

Università degli Studi di Cagliari



Dottorato di Ricerca in Scienze Chimiche

XIX ciclo (aa.aa. 2003-2004, 2004-2005, 2005-2006)

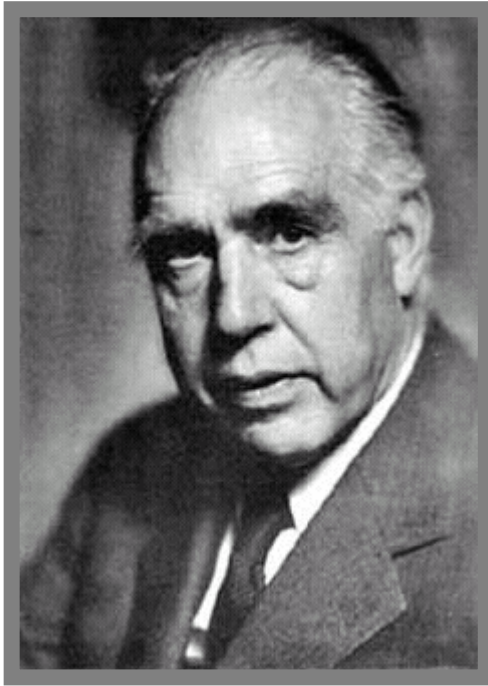
**Studi di metabonomica basata sull’NMR  
in medicina sulla sclerosi multipla negli  
adulti e sull’insufficienza renale  
neonatale.**

Settore Disciplinare Chimica-Fisica (Chim-02)

Autore: Luigi Barberini

Supervisore Prof. Adolfo Lai, Dipartimento di Scienze Chimiche Università di Cagliari.

Coordinatore del Corso di Dottorato Prof. Italo Ferino



“Prediction is very difficult,  
especially about the future.”

“Isolated material particles are  
abstractions, their properties being  
definable and observable only through  
their interactions with other systems.”

Niels Bohr *Atomic Theory and the Description of Nature*, 1934

## Sommario

Introduzione.....	4
1. La Metabonomica.....	8
2. La Sclerosi Multipla: descrizione della patologia.....	13
3. Analisi metabonomica 1H-NMR di liquor cerebrospinale di pazienti con Sclerosi Multipla.....	23
3.1. Preparazione dei campioni.....	23
3.2. Spettri NMR protonici del liquido cerebrospinale.....	23
3.3. Analisi degli spettri NMR.....	25
3.4. Proprietà dei metaboliti.....	34
3.5. La strategia di analisi dei dati NMR protonici del liquido cerebrospinale.....	45
3.6. Risultati.....	55
4. Analisi metabonomica 1H-NMR su urine di neonati “pretermine”.....	70
5. Conclusioni e sviluppi futuri.....	79
Glossario.....	85
Sigle usate nella Tesi.....	89
Appendice 1. Routines per l’ambiente di calcolo Matlab.....	90
Bibliografia.....	91
Indice delle Tabelle.....	94
Indice delle Figure.....	95
Ringraziamenti.....	98

## ***Introduzione.***

Il presente lavoro rientra nell'ambito delle attività avviate presso l'Università di Cagliari sulla "Metabonomica", ovvero lo studio delle alterazioni delle funzionalità metaboliche di sistemi biologici per scopi diagnostici.

Data l'importanza e l'interesse che queste tecniche stanno riscuotendo nell'ambito della ricerca in campo diagnostico medicale è stato sviluppato un progetto di ricerca interdisciplinare sull'analisi metabonomica di patologie di varia natura, progetto che attualmente vede la collaborazione tra il Dipartimento di Scienze Chimiche, il Dipartimento di Scienze Cardiovascolari e Neurologiche, la Radiologia, il Dipartimento di Tossicologia dell'Università di Cagliari, del Centro Studi Sclerosi Multipla della ASL n°8 di Cagliari ed il Dipartimento di Patologia Neonatale e Terapia Intensiva Neonatale.

Il gruppo di ricerca impegnato in questo settore ha attualmente in corso esperimenti di metabonomica mediante spettroscopia di risonanza magnetica nucleare, spettrometria di massa e cromatografia su biofluidi per la caratterizzazione di patologie del sistema nervoso, di patologie neonatali rare e di alcune forme di tumore.

La specialità della tecnica proposta, riconosciuta a livello accademico internazionale dalla enorme quantità di articoli e di riviste specifiche del settore nate e sviluppatesi di recente (Metabolomics, Springer), ha già portato notevoli risultati in oncologia, nello studio delle disfunzioni metaboliche dell'alimentazione e sta dando interessanti risultati nella caratterizzazione delle patologie cardiovascolari mediante spettroscopia NMR su campioni di sangue ed urine.

Inoltre, risultano estremamente interessanti le applicazioni nella cosiddetta "farmacometa-bonomica", cioè nello studio della risposta dei sistemi biologici a

trattamenti terapeutici. Con le tecniche di metabonomica ad esempio si possono monitorare gli effetti di nuovi farmaci su modelli animali con lo scopo primario di valutare quantitativamente l'eventuale tossicità delle nuove molecole.

L'aspetto innovativo delle tecniche di analisi metabonomiche riguarda la possibilità di studiare i sistemi biologici nel loro complesso; in generale, l'alto grado di organizzazione e coordinamento di tali sistemi fa sì che tutti i processi funzionali siano descritti da un elevato numero di variabili correlate tra di loro; il carattere "multivariante" delle funzioni che quindi descrivono il comportamento dei sistemi biologici può essere studiato opportunamente solo attraverso le tecniche matematiche di riconoscimento di modelli e profili, definite in gergo tecnico di "Pattern Recognition". Queste tecniche prevedono in genere l'individuazione del set minimo di variabili dello spazio metabonomico, lo spazio virtuale di definizione dei profili metabolici, univocamente legate al fenomeno in studio, cioè alla patologia in esame, ed alla sua caratterizzazione in termini di modelli descrittivi-predittivi. Si tratta cioè di rilevare dai dati una sorta di "impronta digitale" del metabolismo collegabile in maniera univoca al fenomeno biologico in studio.

Il lavoro presentato in questa tesi raccoglie i risultati ottenuti e le metodiche approntate in tre anni di ricerche svolte per l'individuazione e lo studio dei fenotipi caratterizzanti i pazienti adulti affetti da Sclerosi Multipla ed i pazienti neonati "pretermine" con patologie renali; lo studio è effettuato mediante spettroscopia protonica di risonanza magnetica "in vitro" eseguita su campioni di liquido cerebrospinale per lo studio della Sclerosi Multipla (SM) e sulle urine per gli studi sui neonati pretermine.

Scopo di questo lavoro è stato quello di acquisire una migliore conoscenza dei meccanismi metabolici alla base delle patologie e del loro decorso, e di predisporre strumenti di diagnosi precoce per ottimizzare la terapia da somministrare ai pazienti.

La tesi fondamentale è organizzata in cinque capitoli; il capitolo primo introduce i concetti fondamentali della metabonomica, il secondo capitolo descrive la patologia della sclerosi multipla; il terzo capitolo descrive l'analisi metabonomica basata sull'NMR per la diagnosi di SM. Il quarto capitolo presenta il lavoro sperimentale ed i risultati ottenuti nello studio metabonomico delle urine di neonati "pretermine" con patologie renali. Nel capitolo quinto, infine, sono esposte le conclusioni di questi studi e gli sviluppi futuri che ci si attende dallo sviluppo e dall'applicazione delle metodiche metabonomiche in campo medicale.

Per rendere più fluida l'esposizione degli argomenti nei vari capitoli è stato predisposto un glossario, riportato in coda alla tesi, dove si possono trovare definizioni e spiegazioni della terminologia specifica usata e rimandi bibliografici opportuni.

In appendice si riportano alcune procedure di calcolo utilizzate in questi lavori e messe a punto in ambiente MATLAB (The Mathworks <http://www.mathworks.com/>), l'ambiente di calcolo utilizzato per buona parte dell'analisi dei dati sperimentali. È consuetudine tra i programmatori in Matlab ridistribuire i programmi per l'uso da parte di altri utenti.

Il lavoro presentato è stato svolto presso la Neurologia del Policlinico di Monserrato, diretta dal Prof. F. Marrosu, e presso il Laboratorio NMR del Dip.to di Scienze Chimiche, diretto dal Prof. A. Lai, dove sono state effettuate le misure NMR "in vitro" sui fluidi biologici. I campioni di liquor cerebrospinale sono stati forniti dal Centro Studi Sclerosi Multipla di Cagliari, diretto dalla Prof.ssa M. Marrosu. I campioni di urine sono stati forniti dal Dipartimento di Patologia Neonatale e Terapia Intensiva Neonatale del Prof V. Fanos, in collaborazione con il Dipartimento di Tossicologia dell'Università di Cagliari

che svolge da anni interessanti ricerche su metodi innovativi di diagnostica strumentale. Il prof L. Attori del Dipartimento di Tossicologia, che ha collaborato a questo lavoro, segue le ricerche sulla ipofunzionalità renale dei neonati pretermine ed alcuni lavori sui tumori del polmone.

## **1. La Metabonomica.**

La metabonomica è una disciplina basata sull'uso di metodi matematici e statistici per la soluzione di problemi multivariati relativi a dati riguardanti l'espressione di tutti i metaboliti dei sistemi vitali nell'uomo e negli animali (metabonoma animale); recentemente sono state prodotte anche applicazioni per la caratterizzazione di funzioni metaboliche nelle piante [1]. I dati sono prodotti mediante tecniche di indagine chimico-fisiche come la spettroscopia a risonanza magnetica, la cromatografia e la spettrometria di massa applicate a campioni di biofluidi o di tessuti solidi opportunamente selezionati [2]. Queste metodologie trovano oggi numerose possibilità di impiego nel campo delle scienze mediche dove le variabili che possono essere rilevate su soggetti umani ed animali che presentano una determinata patologia sono numerose; può anche accadere che non tutte queste variabili siano necessariamente utili per descrivere il problema considerato, e quelle di interesse possono dipendere da meccanismi di interazione complessi tra i componenti di diversi percorsi metabolici.

Questo discorso è valido sia per la descrizione di patologie in essere, sia per l'identificazione di stadi pre-patologici (non ancora manifesti con disabilità evidenti del paziente) collegabili a fattori di rischio di insorgenza delle malattie, cioè stati descrivibili da variabili dipendenti da alterazioni di origine genetica o ambientale che concorrono ad aumentare o a diminuire la possibilità di insorgenza di malattie e la cui conoscenza può aiutare a mettere in atto piani di prevenzione.

Come già detto in tutti i casi in cui si considerano sistemi biologici le variabili sono numerose e talvolta anche correlate tra loro; in questo caso l'utilizzo dei metodi chemiometrici può aiutare a fornire una visione globale – olistica - del problema, evidenziando le relazioni tra le variabili e la loro importanza relativa,



e può inoltre mettere in evidenza differenze e similarità tra i campioni (nel nostro caso tra i soggetti affetti da una determinata patologia) in base alla loro distribuzione nello spazio multi-dimensionale descritto dall'insieme delle variabili metabonomiche. Considerare i singoli processi biologici come processi isolati esprime una visione riduzionistica delle funzioni vitali, un'astrazione che permette a volte di semplificare notevolmente il problema in studio ma che porta inevitabilmente a modelli di limitato valore. I processi biologici sono rappresentabili attraverso reti che connettono organi e funzionalità dei sistemi viventi con meccanismi molto complicati. In una ipotetica “piramide della vita” tutte le informazioni di base dei sistemi viventi sono contenute nel codice genetico e lo studio dell'espressione genica permette di rilevare alterazioni del codice da mettere in relazione con le patologie. Nel livello superiore lo studio della proteomica aggiunge un importante tassello nella caratterizzazione degli effetti delle alterazione del codice genetico sulle funzioni dei singoli sistemi biologici isolati. Alla vertice di questa piramide troviamo l'espressione dei metaboliti che rappresenta una sorta di indicatore sul modo di evolvere del processo biologico in studio.

