è pari ad almeno il 90%.



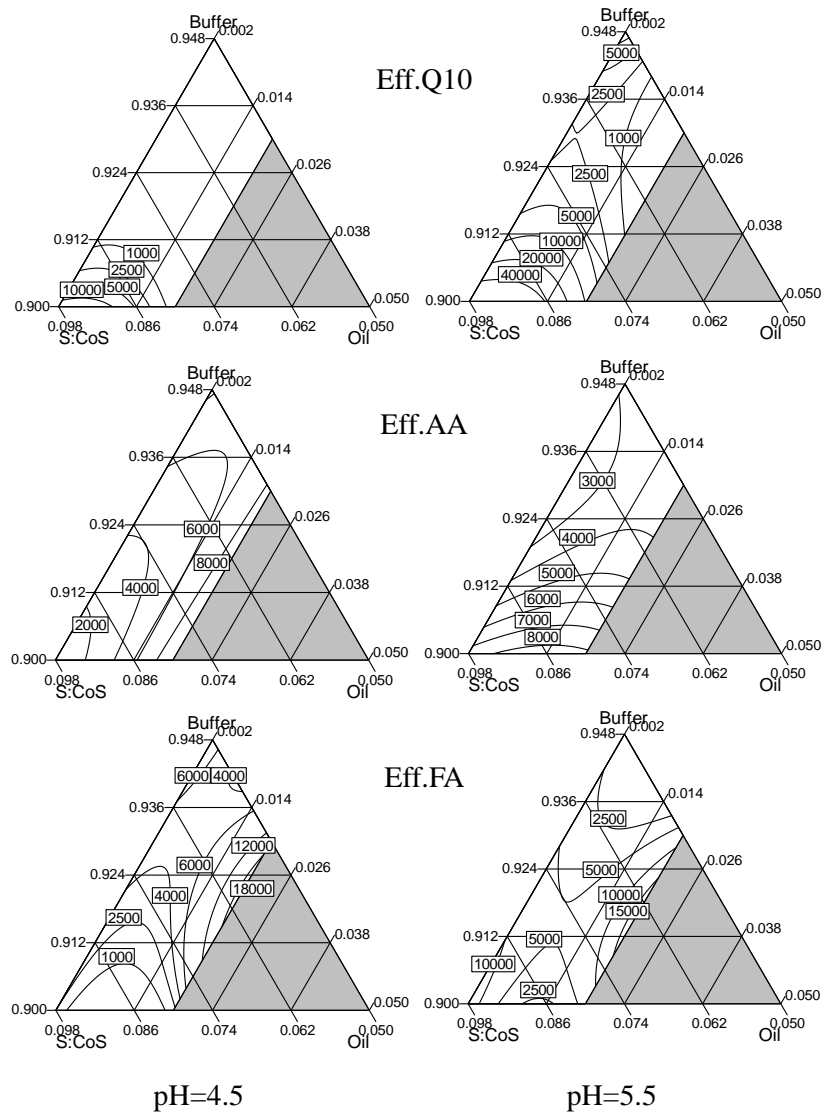
Contour plots e perturbation plots

Per ognuna delle otto combinazioni dei livelli inferiori e superiori delle PV è stata generata una superficie di isorisposta (*contour plot*) per Eff.Q10 , Eff.AA e Eff.FA nel dominio sperimentale delimitato dai fattori di miscela. I *contour plots* mostrano come il cambiamento delle proporzioni degli MC (cioè la composizione della microemulsione) influenza la risposta. Per ognuna delle tre risposte considerate, i *contour plots* differiscono in modo significativo nelle otto combinazioni delle PV. Questo indica che

le PV hanno effetti significativi sulle proprietà di miscela degli MC; in altre parole, gli MC e le PV presentano effetti di interazione tra loro.

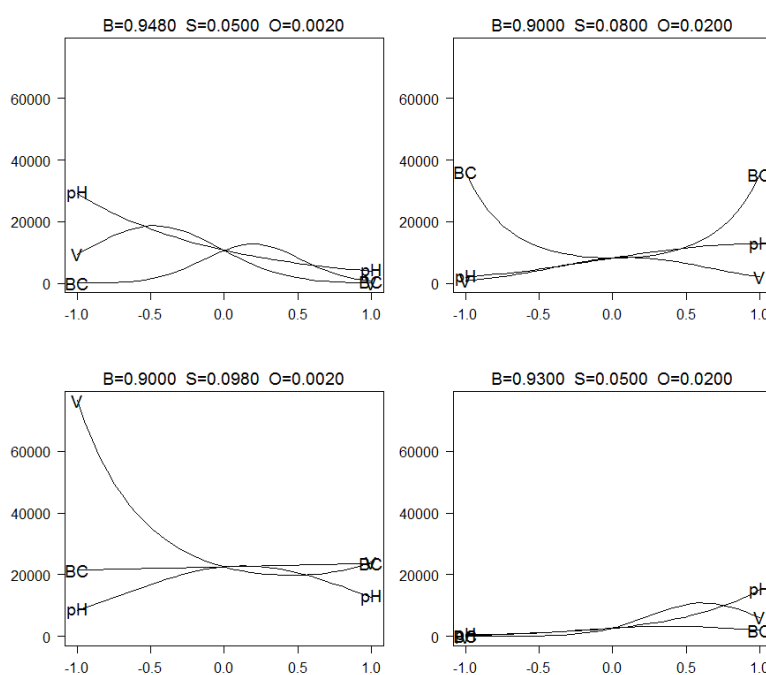
La figura seguente mostra a titolo di esempio due degli otto contour plots ottenuti per due delle combinazioni delle variabili PV (pH=4.5 e pH=5.5, con BC=105 mM e V=22 kV) per ciascuna delle risposte Eff.Q10, Eff.AA, Eff.FA, in modo da focalizzare le differenze nell'andamento e nei valori massimi delle risposte predette nel caso di diverso valore delle PV e nel caso dei diversi analiti.

Nelle figure successive sono mostrati alcuni *contour plots* ottenuti per ciascuna delle risposte.



Contour plots di Eff.Q10, Eff. AA e Eff.FA relative ai modelli MPV nel dominio sperimentale dei componenti di miscela per due combinazioni delle variabili PV (pH 4.5 e 5.5, con BC=105 mM e V=22 kV).

Per ognuna delle quattro combinazioni corrispondenti ai quattro vertici del dominio sperimentale degli MC è stato generato un *perturbation plot* per Eff.Q10, Eff.AA e Eff.FA. I *perturbation plots* mostrano tre curve che fanno vedere l'effetto della variazione di ognuna delle tre PV dal suo livello inferiore al suo livello superiore, con le altre PV fissate ai loro valori centrali. La figura seguente riporta i *perturbation plots* relativi a Eff.Q10. È possibile notare come in generale le PV hanno effetti relativamente bassi su Eff.Q10, con alcune eccezioni. In particolare, quando B(=0.948) è al suo livello più alto, mentre S (=0.050) e O (=0.002) sono al loro livello più basso, Eff.Q10 diminuisce all'aumentare del pH. Quando B(=0.900) si trova al livello inferiore, mentre S (=0.080) assume un valore relativamente alto e O (=0.020) assume il suo valore più alto, Eff.Q10 aumenta quando BC va dal livello centrale ai suoi valori inferiore e superiore. Quando B(=0.900) e O (=0.002) sono ai loro valori inferiori, mentre S(=0.098) corrisponde al suo valore più alto, Eff.Q10 aumenta col diminuire di V. Alla formulazione della microemulsione corrispondente a B=0.930, S=0.050 e O=0.020 nessuna delle PV ha un effetto importante su Eff.Q10. I diversi effetti delle PV nel caso delle diverse formulazioni derivano dagli effetti di interazione degli MC e delle PV su Eff.Q10.



Perturbation plots relativi al modello MPV per Eff.Q10 per ciascuno dei quattro vertici del dominio sperimentale degli MC. Viene predetto l'effetto della variazione di ogni PV dal suo valore inferiore (codificato con -1.0) al suo livello superiore (codificato con +1.0), con le altre PV fissate al loro livello centrale (codificato con 0.0).