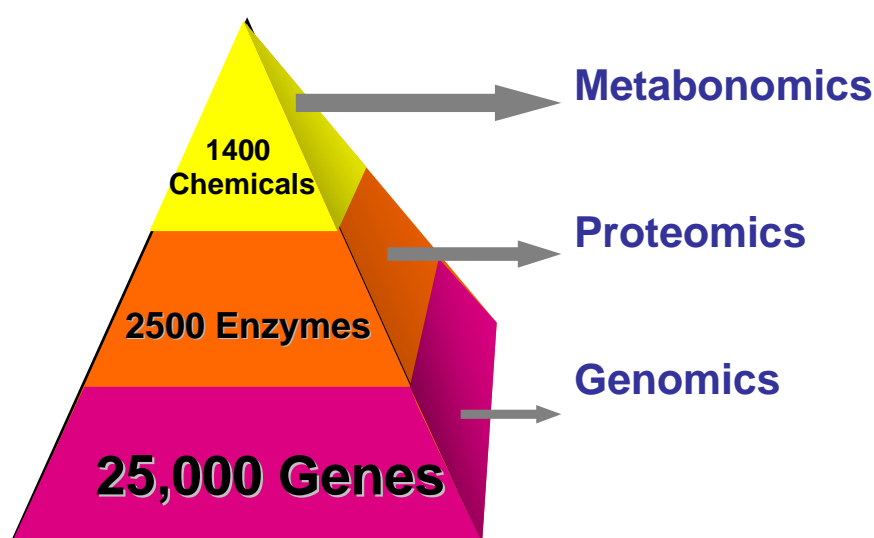
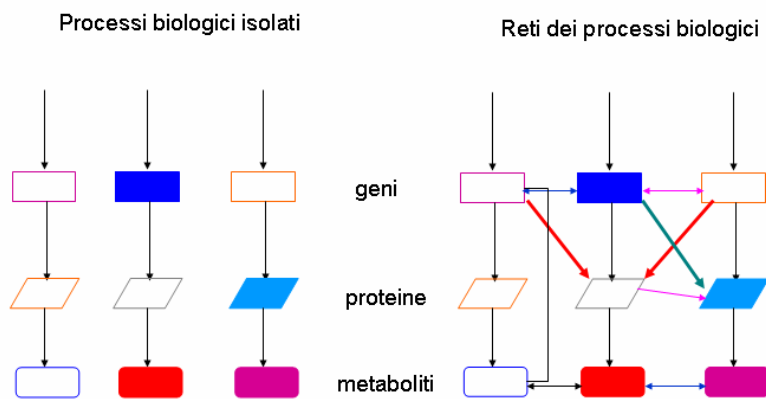


Figura 1. La piramide dei processi biologici

Ma mentre l'espressione genica è un processo di natura interna dei sistemi in studio, le espressioni proteomiche e metaboliche riguardano anche fattori esterni considerati oggi molto importanti in medicina, quali l'alimentazione, i fattori ambientali, etc.

Inoltre le espressioni metaboliche e proteomiche dipendono anche dalle interazioni tra i vari processi biologici e quindi permettono di studiare anche i processi di adattamento e reazione che i sistemi biologici mettono in atto, in collaborazione, in presenza di patologie.



**Figura 2 Complessità dei sistemi biologici.**

L'espressione metabolica si pone a valle dei network dei sistemi biologici e ha interazioni con tutti i livelli superiori.

Dunque, le informazioni relative all'espressione dei geni e delle proteine nei sistemi biologici (definite rispettivamente **genomica** e **proteomica**) non sono in grado di fornire tutte le necessarie informazioni per la conoscenza delle funzioni dei sistemi cellulari integrati, in quanto non considerano lo stato metabolico dinamico dell'intero sistema biologico; tale stato, nella scala delle funzioni biologiche, si pone a valle di tutte le interazioni nei network sistemici ed modulato anche dalle azioni delle alterazioni dei geni e delle proteine. La

metabonomica raccoglie informazioni che completano l'analisi genomica e proteomica poiché permette di investigare anche sulle alterazioni prodotte da fattori esterni che possono concorrere in maniera significativa al quadro clinico [3].

I metodi di analisi chimico-fisici applicati per la caratterizzazione di campioni di fluidi o di tessuti biologici producono matrici di dati di grandi dimensioni. Inoltre, la metabonomica prevede, in genere, lo studio di numerosi campioni per estrarre dai dati tutte e solo le informazioni di interesse per la caratterizzazione della patologia. A causa, quindi, della complessità dei calcoli richiesti è stato possibile applicare le tecniche di metabonomica a problemi di carattere medico solo di recente con lo sviluppo di processori di calcolo sempre più potenti ed algoritmi di analisi sempre più efficienti.

Le interdipendenze delle variabili metabonomiche sono in generale rappresentate principalmente con funzioni lineari perché la linearità permette una serie di semplificazioni enormemente utili in fase di calcolo. Ma in molti casi si è recentemente scoperto che i sistemi multivariati presentano fenomeni di interazione non lineare tra le variabili capaci di spiegare molte delle proprietà di sistema macroscopiche direttamente collegabili alle patologie ed alle loro evoluzioni temporali. Si parla allora di “network metabolici complessi”; in generale nei network metabolici gli elementi di base, detti nodi, rappresentano le molecole che prendono parte ai processi biologici; mentre le connessioni tra i nodi, le interazioni, esprimono le possibili reazioni bio-chimiche. Tali interazioni sono responsabili, nel loro insieme, dell'instaurarsi di precise dinamiche metaboliche con specifiche proprietà di natura complessa e caotica [4, 5].

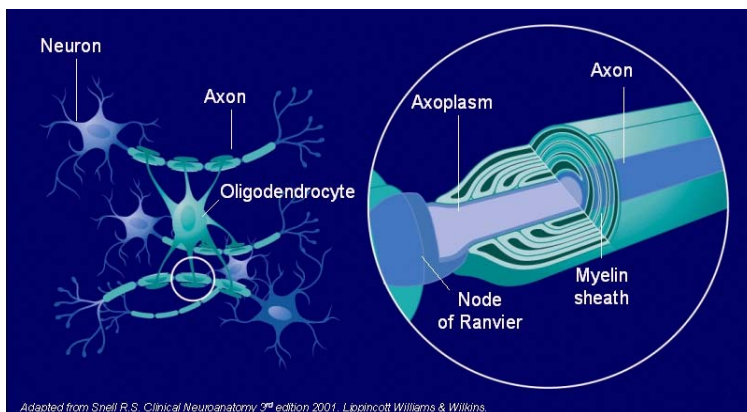
La scoperta dell'esistenza di fenomeni caotici ha portato a considerare in modo nuovo anche il mondo della complessità in biologia: se infatti la concezione riduzionista - che descrive le caratteristiche di un sistema come la somma di quelle dei suoi componenti - perde significato di fronte alle manifestazioni del caos, allora si capisce come i sistemi biologici complessi debbano essere considerati da un punto di vista globale, perché le loro caratteristiche macroscopiche non sono riconducibili alla semplice somma delle proprietà delle loro parti microscopiche. I sistemi biologici sono un esempio affascinante di complessità. Tuttavia i sistemi biologici possono presentare una relativa stabilità per alcuni comportamenti che, in vivo, si mantengono molto lontani dall'equilibrio termico; una situazione che si realizza grazie a processi di cooperazione tra diverse parti del sistema che, pur non essendo a diretto contatto, sono collegate tra loro in modo funzionale, e si trasferiscono l'energia senza dissipazioni. Questo è il motivo per cui è possibile osservare comportamenti caotici in sistemi biologici o in loro parti, fenomeni che si manifestano quando il sistema è sottoposto a particolari stimoli, ma anche in condizioni normali: dinamiche caotiche, per esempio, compaiono nell'attività dei neuroni stimolati con certi tipi di segnale, in quella degli enzimi, in quella cerebrale durante il sonno. Ordine e caos si intrecciano in molte delle manifestazioni della vita.

Queste considerazioni teoriche sono la base dell'approccio metabonomico per la diagnostica moderna in medicina; i modelli che si costruiscono hanno la capacità di raccordare i diversi livelli "omici" (genomica, proteomica, metabonomica) caratterizzanti i biosistemi e rappresentare le reciproche interazioni e dipendenze che generano la complessità di questi sistemi. Solo con questo tipo di approccio è possibile identificare profili specifici di patologie e di malati. Dunque la metabonomica che determina i profili dei processi biochimici

sistemici e la regolazione di funzioni nell'intero sistema biologico complesso può contribuire alla comprensione del perché si instaurano certi stati patologici anziché altri. Importante per l'analisi metabonomica è l'individuazione del biofluido da cui estrarre il profilo metabolico di interesse. Molti sono i fluidi corporei e gli estratti che si possono ricavare per questo scopo: sangue, saliva, urine, liquido cerebrospinale, estratti solidi ridisciolti in soluzioni opportune. È chiaro che oltre alla scelta del biofluido occorre anche considerare quale metodica di indagine strumentale si deve usare per ricavare i profili metabolici; la spettroscopia di risonanza magnetica nucleare, la metodica scelta per lo studio presentato, mostra numerosi vantaggi poiché permette lo studio qualitativo di miscele complesse di metaboliti senza ricorrere a manipolazioni chimiche dei campioni; inoltre è stata scelta anche perché il gruppo di ricerca ha avuto la disponibilità di uso del tomografo del Policlinico Universitario di Monserrato per effettuare la spettroscopia protonica "in vivo" che ci permette di acquisire, mediante le tecniche di Imaging, le popolazioni metaboliche di specifici settori del cervello. Questo permetterà in futuro di svolgere esperimenti con l'integrazione delle due metodiche di spettroscopia, in vivo ed in vitro, per il potenziamento della capacità diagnostica in fase precoce della sclerosi multipla.

## ***2. La Sclerosi Multipla: descrizione della patologia.***

La Sclerosi Multipla (SM) è definita come una malattia infiammatoria degenerativa del Sistema Nervoso Centrale (SNC) caratterizzata dalla formazione di regioni di demielinizzazione (placche) della sostanza bianca, l'insieme delle fibre di prolungamento delle cellule neuronali, chiamate assoni, attraverso cui si propagano gli impulsi elettrici nervosi.

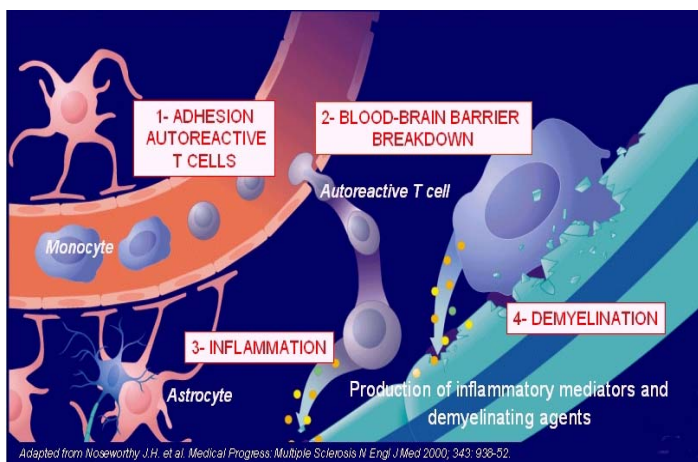


**Figura 3. Il rivestimento mielinico dell'assone, sede dell'attacco della malattia.**

Le placche di demielinizzazione correlano variabilmente sul piano clinico con una disabilità ingravescente per il paziente quantitativamente espressa dall'indice EDSS (Expanded Disability Status Scale). Nonostante il dato di una variabilità individuale renda difficile l'interpretazione della malattia sia in sede diagnostica sia prognostica, possiamo senz'altro definire la SM come una patologia ad andamento cronico-ricidivante.