Spazio del disegno, regione ottimale e condizioni ottimizzate

L'obiettivo dell'ottimizzazione è stato di individuare lo spazio del disegno e al suo interno una regione ottimale entro la quale selezionare la formulazione della microemulsione e i valori delle PV che portassero ai valori desiderati di efficienza. In base all'esame visuale degli elettroferogrammi ottenuti nel disegno MPV, sono stati specificati i valori minimi accettabili e i valori minimi ottimali per ogni risposta:

Eff.Q10: Valore minimo accettabile=11000, Valore minimo ottimale=20000

Eff.AA: Valore minimo accettabile=3500, Valore minimo ottimale=6000

Eff.FA: Valore minimo accettabile=5000, Valore minimo ottimale =8500

Lo spazio del disegno è stato individuato usando i valori minimi accettabili nelle equazioni

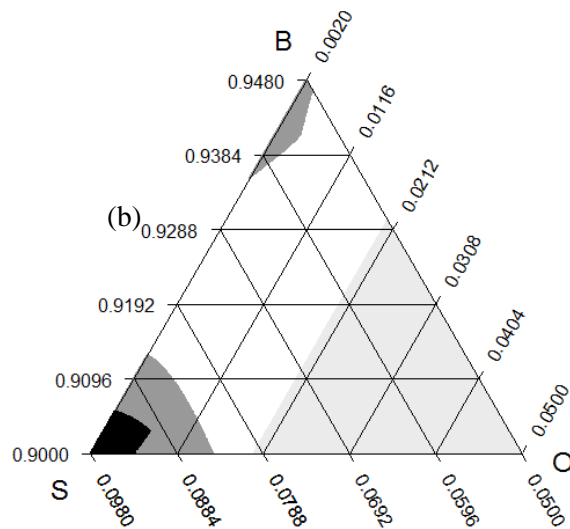
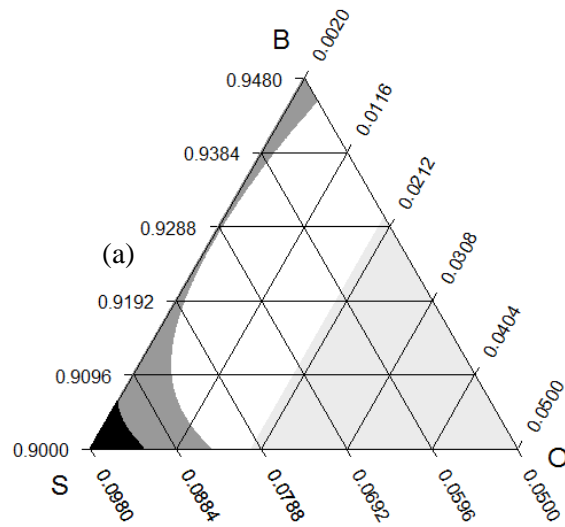
$$M(\text{Eff.Q10}) \geq 11000, M(\text{Eff.AA}) \geq 3500 \text{ e } M(\text{Eff.FA}) \geq 5000$$

mentre la regione ottimale all'interno del disegno è stata individuata attraverso le equazioni

$$M(\text{Eff.Q10}) \geq 20000, M(\text{Eff.AA}) \geq 6000 \text{ e } M(\text{Eff.FA}) \geq 8500$$

dove le notazioni $M(\text{Eff.Q10})$, $M(\text{Eff.AA})$ e $M(\text{Eff.FA})$ rappresentano i modelli MPV già descritti.

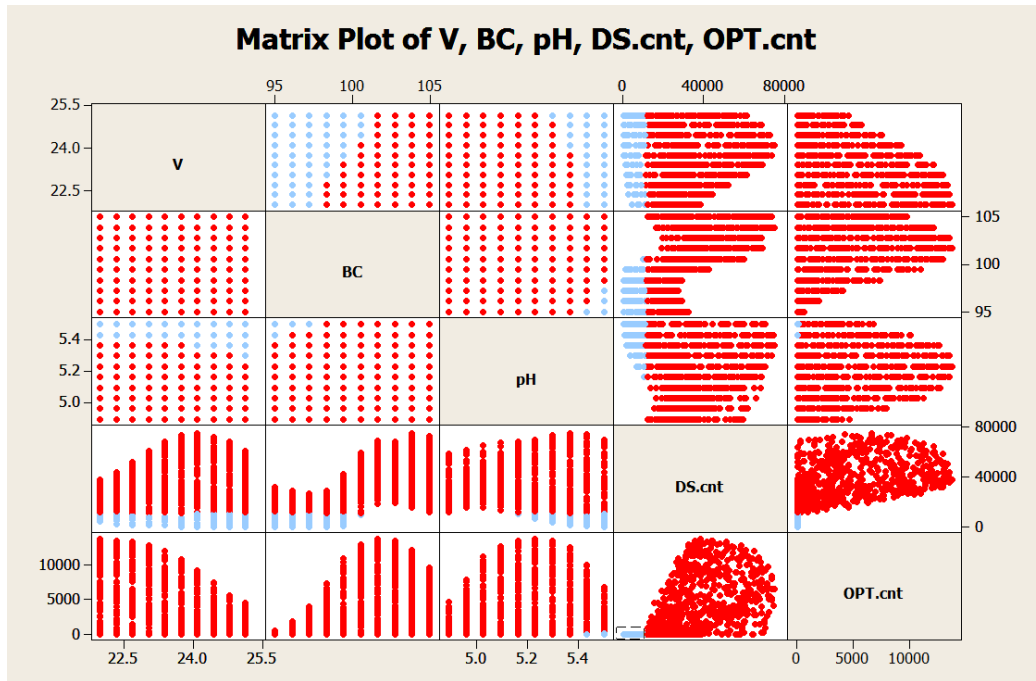
Lo spazio del disegno e la regione ottimale sono regioni 5-dimensionali specificate dai valori codificati degli MC e delle PV $((x_1, x_2, x_3, z_1, z_2, z_3)$, dove $x_1+x_2+x_3=1$) usati nei modelli. Proprio perché queste regioni sono a 5 dimensioni, non possono essere illustrate nella loro interezza. La figura sotto riportata illustra lo spazio del disegno e la regione ottimale quanto a formulazioni della microemulsione per due combinazioni selezionate di valori delle PV. Il grafico indicato con (a) è relativo alla combinazione dei valori delle PV pari a $(V, BC, pH) = (23.8, 100.9, 5.04)$, corrispondenti al punto del disegno MPV #57, l'unico che rientra in tutti i limiti della regione ottimale. Il grafico indicato con (b) è relativo alla combinazione dei valori PV (23, 101, 5.3) discussi di seguito.



Spazio del disegno (grigio) e regione ottimale (grigio scuro) quanto a formulazioni della microemulsione per due combinazioni selezionate delle PV (V, BC, pH): (a) Combinazione (23.8, 100.9, 5.04) corrispondente al punto del disegno #57, (b) combinazione (23, 101, 5.3) selezionata come ottimale all'interno della regione ottimale. La parte in grigio chiaro si trova fuori dal dominio sperimentale.

Per la scelta delle condizioni ottimizzate è stato di ausilio il grafico detto *scatterplot matrix*. Questo grafico mostra le dimensioni dello spazio del disegno (DS) e della regione

ottimale (OPT) in termini di numero di punti in una griglia che copre il dominio sperimentale degli MC, con indicati i valori delle PV. Per ogni coppia di PV si ottiene un grafico *scatterplot*, in modo da formare una matrice con i grafici in alto a destra che sono immagini speculari dei grafici in basso a sinistra.



L'esame della *scatterplot matrix* permette di trovare in modo semplice buone combinazioni per le PV. Infatti, i punti celesti indicano le combinazioni delle PV che non corrispondono ad alcuna regione ottimale (0 punti). Viceversa, è possibile individuare quale combinazione delle PV fornisce DS e OPT di ragionevoli dimensioni (idealmente tutti e due dovrebbero essere più grandi possibili, anche se i plot OPT vs. DS mostrano che i loro massimi assoluti non possono essere raggiunti contemporaneamente).

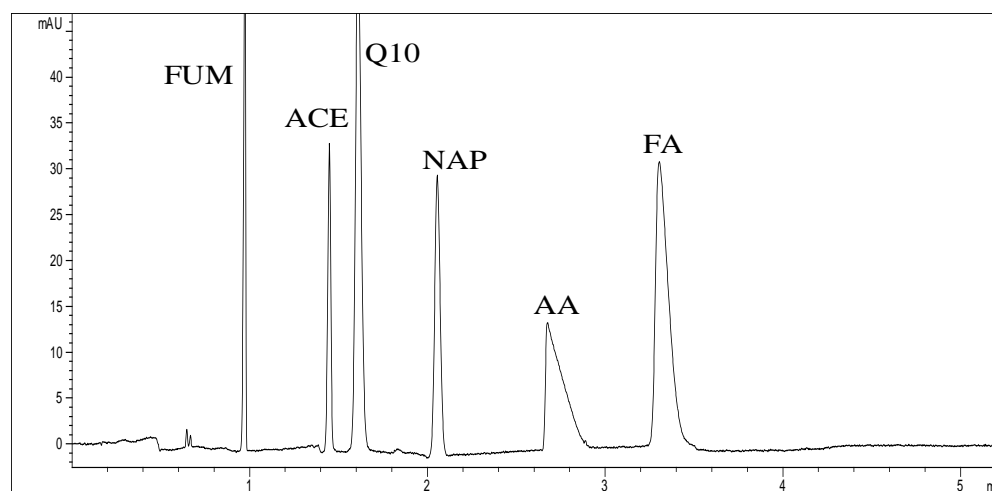
I livelli ottimali degli MC e delle PV all'interno della regione ottimale (mostrate in termini di variabili codificate e non codificate) sono di seguito riportati.

$$\begin{aligned}
 x_1 &= 0 & B &= 0.900 \\
 x_2 &= 1.000 & S &= 0.098 \\
 x_3 &= 0 & O &= 0.002
 \end{aligned}$$

$z_1 = -0.6$ $V = 23$
 $z_2 = 0.2$ $BC = 101$
 $z_3 = 0.6$ $pH = 5.3$

Questi valori sono stati scelti all'interno della regione ottimale sulla base della *scatterplot matrix* e sulla base di considerazioni pratiche. In particolare, il basso livello di olio ha permesso di ridurre il tempo di lavaggio del capillare e di preparare la microemulsione in un tempo minore. Sono stati preferiti un alto valore di tensioattivo/cotensioattivo e un basso valore di fase acquosa in modo da aumentare la forza trascinante delle gocce negative della microemulsione verso l'uscita dal capillare. Un basso valore di voltaggio ha permesso di mantenere più bassa la corrente generata, mentre è stato scelto un alto valore di pH per aumentare la mobilità elettroforetica di Q10, diretta verso l'anodo.