La sintomatologia d'esordio della malattia è caratterizzata da episodi reversibili classificati come "definiti" o "possibili", in relazione al grado di compatibilità con una diagnosi su base clinica di SM. La reversibilità è da intendersi in senso relativo, poiché è frequente che a distanza di mesi dalla remissione apparente della sintomatologia, questa si ripresenti in forma variamente progressiva, fino a definire il quadro della malattia conclamata. Questo comporta la comparsa, accanto a sintomi di natura sensitiva (che rappresentano quasi la norma nei pazienti più giovani), di disturbi di tipo motorio consistenti nell'aumento del tono muscolare (spasticità), di atassia cerebellare e disturbi di tipo neurovegetativo (incontinenza, impotenza), che finiscono per compromettere gravemente la qualità della vita dei pazienti. I disturbi di tipo psichiatrico frequentemente lamentati dal paziente, soprattutto la depressione, sono da riferire anche alla difficile elaborazione di una malattia vissuta senza speranza

di guarigione. La classificazione attuale distingue, in funzione del decorso, tra forme primariamente progressive, forme recidivanti-remittenti, forme secondariamente progressive e forme “benigne”, in funzione dell’evoluzione temporale della malattia. Secondo quanto le conoscenze acquisite e gli studi sperimentali suggeriscono, la SM riconosce un’eziologia multifattoriale (la cui interpretazione non è peraltro univoca) in cui il peso di un substrato genetico, ancora in via di definizione, si sovrappone a influenze ambientali; sembrerebbe inoltre che gli agenti in grado di scatenare le periodiche poussées possano essere di natura virale. Secondo la teoria patogenetica autoimmune, le placche di demielinizzazione sono il risultato di un’aggressione dei linfociti T nei confronti degli oligodendrociti in risposta a stimoli ambientali agenti sul substrato di un particolare pattern HLA, come confermano studi recenti effettuati proprio su popolazioni sarde [6].



**Figura 4** Le fasi dell'inflammatione.

L’importanza della malattia è facilmente comprensibile alla luce degli studi epidemiologici, che mostrano come in Sardegna si registrino valori di prevalenza attorno ai 145/100.000 abitanti, sensibilmente superiori alla soglia che definisce un’area ad alto rischio (30/100.000). Tali studi hanno contribuito ad affermare nella comunità scientifica l’idea della Sardegna come realtà

epidemiologica a sé stante, mettendo in rilievo sul piano assistenziale il costo umano e sociale che la malattia ha e avrà maggiormente in futuro. Da queste considerazioni sorge la necessità di affinare e rendere tempestive le procedure di diagnosi e monitoraggio evolutivo della malattia, limitando al minimo diagnosi errate e falsi negativi. La sistematizzazione dei criteri diagnostici secondo Poser, nella revisione di McDonald [7] , ha messo ordine nell'approccio talvolta caotico che una malattia così complessa suggeriva, distinguendo criteri clinici e paraclinici. Questi ultimi forniscono un valido supporto laboratoristico e strumentale al ragionamento critico differenziale sui sintomi. L'evidenza di bande oligoclonali nel liquido cerebrospinale, un pattern patologico nei potenziali evocati e un quadro di lesioni in atto rilevate con le tecniche di Imaging di risonanza magnetica (RM) costituiscono oggi reperti altamente significativi nella diagnosi di SM. Gli sviluppi delle conoscenze nel campo dell'imaging hanno imposto la RM come standard nella diagnosi dei casi che si presentano in maniera monosintomatica. Anche il monitoraggio della malattia attraverso la valutazione del carico lesionale (MRI avanzato) e delle popolazioni dei metaboliti cerebrali (MRS in vivo) è migliorato in sensibilità e specificità grazie al miglioramento delle tecniche di RM. Accanto alle tradizionali tecniche RM viene valutata negli ultimi anni l'affidabilità in diagnostica di tecniche non convenzionali. In particolare, le tecniche attualmente più promettenti in questo settore della RM sono la spettroscopia NMR protonica in vivo ed in vitro. Accanto allo sviluppo di nuove tecniche in grado di migliorare l'affidabilità in diagnostica della RM, un altro capitolo di ricerca importante riguarda i processi di estrazione delle informazioni dalle basi di dati disponibili. La digitalizzazione delle indagini medicali ha aperto un settore di ricerca nella determinazione di algoritmi di analisi automatica o semiautomatica che permette in linea di principio di poter esaminare grandi quantità di dati tramite tecniche di analisi statistica multivariata che permettono di estrapolare le connessioni tra le



variabili dei processi biofisiologici che originano le modificazioni strutturali e funzionali in patologie complesse come la SM. Questo lavoro, attraverso la presentazione dei dati inerenti lo studio comparato all’NMR del liquido cerebrospinale in soggetti affetti da SM vs soggetti sani, intende evidenziare le potenzialità di questa analisi quantitativa nell’offrire un supporto alla diagnosi precoce di SM. Negli ultimi anni diversi fattori hanno portato allo sviluppo di un approccio diagnostico più intensivo e fortemente basato sull’uso delle moderne tecniche di indagine biomolecolare, uno sforzo anche economico sorretto dalla Regione Sardegna, dai programmi di Ricerca Sanitaria Nazionale, volto allo sviluppo di tecniche diagnostiche per la diagnosi precoce della SM. Tra tali fattori c’è anzitutto la disponibilità di trattamenti farmacologici che hanno dimostrato di essere potenzialmente efficaci nel modificare la storia di malattia in una fase estremamente precoce, in altre parole quella che è definita attualmente con il termine di “sindrome clinicamente isolata” (il termine anglosassone è Clinically Isolated Syndrome – abbreviata in CIS). Si tratta di un primo disturbo neurologico compatibile con una lesione demielinizzante, vale a dire, la prima manifestazione per chi svilupperà effettivamente una SM. Le persone che hanno una CIS, però, non svilupperanno necessariamente una SM: hanno solo una probabilità maggiore rispetto alla popolazione generale di sviluppare una SM. La metabonomica basata sull’NMR di biofluidi è potenzialmente un approccio in grado di fornire una metodica ad alta sensibilità e specificità nella individuazione di profili metabolici patologici per SM [8, 9, 10]; inoltre permetterebbe di seguire anche l’evoluzione delle funzioni metaboliche in seguito a trattamento farmacologico riportando di fatto parametri biofunzionali indicatori dello stato globale di malattia, patologia e risposta del biosistema alle alterazioni indotte, risposta la cui natura può variare da individuo ad individuo ed esprime un carattere cronico o recidivante.

Il carattere cronico recidivante della patologia SM può manifestarsi in diverse forme anche a distanza di tempo e viene definito attraverso le informazioni fornite da un vasto quadro di indagini strumentali alcune da ripetere nel tempo e su un periodo non inferiore ad un anno. L'evoluzione della malattia può avvenire in diversi modi con diverso impatto sulla qualità di vita del paziente.

Risulta quindi di estremo interesse per il Medico un quadro descrittivo quanto più possibile precoce della patologia che possa prevedere anche l'evoluzione clinica a cui andrà incontro il paziente e mirare così anche il trattamento farmacologico.

Nello studio delle dinamiche che sottendono a questi percorsi evolutivi della malattia risulta importante lavorare con tecniche che permettano di estrarre informazioni sul comportamento di insieme del sistema biologico in studio. Nel caso specifico della SM non esiste, attualmente, un esame specifico che permetta di diagnosticare con totale sicurezza la patologia.

I criteri di MacDonald, lo standard approvato dalla comunità scientifica dei neurologi per dare un quadro diagnostico ad alta compatibilità con la diagnosi di SM, prevedono, come già detto, l'integrazione dei dati provenienti da diverse metodiche di indagine; occorre inoltre che questi esami, soprattutto quelli di imaging, siano ripetuti ogni tre mesi per un periodo di tempo non inferiore ad un anno.

L'analisi metabonomica proposta per questa patologia permetterebbe di sviluppare una visione di funzionalità sistemica a livello del sistema nervoso centrale (SNC) studiando le alterazioni delle popolazioni metaboliche del liquido cerebrospinale (CSF chiamato anche in letteratura Liquor cefalorachidiano o semplicemente liquor), un fluido prodotto all'interno dei ventricoli e che fluisce diffusamente nei distretti cerebrali funzionando sia da sistema di trasporto che da liquido di riempimento e di ammortizzamento del cervello e del midollo spinale.

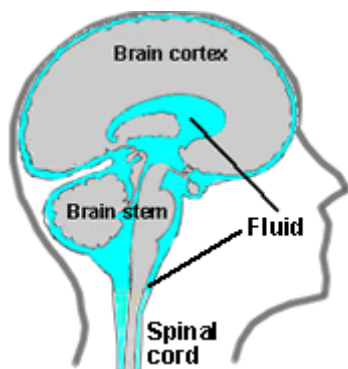


Figura 5. Circuiti di flusso cerebrale del liquor.

Questo rende il CSF un biofluido promettente per la ricerca di biomarkers ad alta specificità per la SM in grado di caratterizzare il fenotipo patologico rispetto ai controlli.

La possibilità di individuare fin dall'esordio il tipo e l'evoluzione della patologia permetterebbe una ottimizzazione del processo di cura con enormi vantaggi per il paziente e per la collettività. La SM e le sue varie forme di esordio e di evoluzione sono ancora oggi difficilmente classificate e la ricerca di biomarkers specifici è uno degli obiettivi più perseguiti dai ricercatori impegnati nel settore. Oggi la SM si controlla, non si cura, mediante farmaci tipo l'Interferone Beta, i cui effetti collaterali nella fase di somministrazione sono spesso pesanti e di difficile sopportazione. In questa fase risulta molto importante l'ottimizzazione del trattamento al fine di evitare inutili sacrifici ai pazienti.

L'ottimizzazione della terapia passa però attraverso una specifica caratterizzazione del fenotipo metabolico a cui appartiene il paziente e dipende dalle specifiche risposte del sistema metabolico ed alle interazioni tra i vari network metabolici.

Il liquido cerebrospinale serve anche a drenare i prodotti del metabolismo cellulare del SNC; viene prelevato mediante puntura lombare (rachicentesi) e può essere analizzato attraverso diverse tecniche di indagine strumentale.

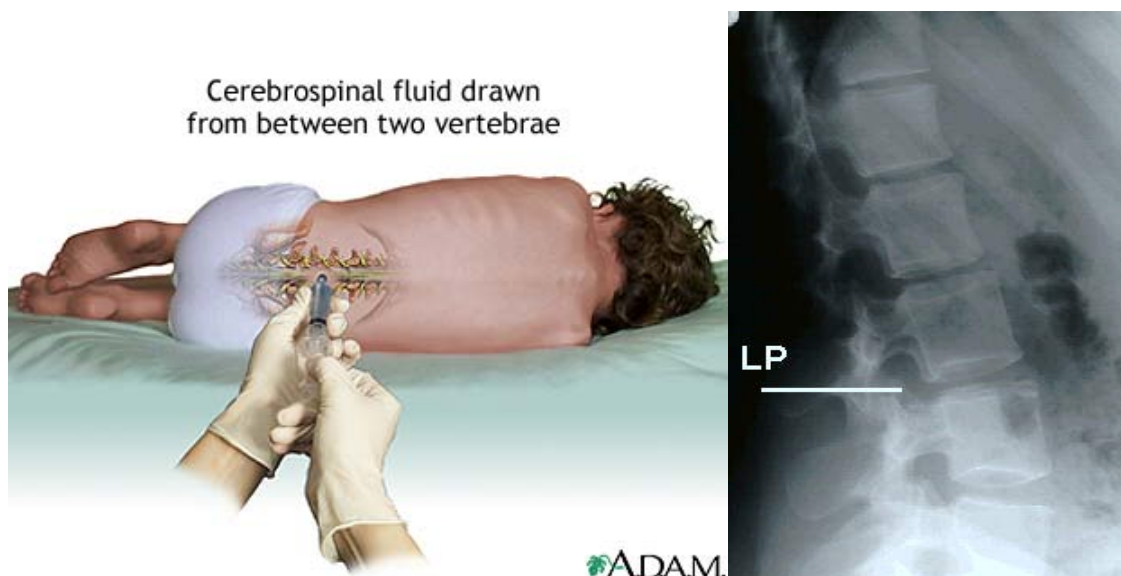


Figura 6. Rachicentesi o prelievo lombare di liquido cerebrospinale.

Questi stati funzionali dipendono dall'espressione genica e proteomica e sono funzione anche di sollecitazioni esterne. Tuttavia, come abbiamo già avuto modo di dire, genomica e proteomica, pur avendo portato notevoli avanzamenti nella comprensione della genesi di alcune patologie, non riescono a fornire informazioni complete e definitive sul comportamento globale dei sistemi biologici; molti dei disordini neurodegenerativi oggi noti non sono ancora del tutto compresi nei loro meccanismi di azione e sviluppo che coinvolgono particolari dinamiche funzionali quali la plasticità neuronale. L'analisi dei meccanismi di plasticità strutturale e funzionale in particolari strutture del cervello, quali l'ippocampo, l'amigdala, la sostanza bianca ad esempio, mediante la determinazione di specifici markers è l'oggetto degli studi moderni per la caratterizzazione di molte patologie neuronali fortemente ingratescenti come le Demenze ed il Parkinson, oltre che la sclerosi multipla; nel grafico che segue si riportano alcuni dei mediatori chimici dei molteplici meccanismi di segnalazione neuronale in regioni encefaliche funzionalmente critiche, meccanismi coinvolti nella regolazione della plasticità sinaptica e delle strategie neuroprotettive.

Del resto i metaboliti a disposizione dopo una analisi NMR del liquor sono molte decine, alcuni con ruolo ben conosciuto, altri di cui si conosce solo il coinvolgimento nei processi patologici ma non si conosce completamente il complesso delle reazioni in cui sono coinvolti. Nel complesso meccanismo di trasduzione del segnale che risiede a valle dei sistemi di trasmissione limbici ad esempio, le MAPKs costituiscono fattori chiave per la comprensione delle caratteristiche funzionali delle cellule nervose, considerato il loro ruolo nella regolazione dei processi mitotici e nella morfogenesi dendritica (Kumar et al., 2005). Tra le MAPKs, i pathways cellulari di ERK, JNK e p38 sono tra i più finemente caratterizzati e coinvolti nella risposta a determinate condizioni fisiopatologiche quali stati di stress, ipossia o preconditionamento ischemico (Zhu et al., 2005). Nelle cellule neuronali, lo stress coinvolge anche fattori cellulari quali le heat shock proteins (Feder and Hofmann, 1999) che attivano specifici pathways cellulari che inducono l'attivazione di determinati fattori trascrizionali (Calabrese et al., 2002).

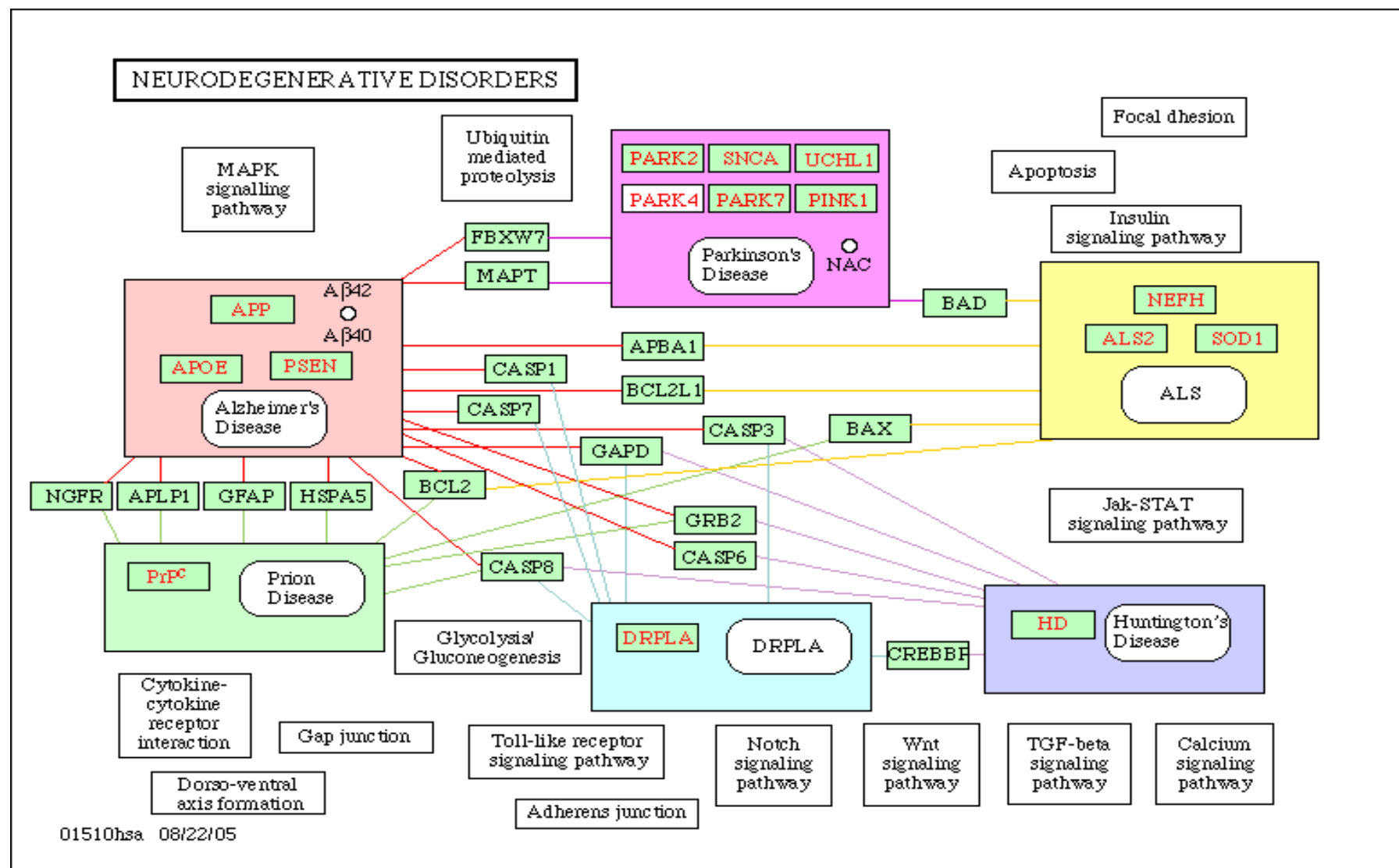


Figura 7. Disordini neurometaboliti e percorsi delle alterazioni geniche e di complessi proteici che inducono alterazioni funzionali e generano i disordini neuronali

### **3. Analisi metabonomica 1H-NMR di liquor cerebrospinale di pazienti con Sclerosi Multipla.**

L'analisi 1H-NMR del liquor permette di individuare e quantificare contemporaneamente decine di metaboliti e permette quindi, in linea di principio, di studiare tutti i processi biologici che sottintendono allo stato funzionale delle diverse aree cerebrali.

#### **3.1. Preparazione dei campioni.**

L'esame 1H-NMR dei campioni di liquor conservati a  $-20^{\circ}$  è stato effettuato previa liofilizzati e solubilizzazione in  $D_2O$ .

Il protocollo adottato prevede il prelievo di 1800  $\mu$ l di CSF che dopo essere stato scongelato viene centrifugato per 10 minuti a 3000 rpm per separare eventuali residui solidi prelevando il surnatante allo scopo di avere un campione più omogeneo che rende più facile ovviare ad effetti di disomogeneizzazione sul campione del campo magnetico statico. Il surnatante prelevato viene ricongelato e posto nel liofilizzatore; il liofilizzato viene poi solubilizzato in 600  $\mu$ l di  $D_2O$ . Il pH dei campioni è stato corretto al valore di  $(7,0 \pm 0,5)$  mediante opportuna addizione di NaOD o DCl.

#### **3.2. Spettri NMR protonici del liquido cerebrospinale.**

Gli spettri NMR protonici sono stati registrati con uno spettrometro NMR Varian 400 Unity INOVA che opera alla frequenza di 400MHz.

Le sequenze di eccitazione del campione sono state predisposte in accordo con gli standard indicati in letteratura per l'analisi dei campioni di CSF all'NMR [11]; a causa della degradazione di alcuni metaboliti, le sequenze di acquisizione non possono essere troppo lunghe e quindi anche l'accumulo di segnale non può avvenire con un numero troppo elevato di transienti. La soppressione del picco dell'acqua è stata ottenuta mediante un impulso di presaturazione la cui durata è stata impostata a 2 s con una potenza di impulso di 6 mW.

L'acquisizione è stata realizzata su una finestra spettrale di 6000 Hz; un delay di rilassamento tra i due transienti di 67 s ci pone in condizione di rilassamento completo del campione dato che il tempo di rilassamento più lungo misurato nei nostri campioni di CSF è quello del lattato di 7,3 s.

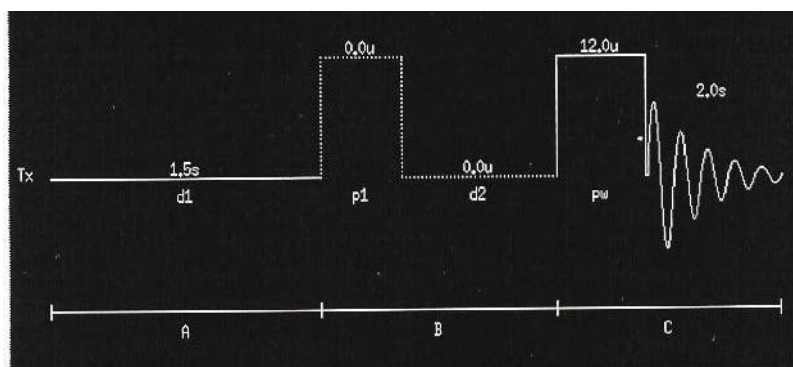


Figura 8. Sequenza con presaturazione del segnale dell'acqua ed acquisizione del FID.

La sequenza è completata dall'impostazione di un tempo di acquisizione  $t_a=1,2$  s, un guadagno pari 20 ed un delay  $d_1= 67$ s. Per i campioni liofilizzati il numero di transienti stimato per l'acquisizione di spettri con un buon rapporto segnale/rumore (SNR) è pari a 64.

Gli spettri acquisiti sono stati processati off-line su PC notebook ASUS, con processore Pentium 4 a 3.2 GHz, equipaggiato con 1 GB di memoria RAM DDR a 400 MHz. La stessa macchina ha eseguito tutti i calcoli e le analisi dei dati per la costruzione dei modelli matematici caratterizzanti le classi dei campioni.