Di seguito è riportato l'elettroferogramma corrispondente alle condizioni ottimizzate, evidenziante una buona efficienza per tutti gli analiti (Eff.Q10=21018, Eff.AA=6443, Eff.FA=10166). I picchi risultano separati alla linea di base in circa 3 minuti.



Elettroferogramma nelle condizioni ottimizzate con modello e disegno sperimentale MPV. BGE, microemulsione composta da 90.00% fase acquosa (acetato 101 mM pH 5.30), 0.20% fase oleosa (*n*-ottano), 9.80% tensioattivo/cotensioattivo (SDS/*n*-butanolo) in rapporto 1:4.5. Lunghezza totale del capillare, 48.5 cm; lunghezza effettiva, 8.5 cm (*short end injection*). Lunghezza d'onda di rivelazione, 215 nm; differenza di potenziale, -23kV; temperatura, 20 °C.

7.4. Caratterizzazione geografica del miele mediante un approccio chemiometrico

Il miele è un alimento famoso fin dall'antichità grazie alle sue notevoli proprietà nutrizionali e agli effetti benefici per la salute dell'uomo [115]. È apprezzato in tutto il mondo, non solo come dolcificante naturale, ma anche come alimento energetico ricco di proprietà probiotiche [116-118]. Il miele viene prodotto dalle api partendo dal nettare dei fiori o dalle secrezioni di parti vive di piante, che esse raccolgono, trasformano, combinano con sostanze proprie e depongono nei loro favi [119]. La composizione del miele è piuttosto variabile e in stretta relazione con la sua origine botanica e geografica. La composizione del nettare e della melata può variare molto a seconda della specie botanica di provenienza ma altri fattori esterni come l'inquinamento ambientale e i processi tecnologici possono modificare molto le proprietà organolettiche e nutrizionali del prodotto. La quantità di sali minerali presenti nel miele ammonta a circa 0.1-0.2% e derivano tutti dal terreno [120-122]. I sali minerali attraverso le radici delle piante arrivano fino al nettare dei fiori che poi viene raccolto direttamente dall'ape. Così, la valutazione delle concentrazioni di sali minerali può essere utile per la classificazione del miele in base alla sua origine geografica. Questo studio ha lo scopo di mostrare come il contenuto di sali minerali del miele possa essere usato come indicatore di provenienza del miele stesso. La novità di questo lavoro è stato il trattamento statistico di un'ampia gamma di campioni di miele che includono cationi e anioni ottenuti mediante cromatografia a scambio ionico (Ion Chromatography, IC). Cinquanta campioni di miele di diversa origine geografica sono stati divisi in tre classi: Paesi Europei dell'Est (Serbia, Slovenia, Kosovo, Macedonia, Croazia e Albania), Paesi Italiani, Paesi Equatoriali e altro (Camerun, Kenya e Brasile). Ogni campione è stato caratterizzato da 13 variabili quali: anioni e cationi ottenuti mediante IC e due parametri che misurano il potere antiossidante dei campioni di miele. Inoltre sono stati analizzati due dataset diversi, uno con i dati originali e un altro con i dati normalmente distribuiti.

Mediante l'algoritmo Duplex [123] i campioni sono stati divisi in un training e test set contando rispettivamente (35) e (15) campioni. Metodi di analisi delle componenti

principali (PCA) e di analisi gerarchica del cluster (HCA), sono stati applicati per cercare le proprietà principali dei dati; metodi di analisi lineare discriminante e quadratica (LDA, QDA) e il metodo di classificazione ad albero (CART), sono stati utilizzati per trovare possibile relazioni tra le variabili come per esempio la concentrazione di sali minerali e l'origine geografica del miele. Prima di PCA, HCA, LDA e QDA i dati sono stati autoscalati mentre prima dell'analisi CART questo non è stato necessario.

Campioni di Miele

Sono stati analizzati un totale di m=50 campioni di miele commerciale e artigianale comprato in loco nei diversi paesi. Ogni gruppo in cui è stata fatta la suddivisione geografica ha incluso campioni di diversa origine botanica (acacia, castagno, pino, eucalipto, millefiori, abete). Il gruppo *Paesi Europei dell'Est* era costituito da: 9 campioni della Serbia, 2 della Slovenia, 7 del Kosovo, 2 della Macedonia, 2 della Croazia e 8 dell'Albania; il gruppo *Paesi Equatoriali e Altro* era formato da: 3 campioni provenienti dal Cameroun, 1 dal Kenya e 2 dal Brasile; il gruppo *Paesi Italiani* era costituito da 14 campioni di diverse regioni italiane.

Determinazione delle Variabili

Ogni campione era caratterizzato da 13 variabili: sodio [Na^+], ammonio [NH_4^+], potassio [K^+], magnesio [Mg^{2+}], calcio [Ca^{2+}], cloro [Cl^-], bromo [Br^-], nitrato [NO_3^-], fosfato [PO_4^{3-}], solfato [SO_4^{2-}], acido gluconico, e la total phenolic concentrations [TPC] e la radical scavenging capacity (RSC). La misura delle specie cationiche e anioniche è stata effettuata mediante cromatografia a scambio ionico. Prima dell'analisi i campioni di miele sono stati disciolti in acqua Milli-Q (1g solubilizzati in 10mL di acqua e agitati vigorosamente per 5 minuti e filtrati attraverso una membrana con filtro 0.45) e questa soluzione ottenuta direttamente analizzata con IC. La concentrazione di ogni ione è stata calcolata mediante una curva di calibrazione iniettando corrette diluizioni di una standard solution 1000ppm. La concentrazione fenolica totale è stata determinata usando una modificazione del tradizionale metodo Folin-Ciocalteu [124, 125,

126] e i risultati sono stati espressi in mg di acido gallico equivalenti (GAE) per Kg di miele. I campioni di diversi tipi di miele sono stati trattati con acqua calda distillata (500mg/5mL acqua), sonicati per 5 minuti fino ad ottenere una soluzione limpida. Poi, 0.1mL della soluzione, corrispondenti a 10mg di miele fresco, sono stati aggiunti a al reattivo Folin-Ciocalteu precedentemente diluito 1:10 in acqua distillata. La miscela è stata agitata con magnete per circa 2 minuti e il contenuto versato in una cuvetta da 1.5mL; assorbanza è stata determinata dopo 20 minuti alla lunghezza d'onda di 750nm. La radical scavenging capacity è stata invece valutata secondo il metodo di Brand-Williams con minori cambiamenti [125-127]. La miscela test conteneva: 1.9mL di 1,1-difenil-2-picrilidrazolo (DPPH) disciolto in etanolo assoluto, 1 mL di tampone acetato 100mM pH 5.5 e 0.1mL di una soluzione di miele contenente da 30 a 600mg/mL del miele di partenza. Il volume finale della miscela era di circa 3mL la quale è stata tenuta in un bagnetto al buio a 25°C per 90 minuti. L'attività di scavenging è stata espressa come percentuale di diminuzione del DPPH di partenza. Le analisi sono sempre state replicate in modo da avere una deviazione standard sempre sotto il 5%.

Data Set e Calcoli

I dati sono stati organizzati in una matrice \mathbf{X} $m \times n$, dove m rappresenta il numero dei campioni di miele mentre n il numero delle variabili. Il vettore risposta y , $m \times 1$ consiste in tre diverse classi correlate all'origine geografica.

Il primo data set è formato dalle variabili originali (data set 1); il secondo data set di dati (data set 2) invece a causa di alcune variabili non normalmente distribuite è stato formato da alcune variabili trasformate con la funzione logaritmica quali: $[Na^+]$, $[NH_4^+]$, $[Mg^{2+}]$, $[Ca^{2+}]$, $[Cl^-]$, $[Br^-]$, $[NO_3^-]$, $[PO_4^{3-}]$, $[SO_4^{2-}]$, and [TPC]. Tutti i calcoli sono stati effettuati mediante il software Matlab 7.1 (Mathworks, Natick, MA).