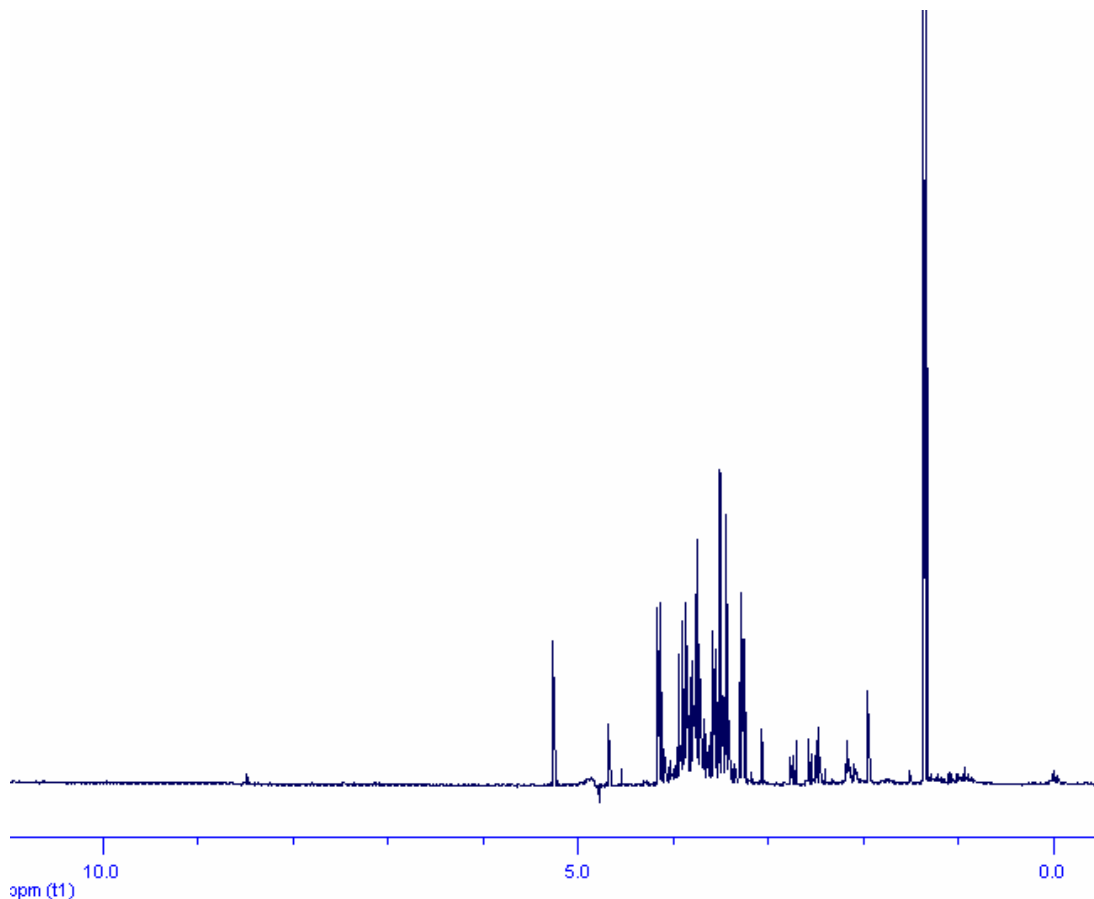
Su ciascun FID acquisito è stata eseguita un "Trasformata di Fourier" con apodizzazione "Lorentziana-Gaussiana" per aumentare la definizione delle risonanze senza andare a discapito del rapporto segnale-rumore.

Su ogni spettro è stato poi eseguita la correzione di fase e la correzione della linea di base.



### 3.3. Analisi degli spettri NMR.

Nella Figura 9. Spettro NMR del H-1 di liquido cerebrospinale. è mostrato lo spettro di un campione (patologico) di liquor. La quantità di informazioni ottenibili dai metaboliti del liquido cerebrospinale è ben espressa dalla complessità degli spettri protonici NMR che da esso si ottengono, dalle numerose risonanze che esprimono i diversi contributi dei metaboliti neuronali.



**Figura 9. Spettro NMR del H-1 di liquido cerebrospinale.**

Con la spettroscopia NMR protonica del liquor e con la spettroscopia “in vivo” si individuano e quantificano, in genere, per le patologie neurologiche, alcuni tra i seguenti metaboliti: l’acido idrossi-isobutirrico, n-acetilaspargato (NAA), colina (Cho), creatina e creatinina (Cr), la glutammina (Gln), la glicina (Gly), il mio-Inositolo (m-I) e l’acido lattico (LA), la tirosina, il citrato, il piruvate, l’alanina (Ala). Molti dei metaboliti presenti nel liquor hanno multipletti di risonanza localizzati in diverse regioni dello spettro NMR.

Nelle figure 10, 11, 12 e 13 si riportano gli spettri NMR nelle regioni [0-1.5]ppm, [1,75-2,75]ppm, [3-4,75]ppm e [5,5-8,75]ppm, mentre nella tabella Tabella 1. Assegnazione delle risonanze 1H- NMR del CSF. sono riportate le assegnazioni delle risonanze principali. L'usuale analisi degli spettri NMR passa attraverso l'assegnazione delle risonanze presenti nello spettro ai metaboliti noti ed alla loro quantificazione mediante la misura dell'area sottesa dalle risonanze, che esprime la concentrazione molecolare del metabolita (dopo l'opportuna normalizzazione al numero di protoni contenuto nel gruppo o nei gruppi metilici, metilenici o metinici delle molecole stesse e dopo aver normalizzato anche rispetto all'area del riferimento usato o all'area totale delle risonanze nello spettro).

Si costruisce così in genere un set di metaboliti presenti nello spettro di cui è nota anche la concentrazione. Il set di metaboliti e le loro concentrazioni rappresentano "l'informazione" estratta dallo spettro NMR e relativa al paziente ed al suo stato clinico. Quindi chemical shift di questi metaboliti e relative concentrazioni rappresentano i parametri "descrittivi" dello spettro e quindi dello stato clinico del soggetto analizzato. Questa procedura viene definita in letteratura ASA, Assigned Signals Analysis.

Il problema della estrazione degli opportuni descrittori di spettro è un punto centrale di grande importanza; in realtà in metabonomica si usano procedure molto diverse che permettono più facilmente di automatizzare l'analisi. Tuttavia nella fase iniziale del progetto si è provveduto ad effettuare l'analisi classica ASA per avere un termine di confronto delle metodologie automatiche di trattamento ed analisi dello spettro sviluppate.

Nell'analisi degli spettri di liquor forniti per lo studio della SM sono state individuate ed assegnate le risonanze di una ventina di metaboliti.

Nelle varie classi cliniche degli spettri analizzati le concentrazioni sono risultate alquanto variabili come era prevedibile, anche in relazione al problema del degradamento di alcuni metaboliti relativi al problema della conservazione ottimale dei campioni di liquor.

È stata rilevata in alcuni casi la presenza di un picco a circa 0 ppm, una risonanza in genere abbastanza larga di cui non è stato possibile determinare l'origine.

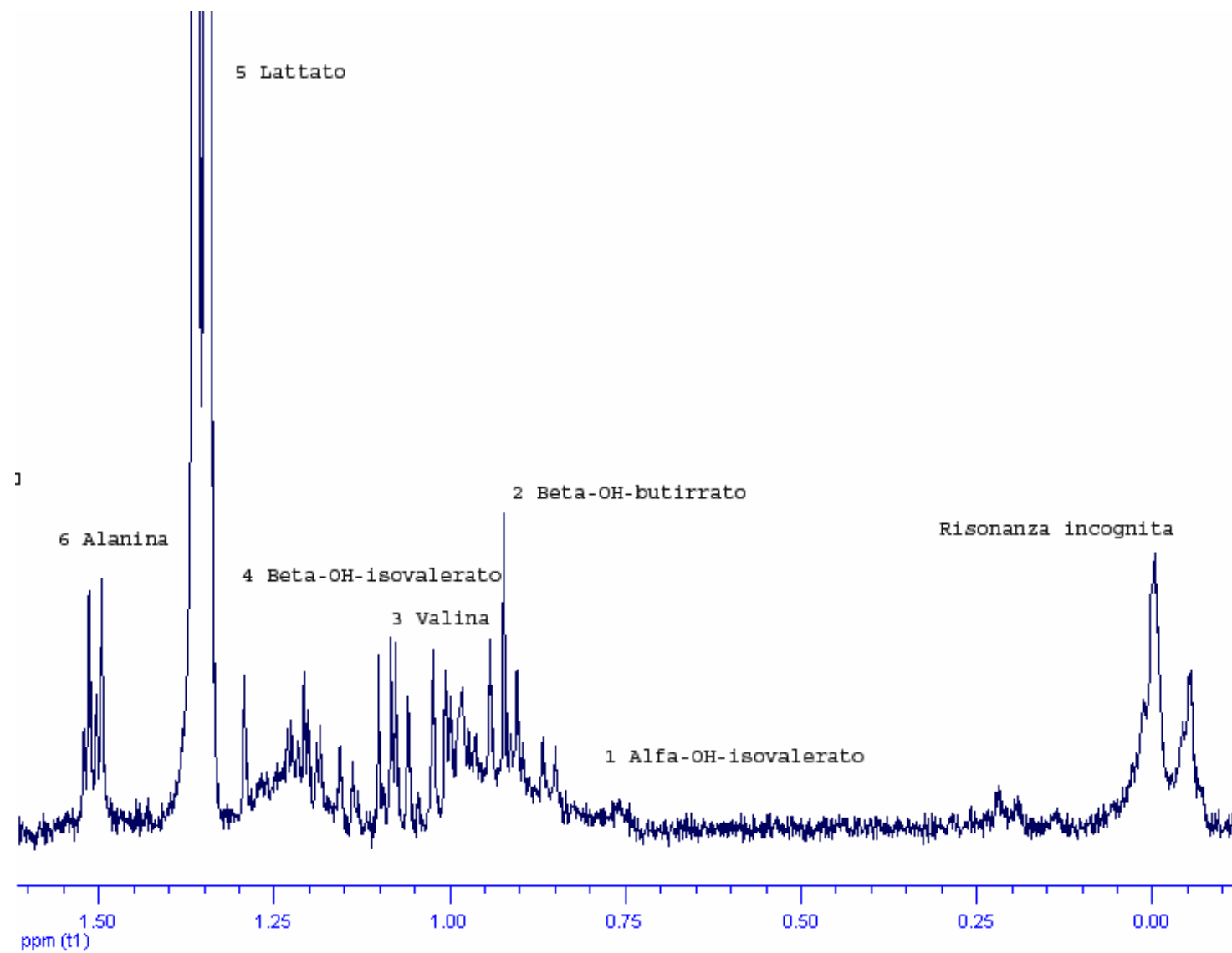


Figura 10. Spettro  $^1\text{H-NMR}$  del CSF nella regione tra 0 e 1,5 ppm.