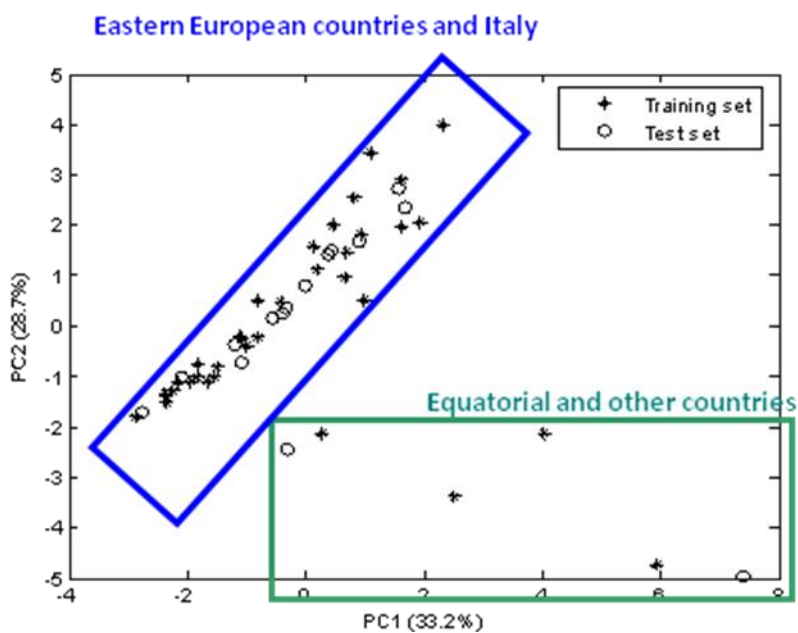
Analisi delle componenti principali e analisi dei Cluster

L'analisi delle componenti principali è stata usata per investigare la struttura multivariata dei dati e mettere in evidenza le variabili più significative.

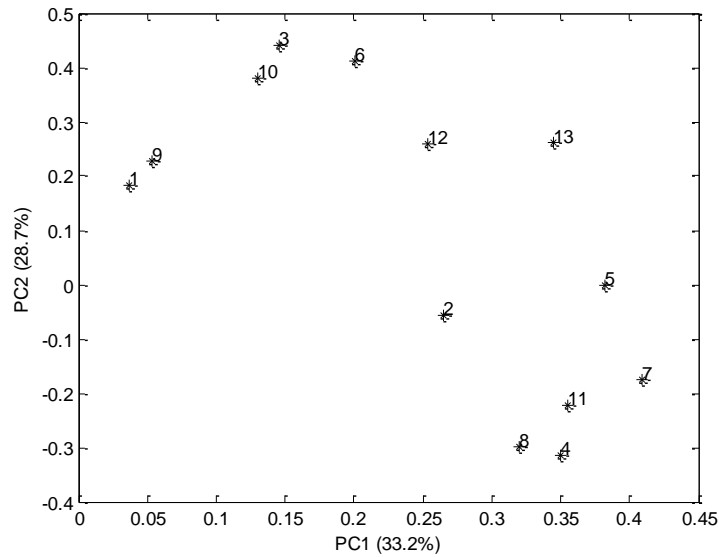
La PCA è stata applicata a una matrice X 50×13 , dove 50 erano i campioni di miele e 13 il numero di variabili, sia per quanto riguarda dataset 1 che per il dataset 2. Prima della PCA i dati sono stati autoscalati.

Sono state scelte le prime due componenti principali per esaminare il dataset 1. La percentuale di varianza totale spiegata dalle singole componenti era pari al 61.94%, rispettivamente 33.23% descritti dalla PC1 e 28.71% descritti dalla PC2. In figura sono riportati rispettivamente lo score e loading plot.

(a)



(b)



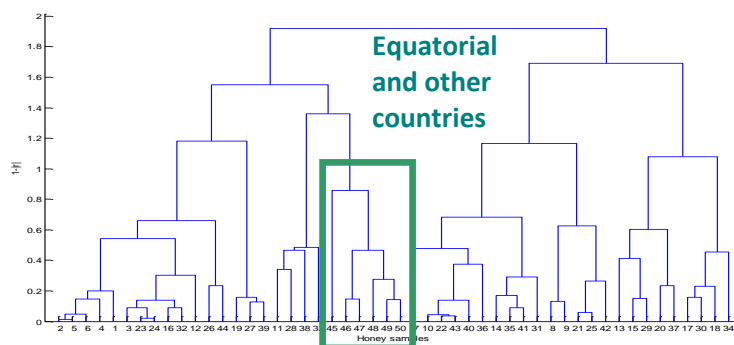
PC1-PC2 (a) score and (b) loading plot of the Original data set 1 after autoscaling. Legend: 1) $[Na^+]$, 2) $[NH_4^+]$, 3) $[K^+]$, 4) $[Mg^{2+}]$, 5) $[Ca^{2+}]$, 6) Gluconic Acid 7) $[Cl^-]$, 8) $[Br^-]$, 9) $[NO_3^-]$, 10) $[PO_4^{3-}]$, 11) $[SO_4^{2-}]$, 12) total phenolic concentration (TPC), and 13) radical scavenging capacity (RSC)

I 50 campioni di miele caratterizzati dalle 13 variabili, sono stati visualizzati in uno spazio bidimensionale descritto dalle due prime componenti principali (PC1 e PC2) in modo da ottenere informazioni riguardo alle similarità e dissimilarità degli oggetti. Come mostra la figura a, i campioni di miele con caratteristiche simili si trovano l'uno vicino all'altro, mentre quelli con proprietà diverse sono lontani. Si può così chiaramente notare due gruppi ben separati: uno omogeneo e corrispondente ai Paesi equatoriali e altro, uno più grande, corrispondente ai Paesi Europei dell'est e Italiani. Questa separazione si osserva in particolare lungo la PC2.

Dall'analisi della figura b invece, appare evidente che le variazioni lungo la PC1 riguardano le variabili 1) $[Na^+]$, 9) $[NO_3^-]$, 5) $[Ca^{2+}]$ e 7) $[Cl^-]$, mentre le variabili 4) $[Mg^{2+}]$, 8) $[Br^-]$, 11) $[SO_4^{2-}]$, 3) $[K^+]$, 6) acido gluconico e 10) $[PO_4^{3-}]$ sono associate alla variazione lungo la PC2. Alcune di queste variabili possono infatti essere correlate

all'inquinamento ambientale [128]. Molti tipi di fertilizzanti vengono utilizzati per aumentare le produzioni agricole: l'obiettivo essenziale della concimazione è quello di aggiustare il terreno ad un livello nutritivo sufficiente per alimentare le piante e di mantenere il suo potenziale nutritivo in funzione dei bisogni futuri delle colture. Fosfati, nitrati, bromuri, solfati, sono tutti elementi che vengono utilizzati o come fertilizzanti o come pesticidi.

La conferma della PCA arriva inoltre dall'analisi gerarchica del cluster dopo l'autoscalatura dei dati. HCA crea dei gruppi con i campioni più simili tra di loro non tenendo conto della classe di cui fanno parte. I dendrogrammi sono creati usando tre distanze di misura (Distanza Euclidea, Distanza di Mahalanobis e $1-|r|$) e sette metodi di collegamento (single linkage, complete linkage, unweighted average linkage, weighted average linkage, centroid's method, median method e Ward's method) [129]. La distanza $1 - |r|$ ha assieme al collegamento chiamato "complete linkage" ha permesso di ottenere il migliore dendrogramma mostrato in figura che permette una visione globale degli oggetti riguardo alla loro somiglianze.

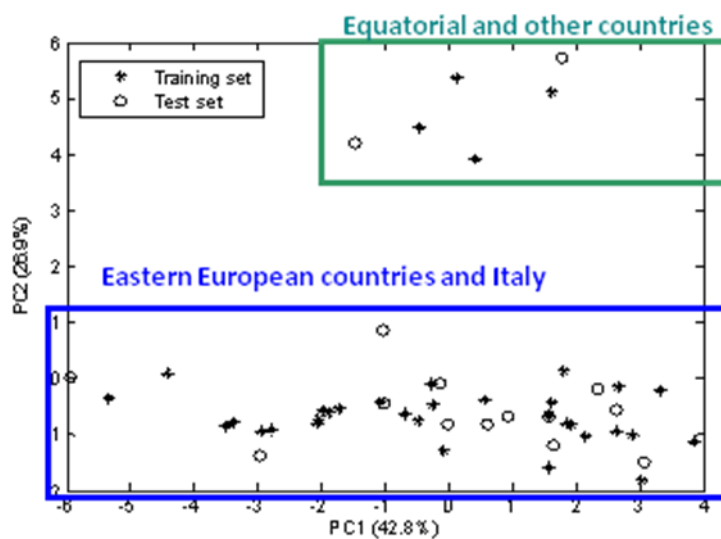


HCA dendrogram obtained using $1 - |r|$ as distance metric and complete linkage of the Original data set 1 after autoscaling.

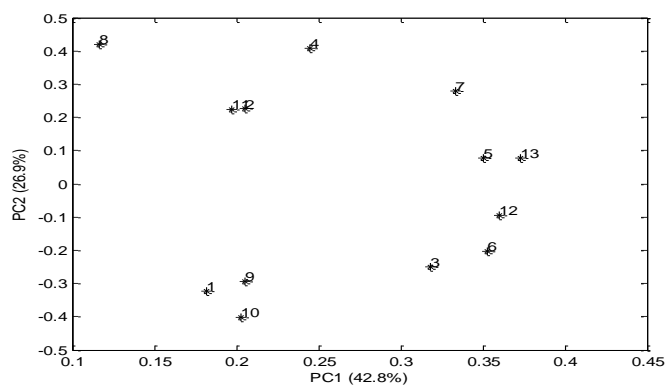
Per quanto riguarda invece il data set 2, la varianza totale spiegata dalle due prime componenti principali è lievemente incrementata pari al 69.70%. In particolare la PC1 ha

spiegato il 42.8% della varianza totale e la PC2 il 26.9%. Nella figura qui sotto riportata sono mostrati il loading e score plot relativo alle PC1-PC2.

(a)

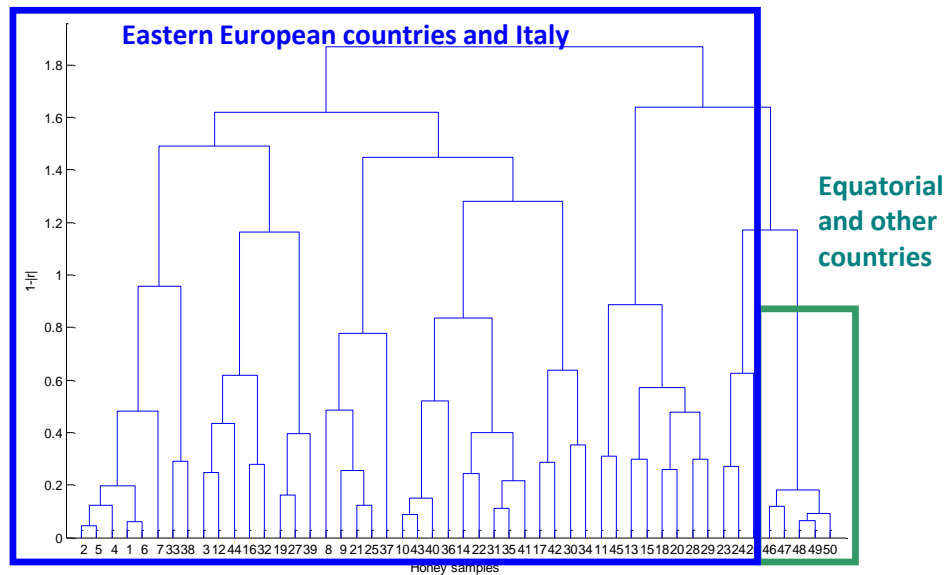


(b)



PC1-PC2 (a) score and (b) loading plot of the Normally distributed data set 2 after autoscaling. Legend: 1 = $\log([Na^+])$, 2) $\log([NH_4^+])$, 3) $[K^+]$, 4) $\log([Mg^{2+}])$, 5) $\log([Ca^{2+}])$, 6) Gluconic Acid, 7) $\log([Cl^-])$, 8) $\log([Br^-])$, 9) $\log([NO_3^-])$, 10) $\log([PO_4^{3-}])$, 11) $\log([SO_4^{2-}])$, 12) $\log(\text{TPC})$, and 13) (RSC).

Anche in questo caso è possibile notare due gruppi ben separati lungo la PC2 corrispondenti ai Paesi Europei dell'est e italiani e ai Paesi Equatoriali e altro (figura a). Il loading plot (figura b) mostra che le variazioni lungo la PC1 riguardano le variabili 8) $\log([Br^-])$, 12) $\log([TPC])$ e 13) RSC mentre le variabili 8) $\log([Br^-])$, 4) $\log([Mg^{2+}])$, 10) $\log([PO_4^{3-}])$, 1) $\log([Na^+])$, e 9) $\log([NO_3^-])$ riguardano la PC2. Anche l'analisi gerarchica del cluster ha dato migliori risultati nel data set 2 rispetto al dataset 1: il migliore dendrogramma è stato ottenuto con la medesima misura di distanza e collegamento ma il gruppo Paesi Equatoriali e altro trovandosi all'estremità del grafico ha indicato una migliore separazione.



HCA dendrogram obtained using $1 - |r|$ as distance metric and complete linkage of the Normally distributed data set 2 after autoscaling

Metodi di Classificazione Multivariata

I dati sono stati divisi in training e test set mediante algoritmo Duplex con il fine di valutare l'abilità predittiva del modello. La divisione è stata fatta individualmente per ogni classe in modo che il 70% di ogni classe fosse nel training set. Così, il training set era costituito da 35 campioni (21 Paesi Europei dell'Est, 10 Italiani, 4 Paesi Equatoriali e altro) e il test set da 15 campioni (9 Paesi Europei dell'Est, 4 Italiani, 2 Paesi Equatoriali e altro). I risultati dei modelli di classificazione mediante LDA, QDA e CART per i due data set considerati sono mostrati in tabella.

(a) Original data set 1

Method	Information	Considered variables	%CCR	%CCR	%CCR
			TRAINING-CV	TRAINING	TEST
LDA	5 variables	[Br^-], [PO_4^{3-}], [NO_3^-], [Mg^{2+}], [Ca^{2+}]	77.14%	82.86%	60.00%
QDA	7 variables	[Br^-], [PO_4^{3-}], [NO_3^-], [Mg^{2+}], [Ca^{2+}], Acido Gluconico, [K^+]	71.43%	88.57%	53.33%
CART*	Gini, Twoing # TN = 3, 2 variables	[Mg^{2+}], Acido Gluconico	82.86%	82.86%	60%
	Deviance, # TN = 4, 3 variables	[Mg^{2+}], [Br^-], Acido Gluconico	91.43%	91.43%	66.67%

(b) Normally distributed data set 2

Method	Information	Considered variables	%CCR	%CCR	%CCR
			TRAINING-CV	TRAINING	TEST
LDA*	5 variables	log([PO_4^{3-}]), log([TPC]), log([Mg^{2+}]), RSC, log ([Ca^{2+}])	94.29%	94.29%	60.00%
QDA	2 variables	log([PO_4^{3-}]), log([TPC])	77.14%	88.57%	66.67%
CART	Gini and twoing, # TN = 3, 2 variables	log([Mg^{2+}]), Acido Gluconico	82.86%	82.86%	60.00%
	Deviance, # TN = 4 3 variables	log([Mg^{2+}]), log([Br^-]), Acido Gluconico	91.43%	91.43%	66.670%

Table Classification in three classes: Method = Applied classification method; Some background information: Considered variables. Percentage correct classification rate (%CCR) of the training set samples using cross-validation (%CCR_{TRAINING-CV}), of the training set samples without CV (%CCR_{TRAINING}), and of the independent test set samples (%CCR_{TEST}), were reported. The models with the best predictive ability are indicated (*). # TN = Number of Terminal Nodes

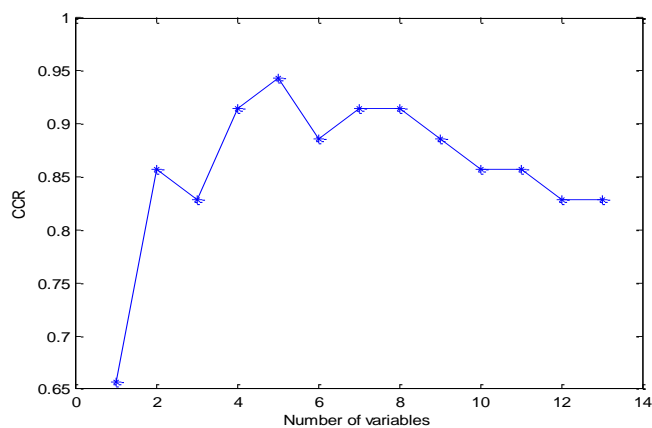
Analisi lineare e quadratica discriminante è stata effettuata dopo aver autoscalato dati. Inoltre la classificazione stepwise è stata utilizzata in modo da ordinare le variabile secondo il loro grado di importanza.