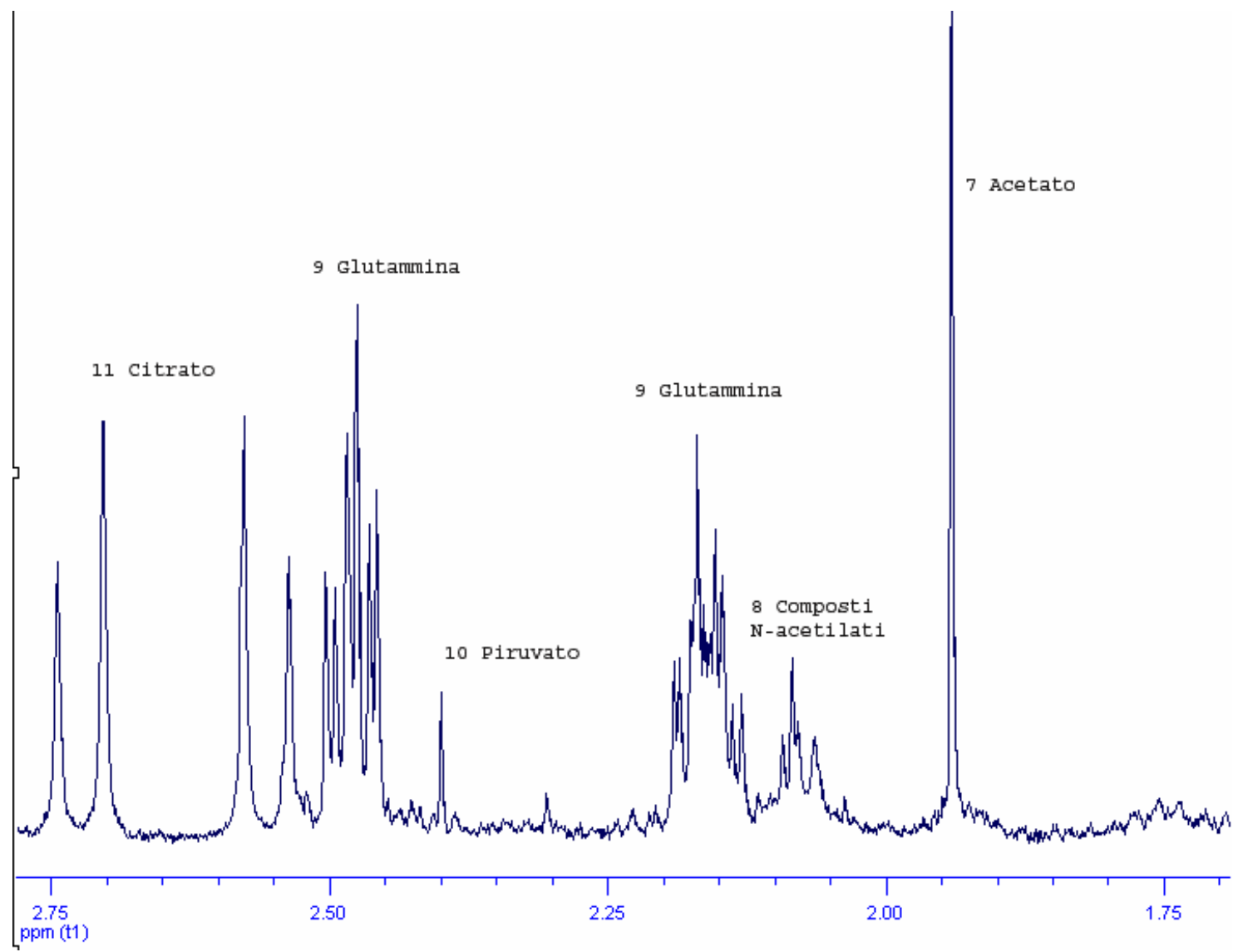


Figura 11. Spettro 1H-NMR del CSF nella regione tra 1,75 e 3,35 ppm.

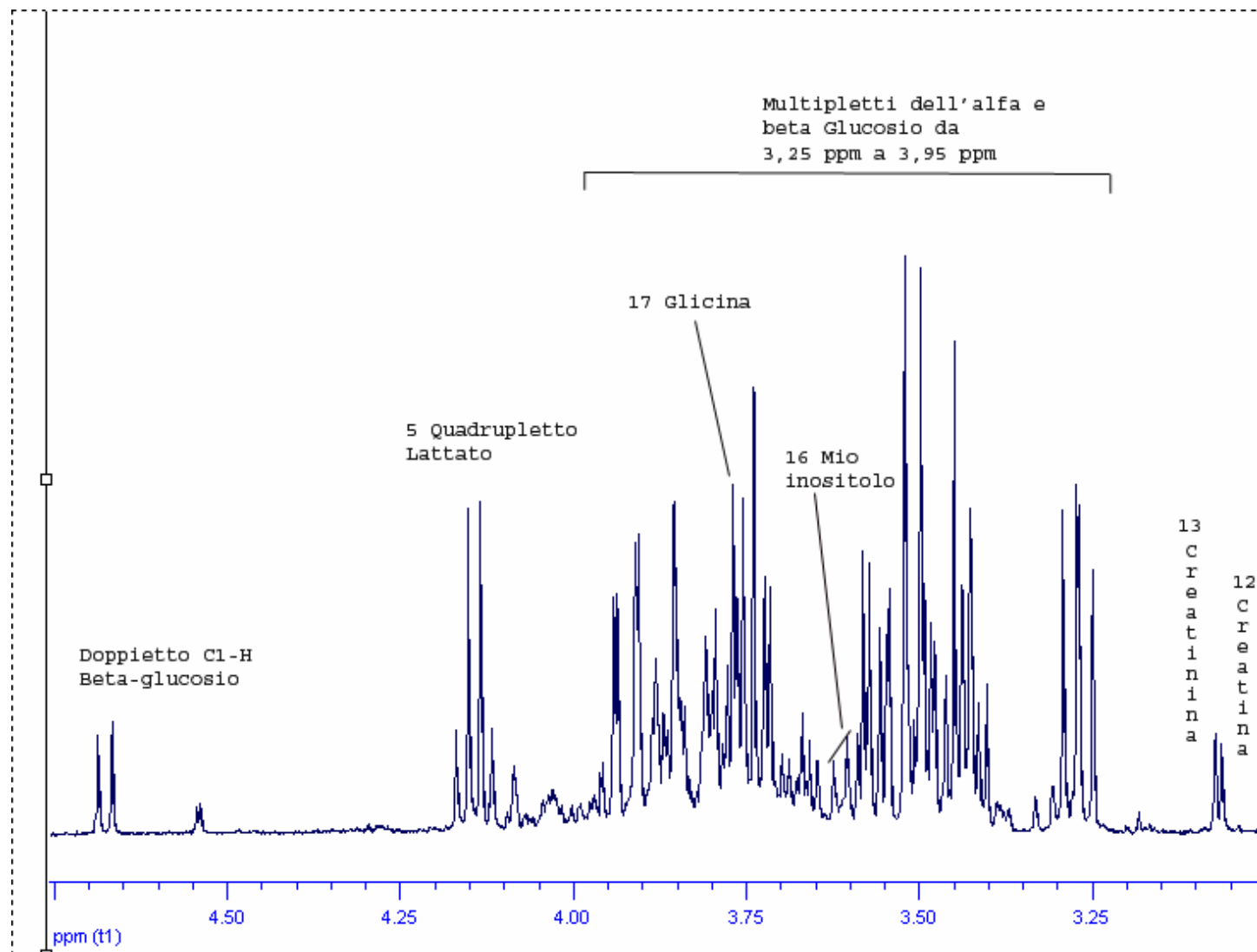
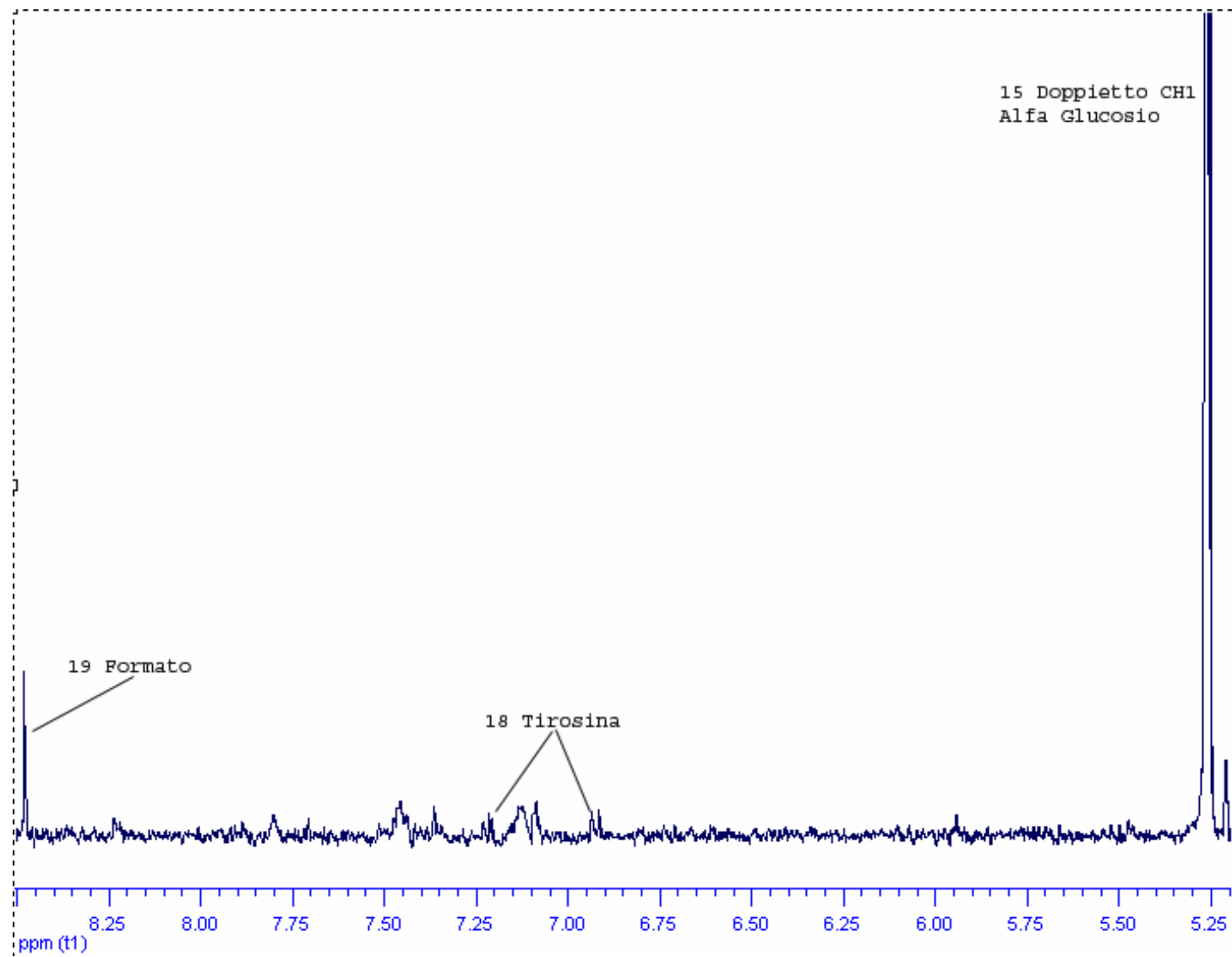


Figura 12. Spettro 1H-NMR del CSF nella regione tra 3 e 4,75 ppm.