Per primo è stato studiato il dataset 1. Sia nel caso dell'LDA che della QDA sono state indicate le variabili responsabili della classificazione del modello: 5 per l'analisi lineare discriminante quali [Br^-], [PO_4^{3-}], [NO_3^-], [Mg^{2+}] [Ca^{2+}] e 7 per l'analisi quadratica discriminante come [Br^-], [PO_4^{3-}], [NO_3^-], [Mg^{2+}], [Ca^{2+}], acido gluconico e [K^+]. Così la classificazione lineare è risultata più semplice con un minor

numero di variabili coinvolte e un coefficiente di correlazione cross validato in percentuale maggiore pari a 77.14%.

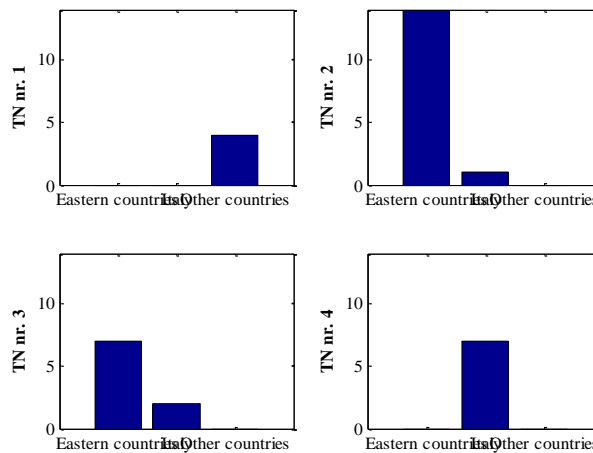
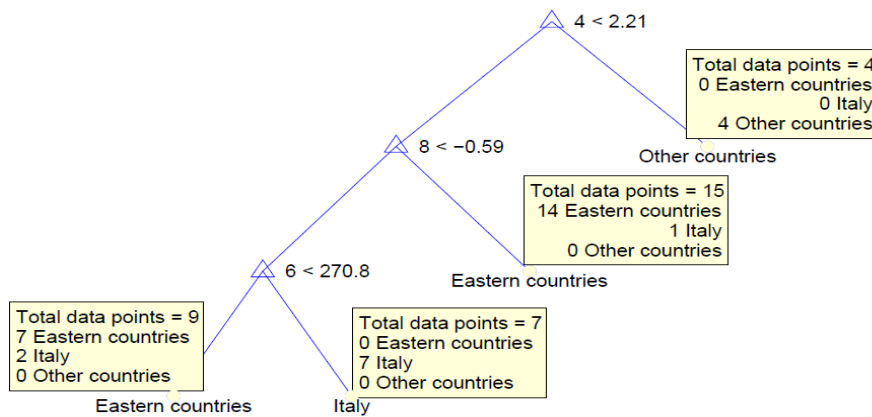
Inoltre LDA e QDA sono state applicate anche al data set 2. In questo caso l'analisi lineare discriminata ha portato ad ottimi risultati arrivando a un coefficiente di correlazione pari al 94.29%.

Il grafico qui sotto riportato mostra la funzione del numero delle cinque differenti variabili coinvolte nella classificazione.



LDA on the Normally distributed data set 2 after autoscaling: CCR as a function of the number of variables

Il metodo CART è stato applicato a entrambi i data set utilizzando dati non auto scalati. I risultati ottenuti molto simili tra di loro hanno permesso di indicare le variabili responsabili della classificazione. Sono stati utilizzati tre diversi criteri di “splitting” (Gini, Twoing, Deviance) con buoni risultati. L'albero decisionale mostrato in figura ottenuto con il criterio deviance mostra la migliore capacità predittiva. L'albero è formato da 4 nodi terminali e da 3 variabili di splitting quali: $[Mg^{2+}]$, $[Br^-]$, acido gluconico per il data set 1 e $\log([Mg^{2+}])$, $\log([Br^-])$, acido gluconico per il data set 2. Si può inoltre notare la distribuzione dei 35 campioni del training set lungo i nodi terminali: 21 Paesi Europei dell'est, 10 Italiani e 4 Paesi Equatoriali e altro. Nell'albero si possono inoltre visualizzare due nodi puri corrispondenti a TN1 con 4 campioni dei Paesi Equatoriali e altro e TN4 con 7 campioni italiani.



CART: (a) Classification tree with the best predictive ability for Original data set 1 and Normally distributed data set 2; and (b) The distribution of the training set samples over the four terminal nodes

In conclusione sulla base della semplicità dell'interpretazione delle risposte e il maggior coefficiente di correlazione cross validato, il miglior metodo di classificazione per entrambi i data set è stato il metodo CART. Inoltre passando dai dati originali autoscalati a quelli normalmente distribuiti si è potuto notare un incremento della % CV e nel caso dell'analisi lineare discriminante si sono ottenuti risultati soddisfacenti.

8. Conclusioni

Il lavoro svolto durante il Dottorato di Ricerca ha permesso di investigare a fondo le potenzialità della cromatografia elettrocinetica capillare in analisi farmaceutica, attraverso lo studio e lo sviluppo di metodi che coinvolgevano diverse sue modalità operative.

L'elettroforesi capillare ha confermato essere uno strumento analitico flessibile nell'utilizzo, dal momento che consente la modulazione di diverse condizioni operative al fine di ottenere la separazione desiderata e tempi brevi di analisi.

L'utilizzo sistematico di un approccio multivariato ha confermato le sue potenzialità soprattutto in presenza di sistemi complessi, non solo per l'ottenimento delle condizioni ottimizzate finali, ma anche per la ricerca iniziale del corretto dominio sperimentale da esplorare.

La chemiometria si è quindi consolidata tecnica statistica complementare alle tecniche analitiche per garantire affidabilità dei dati sperimentali e maggior comprensione dei fenomeni in studio.

È tuttavia importante sottolineare che il corretto uso delle tecniche chemiometriche è di notevole aiuto sia in analisi farmaceutica, portando a importanti benefici in termini di tempo impiegato, costi sostenuti e risultati ottenuti, ma anche in qualsiasi altro studio dove si abbia l'obiettivo di conoscere le relazioni tra le variabili in esame e l'ottimizzazione del sistema sperimentale.

9. Bibliografia

- [1] D.L. Massart, B.G.M Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics: Part A*, 1997, Elsevier, Amsterdam.
- [2] K.D. Altria, M.A. Kelly, B.J. Clark, *Trends Anal. Chem.* 17 (1998) 214-226.
- [3] K.D. Altria, A. Marsh, C. Sänger-van de Griend, *Electrophoresis* 27 (2006) 2263-2282.
- [4] U. Holzgrabe, D. Brinz, S. Kopec, C. Weber, Y. Bitar, *Electrophoresis* 27 (2006) 2283-2292.
- [5] S. Orlandini, S. Fanali, S. Furlanetto, A.M. Marras, S. Pinzauti, *J. Chromatogr. A* 1032 (2004) 253-263.
- [6] S. Orlandini, S. Furlanetto, S. Pinzauti, G. D'Orazio, S. Fanali, *J. Chromatogr. A* 1044 (2004) 295-303.
- [7] S. Furlanetto, S. Orlandini, A.M. Marras, P. Mura, S. Pinzauti, *Electrophoresis* 27 (2006) 805-818.
- [8] S. Orlandini, I. Giannini, R. Gotti, S. Pinzauti, E. La Porta, S. Furlanetto, *Electrophoresis* 28 (2007) 395-405.
- [9] S. Furlanetto, S. Orlandini, I. Giannini, G. Beretta, S. Pinazuti, *Electrophoresis* 30 (2009) 633-643.
- [10] I. Giannini, S. Orlandini, R. Gotti, S. Pinzauti, S. Furlanetto, *Talanta* 80 (2009) 781-788.
- [11] S. Orlandini, I. Giannini, M. Villar Navarro, S. Pinzauti, S. Furlanetto, *Electrophoresis* 31 (2010) 3296-3304.
- [12] G.A. Lewis, D. 113, R. Phan-Tan-Luu, *Pharmaceutical Experimental Design*, Marcel Dekker, New York, 1999.
- [13] J.L. Goupy, *Methods for Experimental Design*, Elsevier, Amsterdam, 1992.
- [14] I.E. Frank, R. Todeschini (Eds.), *The Data Analysis Handbook*, Elsevier, Amsterdam, 1994.
- [15] A. Peissik, *Methodologie de la recherche experimentale: proprietes et caracteristiques des matrices d'experiences pour les modeles polynomiaux du second*

degre, Tesi di Dottorato in Scienze, Marseille, 1995.

[16] S.N. Deming, S.L. Morgan, *Experimental Design: a Chemometric Approach*, Elsevier, Amsterdam, 1993.

[17] T. Lundstedt, E. Seifert, L. Abramo, B. Thelin, Å. Nyström, J. Pettersen, R. Bergman, *Chemometr. Intell. Lab. Syst.* 42 (1998) 3-40.

[18] Y. Vander Heyden, D.L. Massart in M.M.W.B. Hendriks, J.H. de Boer, A.K. Smilde, *Robustness of Analytical Chemical Methods and Pharmaceutical Technological Products*, Elsevier, Amsterdam, 1996.

[19] R. Carlson, *Design and Optimization in Organic Synthesis*, Elsevier, Amsterdam, 1992.

[20] A. Atkinson, A. Donev, R. Tobias, *Optimum Experimental Designs, With SAS*, Oxford University Press, 2007, Oxford.

[21] J.A. Cornell, *Experiments with Mixtures (3rd Edn.)*, John Wiley & Sons, New York, 2002.

[22] R.G. Brereton, *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*, J. Wiley & Sons, Chichester, UK, 2003.

[23] S.L. Costa Ferreira, R.E. Bruns, E.G. Paranhos da Silva, W.N. Lopes dos Santos, C.M. Quintella, J.M. David, J. Bittencourt de Andrade, M.C. Breitzkreitz, I.C. Sales Fontes Jardim, B.B. Neto, *J. Chromatogr. A* 1158 (2007) 2-14.

[24] R.H. Myers, D.C. Montgomery, C. Anderson-Cook, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments (3rd Edn.)*, John Wiley & Sons, New York, 2009.

[25] 25, *Design-Expert V8*, Minneapolis, MN.

[26] 26. *JMP V9*, SAS Institute, Inc., Cary, NC.

[27] 27. *Minitab V16*. Minitab Inc., State College, PA.

[28] C.M. Anderson-Cook, H.B. Goldfarb, C.M. Borror, D.C. Montgomery, K.G. Canter, J.N. Twist, *Pharm. Stat.* 3 (2004) 247–260.

[29] R. Todeschini, *Introduzione alla Chemiometria*, Edises, Napoli, 1998.

[Vandeginste 1998] B.G.M. Vandeginste, D.L. Massart, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics: Part B*, Elsevier, Amsterdam, 1998.

- [31] E. Van Gyseghem, B. Dejaegher, R. Put, P. Forlay-Frick, A. Elkihel, M. Daszykowski, K. Héberger, D.L. Massart, Y. Vander Heyden, *J. Pharm. Biomed. Anal.* 41 (2006) 141-151.
- [32] R.A. Fisher, *Ann. Eugenics* 7 (1936) 179-188.
- [33] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, *Classification and Regression Trees*, Wadsworth, Pacific Grove, California, 1984.
- [34] R.M. Silverstein, *Spectrometric Identification of Organic Compounds*, 2^a Ed., Casa editrice ambrosiana, Milano 2006.
- [35] R.R. Ernst, G. Bodenhausen, A. Wokaun, *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Clarendon Press, Oxford, 1987.
- [36] C.F. Poole, *The Essence of Chromatography*, Elsevier, Amsterdam, 2003.
- [37] S.F.Y. Li, *Capillary Electrophoresis*, Elsevier, Amsterdam, 1992.
- [38] D.C. Harris, *Chimica Analitica Quantitativa*, 2^a Ed., Zanichelli, Bologna, 2005.
- [39] D.N. Geiger, *High Performance Capillary Electrophoresis-An Introduction*, Hewlett-Packard, Waldbronn, 1992.
- [40] F. Tagliaro, F.P. Smith, *TRAC-Trends Anal. Chem.* 15 (1996) 513-525.
- [41] H. Watarai, *Chem. Lett.* 3 (1991) 391-394.
- [42] P. Kumar, K.L. Mittal, *Handbook of Microemulsion Science and Technology*, Marcel Dekker, New York, 1999.
- [43] K.D. Altria, *J. Chromatogr. A* 892 (2000) 171-186.
- [44] S. Terabe, N. Matsubara, Y. Ishihama, Y. Okada, *J. Chromatogr.* 488 (1992) 23-29.
- [45] Y. Ishihama, Y. Oda, K. Uchikawa, N. Asakawa, *Anal. Chem.* 54 (1995) 1995-1475.
- [46] S.J. Gluck, M.H. Benko, R.K. Hallberg, K.P. Steele, *J. Chromatogr. A* 744 (1996) 141-146.
- [47] L. Song, Q. Ou, W. Yu, G. Li, *J. Chromatogr. A* 699 (1995) 371-382.

- [48] M.F. Miola, M.J. Snowden, K.D. Altria, *J. Pharm. Biomed. Anal.* 18 (1998) 785-797.
- [49] S. Katsuta, K. Saitoh, *J. Chromatogr. A* 780 (1997) 152-178.
- [50] M.H. Abraham, C. Treiner, M. Roses, C. Rafols, Y. Ishihama, *J. Chromatogr. A* 752 (1996) 363-369.
- [51] R. Szücs, E. Van Hove, P. Sandra, *J. High Resolut. Chromatogr.* 19 (1996) 189-192.
- [52] A. Berthod, M. De Carvalho, *Anal. Chem.* 51 (1992) 2254-2272.
- [53] X. Fu, J. Lu, A. Zhu, *J. Chromatogr. A* 735 (1996) 341-344.
- [54] S.H. Hansen, C. Gabel-Jensen, D.T.M. El-Sherbiny, S. Pedersen-Bjergaard, *TrAC-Trends Anal. Chem.* 20 (2001) 494-499.
- [55] N. Gorski, M. Gradzielski, H. Hoffmann, *Ber. Bunsenges. Phys. Chem.* 100 (1996) 1109-1117.
- [56] J.H.T. Luong, A.L. Nguyen, *J. Chromatogr. A* 792 (1997) 431-444.
- [57] M. Blanco, I. Valverde, *TRAC-Trend. Anal. Chem.* 22 (2003) 428-439.
- [58] S.A.C. Wren, R.C. Rowe, *J. Chromatogr.* 603 (1992) 235-241.
- [59] S. Nie, R. Dadoo, R.N. Zare, *Anal. Chem.* 65 (1993) 3571-3575.
- [60] S. Terabe, Y. Miyashita, Y. Ishihama, O. Shibata, *J. Chromatogr.* 636 (1993) 47-55.
- [61] M.J. Sepaniak, C.L. Copper, K.W. Whitaker, V.C. Anigbogu, *Anal. Chem.* 67 (1995) 2037-2041.
- [62] O.H.J. Szolar, R.S. Brown, J.H.T. Luong, *Anal. Chem.* 67 (1995) 3004-3010.
- [63] B. Chankvetadze, W. Lindner, G.K.E. Scriba, *Anal. Chem.* 76 (2004) 4256-4260.
- [64] C.J. Easton, S.F. Lincoln, *Chem. Soc. Rev.* 25 (1996) 163-170.
- [65] *ICH Harmonised Tripartite Guideline, Q2(R1), Validation of Analytical Procedures: Text and Methodology*, 2005.

- [66] M.E. Swarz, I.S. Krull, *Analytical Method Development and Validation*, Marcel Dekker, New York, 1997.
- [67] J. Ermer, *J. Pharm. Biomed. Anal.* 24 (2001) 755-767.
- [68] J. Ermer, J. H. McB. Miller (Eds.), *Method Validation in Pharmaceutical Analysis-A Guide to Best Practice*, Wiley-VCH, Weinheim 2004.
- [69] S. Furlanetto, S. Orlandini, P. Mura, M. Sergent, S. Pinzauti, *Anal. Bioanal. Chem.* 377 (2003) 937-944.
- [70] Pharmaceutical CGMPs for the 21st century – A Risk-Based Approach. Final report (2004) U.S Food and Drug Administration.
- [71] Guidance for Industry. PAT - A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance (2004) U.S Food and Drug Administration.
<http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm070305.pdf>
- [72] Guidance for Industry. Q8(R2) Pharmaceutical Development (2009) U.S Food and Drug Administration.
<http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm073507.pdf>
- [73] ICH Harmonised Tripartite Guideline. Pharmaceutical Development Q8(R2) (2009) International Conference on Harmonisation of technical requirements for registration of pharmaceuticals for human use.
- [74] ICH Harmonised Tripartite Guideline. Quality Risk Management Q9 (2005) International Conference on Harmonisation of technical requirements for registration of pharmaceuticals for human use.
- [75] ICH Harmonised Tripartite Guideline. Pharmaceutical Quality Systems Q10 (2008) International Conference on Harmonisation of technical requirements for registration of pharmaceuticals for human use.
- [76] Li, Y., D.Q., Yang, S., Sudini, R., McGuire, M.A., Bhanushali, D.S., Kord, A.S.J. *Pharm. Biomed. Anal.* 52 (2010) 493-507.