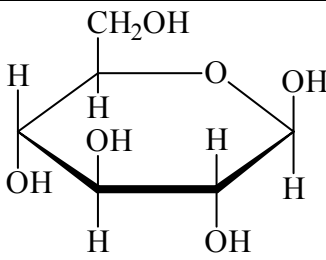
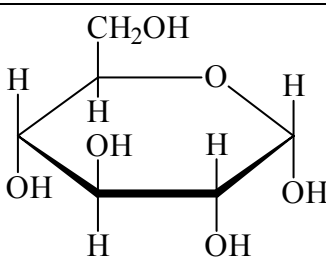
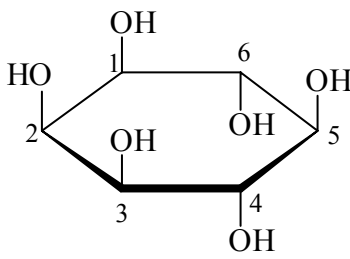


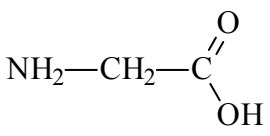
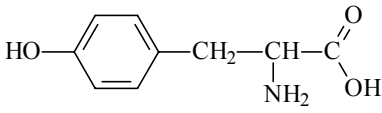
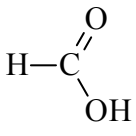
**Figura 13. Spettro 1H-NMR del CSF nella regione tra 5,2 ed 8,75 ppm.**

**Tabella 1. Assegnazione delle risonanza 1H- NMR del CSF.**

	<b>Metabolita</b>	<b>Struttura</b>	<b>Chemical Shift [ppm] e molteplicità delle risonanze.</b>
1	$\alpha$ -OH-isovalerato	$\begin{array}{c} \text{CH}_3 \\ \diagdown \\ \text{CH} \\ \diagup \\ \text{CH}_3 \end{array} - \text{CH} - \underset{\text{OH}}{\text{C}} - \overset{\text{O}}{\parallel} \text{C} - \text{OH}$	0.89 (d, CH <sub>3</sub> ), 0.98 (d, CH <sub>3</sub> ), 4.13 (d, CH-OH)
2	$\beta$ -OH-butirato	$\text{CH}_3 - \underset{\text{OH}}{\text{CH}} - \text{CH}_2 - \overset{\text{O}}{\parallel} \text{C} - \text{OH}$	1.23 (d, CH <sub>3</sub> ), 2.53 (CH <sub>2</sub> )
3	Valina	$\begin{array}{c} \text{CH}_3 \\ \diagdown \\ \text{CH} \\ \diagup \\ \text{CH}_3 \end{array} - \underset{\text{NH}_2}{\text{CH}} - \overset{\text{O}}{\parallel} \text{C} - \text{OH}$	1.00 (d, CH <sub>3</sub> ), 1.04 (d, CH <sub>3</sub> )
4	$\beta$ -OH-isovalerato	$\begin{array}{c} \text{CH}_3 \\ \diagdown \\ \text{C} \\ \diagup \\ \text{CH}_3 \\   \\ \text{OH} \end{array} - \text{CH}_2 - \overset{\text{O}}{\parallel} \text{C} - \text{OH}$	1.33 (s, CH <sub>3</sub> ), 2.55 (s, CH <sub>2</sub> )
5	Lattato	$\text{CH}_3 - \underset{\text{OH}}{\text{CH}} - \overset{\text{O}}{\parallel} \text{C} - \text{OH}$	1.41 (d, CH <sub>3</sub> ), 4.36 (q, CH)
6	Alanina	$\text{CH}_3 - \underset{\text{NH}_2}{\text{CH}} - \overset{\text{O}}{\parallel} \text{C} - \text{OH}$	1.51 (d, CH <sub>3</sub> )
7	Acetato	$\text{CH}_3 - \overset{\text{O}}{\parallel} \text{C} - \text{OH}$	2.08 (s, CH <sub>3</sub> )
8	Composti N-acetilati	$\text{R} - \text{NH} - \overset{\text{O}}{\parallel} \text{C} - \text{CH}_3$	1.94-2.05 (s, CH <sub>3</sub> )
9	Glutammina	$\begin{array}{c} \text{O} \\ \parallel \\ \text{H}_2\text{N} - \text{C} \end{array} - \text{CH}_2 - \text{CH}_2 - \underset{\text{NH}_2}{\text{CH}} - \overset{\text{O}}{\parallel} \text{C} - \text{OH}$	2.16 (m, CH <sub>2</sub> ), 2.47 (m, CH <sub>2</sub> ), 3.91 (t, CH)



10	Piruvato	$\begin{array}{c} \text{O} \\ \parallel \\ \text{CH}_3-\text{C}-\text{C} \\ \parallel \quad \diagdown \\ \text{O} \quad \text{OH} \end{array}$	2.37 (s, CH <sub>3</sub> )
11	Citrato	$\begin{array}{c} \text{O}=\text{C}-\text{OH} \\   \\ \text{H}-\text{C}-\text{H} \\   \\ \text{HO}-\text{C}-\text{COOH} \\   \\ \text{H}-\text{C}-\text{H} \\   \\ \text{O}=\text{C}-\text{OH} \end{array}$	2.91 (sistema AB, CH <sub>2</sub> )
12	Creatina	$\begin{array}{c} \text{NH} \\ \parallel \\ \text{H}_2\text{N}-\text{C}-\text{N}-\text{CH}_2-\text{C} \\ \quad \quad   \quad \quad \quad \parallel \\ \quad \quad \text{CH}_3 \quad \quad \quad \text{OH} \end{array}$	3.05 (s, CH <sub>3</sub> ), 4.06 (s, CH <sub>2</sub> )
13	Creatinina	$\begin{array}{c} \text{NH} \\ \parallel \\ \text{HN}-\text{C}-\text{N}-\text{CH}_2-\text{C} \\ \quad \quad   \quad \quad \quad \parallel \\ \quad \quad \text{CH}_3 \quad \quad \quad \text{O} \end{array}$	3.1 (s, CH <sub>3</sub> ), 4.29 (s, CH <sub>2</sub> )
14	β-Glucoso		3.25-3.95 (vari), 4.64 (d, C <sub>1</sub> -H)
15	α-Glucosio		3.25-3.95 (vari), 5.22 (d, C <sub>1</sub> -H)
16	Mio-Inositolo		3.65 (t, C <sub>5</sub> -H), 4.05 (t, C <sub>2</sub> -H)

17	Glicina		3.77 (s, CH <sub>2</sub> )
18	Tirosina		6.89 (d, H aromatici), 7.19 (d, H aromatici)
19	Formiato		8.25 (s, incerto, volatile)
Legenda: s=singoletti, d=doppio, t=triplo, q=quadruplo, dd doppietto di doppietti, Sistemi AB=interazioni di ordine superiore			

Le aree sono state normalizzate rispetto all'area totale delle risonanze spettrali (quantificazione relativa). Tutte le risonanze assegnate sono state confrontate con i valori della letteratura [10, 11].

### 3.4. Proprietà dei metaboliti.

L'origine ed il ruolo dei vari metaboliti del liquor nei cicli metabolici neuronali è ancora oggetto di studio da parte dei ricercatori soprattutto per la loro correlazione con l'instaurarsi e l'evolversi nel tempo delle patologie neurologiche. Molte informazioni derivano oggi dalla possibilità di effettuare gli esami spettroscopici in vivo direttamente nei tomografi NMR per Imaging.

La totale non invasività dell'esame MRI permette di eseguire facilmente e ripetutamente l'esame spettroscopico senza problemi tecnico-etici e permette di acquisire caratterizzazioni metabolico-funzionale di aree interne del cervello. Chiaramente i bassi valori di campo statico permettono una risoluzione ben inferiore a quelle tipicamente ottenibili negli spettri "in vitro"; le informazioni integrate in vivo in vitro permettono di capire meglio il ruolo dei metaboliti neuronali.

Molti sono i lavori riportati in letteratura che correlano il comportamento dei singoli metaboliti con le varie patologie neurologiche.

L'NAA, ad esempio, è il metabolita più frequentemente e facilmente studiato in 1H-MRS sia in vivo che in vitro per la diagnosi di patologie neuronali. Si

caratterizza con singoletto a 2,02 ppm (regione dei composti N-acetilati in Figura 10). Si ritiene che sia prodotto solo nei neuroni (a livello mitocondriale) e la sua presenza nel tessuto cerebrale è utilizzata come marcatore dello stato delle cellule neuronali in diverse patologie (come la malattia di Alzheimer o l'Epilessia). La sua diminuzione rispetto a soggetti di controllo è da intendere, quindi, come indice di sofferenza o di perdita neuronale. In caso di necrosi cellulare si verifica anche, in contemporanea, la comparsa del picco dei lipidi, generalmente non presente nell'encefalo normale.

Il picco di risonanza della colina è pure inteso come un marker del metabolismo di membrana. Il segnale dei 3 gruppi N-CH<sub>3</sub> cade a 3,19 ppm, mentre i segnali dei due gruppi metilenici CH<sub>2</sub> cadono a 3,65 ed a 4,08 ppm.

Questa ultima serie di segnali risulta in genere di difficile individuazione negli spettri a 400 MHz, anche per la sovrapposizione e vicinanza dei multipletti del glucosio.

Creatina e creatinina costituiscono, in vivo, un unico picco a 3,0 ppm, mentre nello spettro a 400 MHz si riesce a distinguere bene i due singoletti a 3,05 ppm della creatina ed a 3.1 ppm della creatinina dovuti ai protoni dei gruppi metilici. Più difficile invece è individuare i segnali dei protoni metilenici delle due molecole che cadono in zone dove ci sono vicini o addirittura sovrapposti altre risonanze.

Si ritiene che queste creatina e creatinina abbiano principalmente il ruolo di indicatori della riserva energetica cellulare e data la loro stabilità il “picco somma” creatina-creatinina viene utilizzato, in vivo” come normalizzazione delle risonanze di NAA e Cho per costruire i rapporti NAA/CR e Cho/Cr; questi rapporti di concentrazione danno due importanti indici di valutazione delle alterazioni metaboliche neuronali patologiche.

Il lattato è invece il prodotto finale della glicolisi anaerobia, e fornisce così ulteriori informazioni sul metabolismo bioenergetico. È costantemente presente nel liquor cefalorachidiano. Il lattato penetra la membrana tramite le proteine di

trasporto monossilate MCT e viene trasformato in piruvato tramite l'isoenzima latticodeidrogenasi LDH1; poi segue la produzione di energia, di glutammato, di aspartato o di acido gammaminobutirrico GABA. Il ciclo è capillare, parte a livello astrocitario, dove sono coinvolti metaboliti quali glucosio, piruvate e lattato, e prosegue a livello neuronale dove sono coinvolti altri metaboliti quali appunto il lattato, il piruvato, il glutammato, per la produzione di energia; inoltre si collega al ciclo di produzione del glucosio cerebrale (gluconeogenesi) (Figura 14).

Anche il glucosio è un indicatore delle riserve energetiche ed interviene nel metabolismo astrocitario. È molto interessante la sua correlazione con il lattato nel metabolismo energetico neuronale. Infatti, per molto tempo, associando il ruolo di mediatore energetico al solo glucosio, la presenza di lattato nel sistema nervoso è stata scambiata come indice di sofferenza del medesimo che avrebbe provocato danni permanenti (come realmente avviene nelle ischemie); ma il lattato ha anche un ruolo positivo che svolge insieme al Glucosio. Il Lattato è una molecola di scambio tra le cellule della glia ed i neuroni e la sua presenza in un cervello sano è in realtà il frutto di un corretto funzionamento del ciclo energetico cellulare dei neuroni. Risultati ottenuti con gli esperimenti "in vitro" e con mezzi quali la tomografia ad emissione positronica TEP e la risonanza magnetica funzionale fMRI hanno confermato come tutte queste evidenze siano necessarie per una futura completa comprensione dei meccanismi del metabolismo cerebrale. Cade oggi una credenza che voleva il metabolismo dei neuroni connesso direttamente al solo glucosio; oggi appare più corretto affermare che i neuroni attivi ricevono il lattato dagli astrociti e che questo lattato è il principale elemento per la produzione energetica dei neuroni stessi. Chiaramente anche questo non è assolutamente un "principio base", ma solo un'intuizione, importante però perché eleva le cellule della glia (gli astrociti) ad elemento fondamentale nella trasmissione dell'impulso nervoso in quanto mediatore del metabolismo energetico. Da qui si capisce quanto siano importanti

i marker sullo stato delle cellule della glia. Uno di questi marker sembra essere il Mioinositolo.