- [77] Lopes J.A., Costa P., Alves T., Menezes J., *Chemom. Intell. Lab. Sys.* 74 (2004) 269-275.
- [78] Mennini N, Furlanetto S, Cirri M, Mura P (2012) *Eur J Pharm Biopharm* 80:67-75.
- [79] Verma S, Lan Y, Gokhale R, Burgess DJ (2009) *Int J Pharm* 377:185-198
- [80] Adam S, Suzzi D, Radeke C, Khinast JG (2011) *Eur J Pharm Sci* 42:106-115
- [81] Mennini N, Furlanetto S, Cirri M, Mura P (2012) *Eur J Pharm Biopharm* 80:67-75.
- [82] Rosas JG, Blanco M, González JM, Alcalá M (2011) *J Pharm Sci* 100:4432-4441.
- [83] Wu H, White M, Khan MA (2011) *Int J Pharm* 405:63-78.
- [84] Gavin PF, Olsen BA (2008) *J Pharm Biomed Anal* 46:431-444.
- [85] Hibbert DB (2012) Experimental design in chromatography: A tutorial review. *J Chromatogr B* Doi:10.1016/j.jchromb.2012.01.020.
- [86] Hanrahan G, Montes R, Gomez FA *Anal Bioanal Chem* (2008) 390:169-179
- [87] Vogt FG, Kord AS (2011) *J Pharm Sci* 100:797-812
- [88] Piepel G, Pasquini B, Cooley S, Heredia-Langner A, Orlandini S, Furlanetto S (2012) Mixture-process variable approach to optimize a microemulsion electrokinetic chromatography method for the quality control of a nutraceutical based on coenzyme Q10. *Talanta* Doi:10.1016/j.talanta.2012.03.064.
- [89] Molnár I, Rieger HJ, Monks KE (2010) *J Chromatogr A* 1217:3193-3200.
- [90] McBrien M (2010) *Chromatogr Today* 3:30-34.
- [91] Debrus B, Lebrun P, Ceccato A, Caliaro G, Rozet E, Nistor I, Oprean R, Rupérez FJ, Barbas C, Boulanger B, Hubert P (2011) *Anal Chim Acta* 691:33-42.
- [92] Debrus P, Lebrun P, Mbinze Kindenge J, Lecomte F, Ceccato A, Caliaro G, Mavar Tayey Mbay J, Boulanger B, Marini RD, Rozet E, Hubert Ph (2011) *J. Chromatogr A* 1218:5205-5215.
- [93] Lionberger RA, Lee SL, Lee LM, Raw A, Yu LX (2008) *AAPS J* 10:268-276.
- [94] Borman P, Nethercote P, Chatfield M, Thompson D, Truman K (2007) *Pharm Tech* 31:142-152.
- [95] Ishikawa K (1985) *What is Total Quality Control? The Japanese Way*, Prentice-Hall, Englewood Cliffs, NJ, 63-64.
- [96] Monks KE, Rieger H-J, Molnár I (2011) *J Pharm Biomed Anal* 56:874-879.

- [97] S. Orlandini, S. Pinzauti, S. Furlanetto, Application of quality by design to the development of analytical separation methods. *Anal. Bional. Chem.*, in press, DOI 10.1007/s00216-012-6302-2.
- [98] Huang J, Kaul G, Cai C, Chatlapalli R, Harnandez-Abad P, Ghosh K, Nagi A (2009) *Int J Pharm* 382:23-32.
- [99] Molnár I, Monks KE (2011) *Chromatographia* 73:S5-S14.
- [100] Awotwe-Otoo D, Agarabi C, Faustino PJ, Habib MJ, Lee S, Khan MA, Shah RB (2012) *J Pharm Biomed Anal* 62:61-67.
- [101] Krull I, Swartz M, Turpin J, Lukulay PH, Verseput R (2008) *LCGC Am* 26:1190-1197.
- [102] Krull I, Swartz M, Turpin J, Lukulay PH, Verseput R (2009) *LC GC Am* 27:48-61.
- [103] Orlandini S, Gotti R, Giannini I, Pasquini B, Furlanetto S (2011) *J. Chromatogr. A* 1218:2611-2617.
- [104] Furlanetto S, Cirri M, Piepel G, Mennini N, Mura P (2011) *J Pharm Biomed Anal* 55:610-617
- [105] Lebrun P, Krier F, Mantanus J, Grohganz H, Yang M, Rozet E, Boulanger B, Evrard B, Rantanen J, Hubert P (2012) *Eur J Pharm Biopharm* 80:226-234
- [106] Derringer G, Suich R (1980) *J Qual Technol* 12:214-219
- [107] Lebrun P, Govaerts B, Debrus B, Ceccato A, Caliaro G, Hubert P, Boulanger B (2008) *Chemom Intell Lab Syst* 91:4-16
- [108] De Backer B, Debrus B, Lebrun P, Theunis L, Dubois N, Decock L, Verstraete A, Hubert Ph, Charlier C (2009) *J Chromatogr B* 877:4115-4124
- [109] Monks K, Molnár I, Rieger H-J, Bogáti B, Szabó E (2012) *J Chromatogr A* 1232:218-230
- [110] ICH Harmonised Tripartite Guideline. Pharmaceutical Quality Systems Q10 (2008) International Conference on Harmonisation of technical requirements for registration of pharmaceuticals for human use
- [111] J.A. Cornell, *Experiments and Designs, Models and the Analysis of Mixture Data*, J. Wiley & Sons, 2nd Ed., New York, USA, 1990.
- [112] B. Bourguignon, D.L. Massart, *J. Chromatogr. A* 586 (1991) 11-20.
- [113] D. 113, J. Nony, R. Phan-Tan-Luu, *NEMROD-W*, LPRAI sarl, Marseille.

- [114] S. Lucangioli, S. Flor, M. Contin, V. Tripodi, *Electrophoresis* 30 (2009) 1899-1905.
- [115] J.M. 115, A.M. González-Paramás, C. Santos-Buelga, M. Battino, *J. Agric. Food Chem.* 58 (2010) 9817–9824.
- [116] H.M. Abu-Tarboush, H.A. Al-Kahtani, M.S. El-Sarrage, *Food Chem.* 46 (1993) 13-17.
- [117] P. Przybylowski, A. Wilczynska, *Food Chem.* 74 (2001) 289-291.
- [118] I. Bratu, C. Georgescu, *J. Cent. Eur. Agric.* 6 (2005) 467-470.
- [119] K.A. Osman, M.A. Al-Doghairi, S. Al-Rehiyani, M.I.D. Helal, *J. Food Agric. Environ.* 5 (2007) 142-146.
- [120] R. Fernández-Torres, J.L. Pérez-Bernal, M.Á. Bello-López, M. Callejón-Mochón, J.C. Jiménez-Sánchez, A. Guiraúm-Pérez, *Talanta* 65 (2005) 686-691.
- [121] O.M. Hernandez, J.M.G. Fraga, A.I. Jimenez, F. Jimenez, J.J. Arias, *Food Chem.* 93 (2005) 449-458.
- [122] P.Pohl, *TRAC-trend. Anal. Chem.*, 28 (2009) 117-128.
- [123] R.D. Snee, *Technometrics*, 19 (1977) 415-428.
- [124] G. Beretta, P. Granata, M. Ferrero, M. Orioli, R. Maffei Facino, *Anal. Chim. Acta* 533 (2005) 185-191.
- [125] J. Bertoneclj, T. Golob, U. Kropf, M. Koros̃ec, *Int. J. Food Sci. Technol.* 2011, 46, 1661–1671.
- [126] W. Brand-Williams, M.E. Cuvelier, C. Berset, *Lebensm.-Wiss. u.-Technol.*, 28 (1995) 25-30.
- [127] S. Chung, S. Kim, H. Park, G. Jhun, J. Chen, *Water Air Soil Pollut.* 3, (2003) 21–40.
- [128] P. Haygarth, C.Jarvis, *Adv. Agron.* 66 (1999) 195-249.

[129] S. Caetano, J. Aires-de-Sousa, M. Daszykowski, Y. Vander Heyden, *Anal. Chim. Acta* 544 (2005) 315-326.

10. Appendice

ESTRATTI DEI LAVORI PUBBLICATI

- S. Furlanetto, S. Orlandini, I. Giannini, **B. Pasquini**, S. Pinzauti, Microemulsion electrokinetic chromatography: An application for the simultaneous determination of suspected fragrance allergens in rinse-off products, *Talanta*, 83 (2010) 72-77.

- Serena Orlandini, Roberto Gotti, Iacopo Giannini, **Benedetta Pasquini**, Sandra Furlanetto, Development of a capillary electrophoresis method for the assay of ramipril and its impurities: an issue of cis-trans isomerization, *J. Chromatography A*, 18 (2011) 2611-2617.

- F. Melani, I. Giannini, S. Furlanetto, **B. Pasquini**, S. Orlandini, S. Pinzauti, Evaluation of the Separation Mechanism of Electrokinetic Chromatography with a Microemulsion and Cyclodextrins using NMR and Molecular Modeling, *Electrophoresis*, 32 (2011) 3062-3069.

- G. Piepel, **B. Pasquini**, S. Cooley, A. Heredia-Langner, S. Orlandini, S. Furlanetto, Mixture-process variable approach to optimize a microemulsion electrokinetic chromatography method for the quality control of a nutraceutical based on coenzyme Q10, *Talanta*, 97 (2012) 73-82.