Il mio-inositolo è conosciuto come regolatore dell'osmolarità intracellulare ma il suo significato spettroscopico è sostanzialmente quello di un efficace marker dello stato della glia. È caratterizzato da un tripletto di risonanze a 3,27 ppm e da un tripletto a 4,05 ppm.

La glutammina si evidenzia nello spettro con due multipletti distinti dovuti ai protoni metilenici a 2,1-2,5 ppm e ad un tripletto dovuto al CH a 3,9 ppm. È importanti per la sintesi delle basi puriniche e pirimidiniche ed è anch'essa collegati al metabolismo astrocitario del lattato; le aree delle risonanze della glutammina aumentano in caso di particolari patologie come l'encefalopatia epatica, sia acuta che cronica, specialmente secondaria a cirrosi. In questa malattia, su cui il nostro gruppo lavora con un protocollo di spettroscopia NMR "in vivo" [12], si rilevano alterazioni funzionali sia a livello epatico sia a livello cerebrale, alterazioni dovute ad accumuli di metalli pesanti.

I dati metabolici si intrecciano in maniera molto complicata. Il tutto si ricollega all'alto grado di organizzazione dei sistemi biologici. Abbiamo visto la correlazione tra lattato e glucosio. Consideriamo allora un percorso diverso e partiamo dal ciclo metabolico della Glicolosi/Gluconeogenesi.

Tra i vari mediatori del segnale neuronale coinvolti da questi processi si trovano anche l'acetilcolina e l'etanolo.

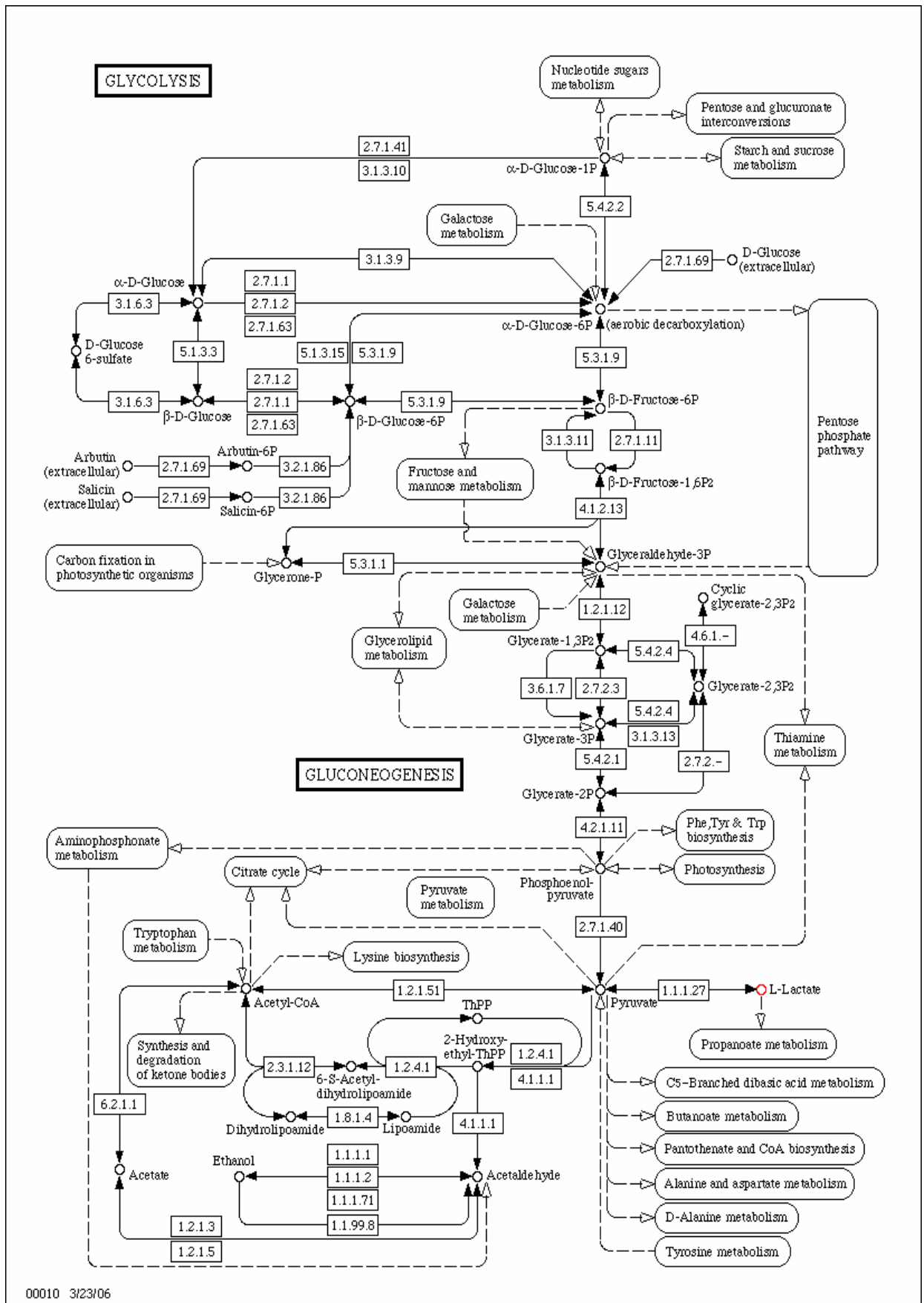


Figura 14. Ciclo glicolisi/gluconeogenesi.

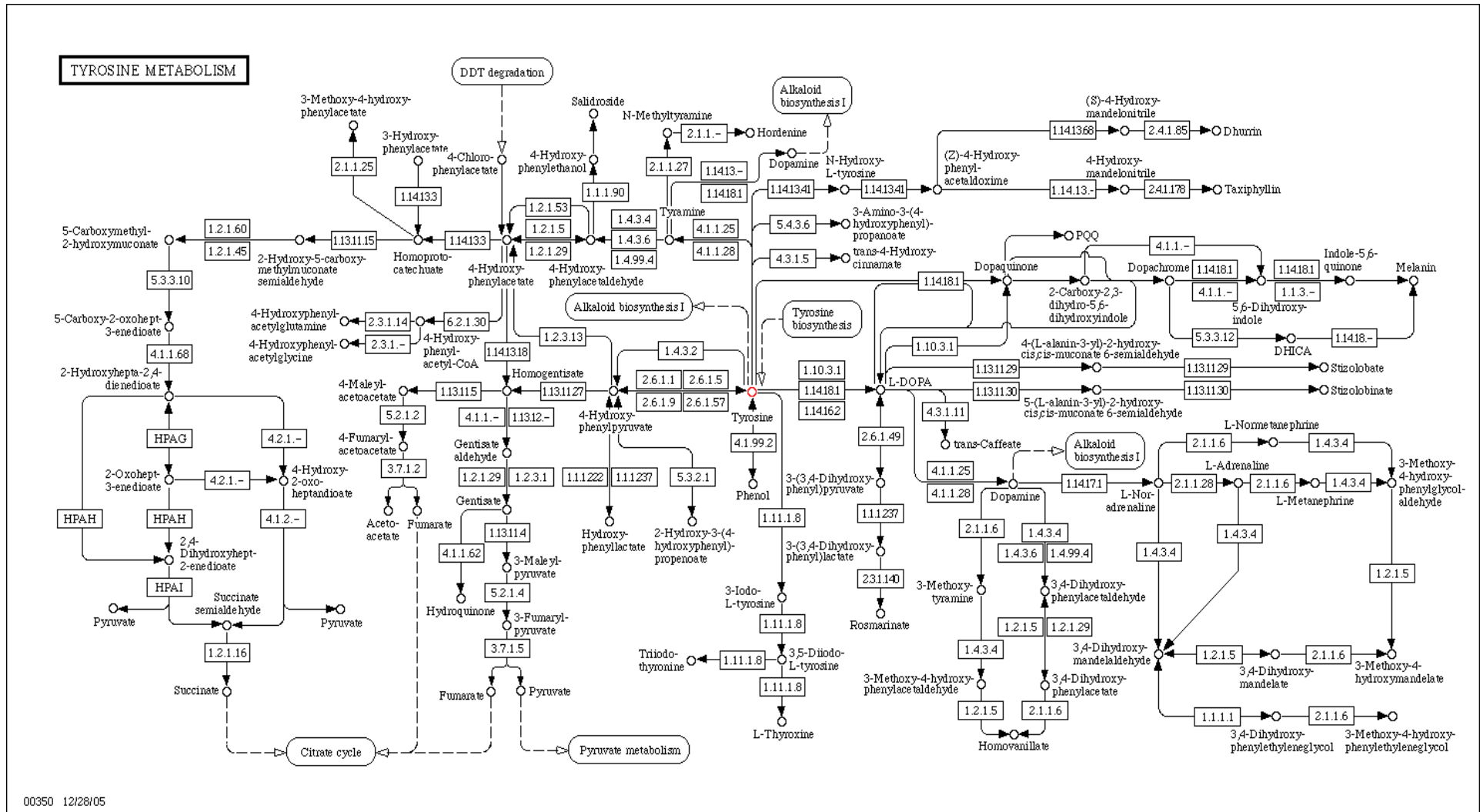
I nodi di questo importantissimo ciclo metabolico connettono anche altri importanti neurometaboliti già visti come la tirosina, il citrato, l'alanina, il piruvato ma anche la lisina, l'acetato ed il lattato.

Il citrato ed il piruvato giocano un ruolo importantissimo nel ciclo di Krebs derivante dal metabolismo aminoacidico, mentre il lattato, come si è già visto, è un marker molto importante nel metabolismo anaerobio.

Analizzando poi ad esempio il ciclo metabolico della tirosina, riportato in Figura 15, si ritrova la correlazione di questo metabolita con il piruvato ed il citrato.

La tirosina è un importante precursore di vari composti tra cui la tiroxina (un ormone tiroideo) e le catecolammine (dopamina, noradrenalina e adrenalina).

Dopamina e noradrenalina sono neurotrasmettitori coinvolti in molte gravi patologie del SNC quali l'epilessia e le malattie degenerative (Parkinson e le Demenze).



**Figura 15. Ciclo metabolico della Tirosina e metaboliti collegati (v. citrato e piruvato).**



Anche nel metabolismo della tirosina si riscontrano connessioni a gruppi metabolici quali quelli del piruvato e del citrato e quelli appunto della dopamina e noradrenalina. Conoscendo quindi la concentrazione locale dei metaboliti neuronali si possono fare valutazioni sulla funzionalità di diverse vie metaboliche quali quelle del sistema glutamminergico e del sistema GABAergico, sul metabolismo energetico neuronale, sulla osmoregolazione del SNC, oltre alla cellularità neuronale e gliale.

Il problema consiste nel fatto che le valutazioni sui singoli metaboliti però non hanno sempre un carattere patologico specifico.

Una riduzione delle concentrazioni di NAA (marker specifico dei neuroni) si osserva, ad esempio, sia nei pazienti con Demenza che con Malattia di Alzheimer, in aree cerebrali ischemiche o nei tumori (il suo riequilibrio dopo idonea terapia è proporzionale al recupero funzionale e alla normalizzazione morfologica). Il lattato diminuisce in tutte le situazioni che comportano ipossia-anossia ed aumenta nei tumori a crescita rapida; negli studi “in vivo” come detto l’insieme glutamina+glutammato definito in letteratura Glx aumenta nella patologia epatica; la colina, che aumenta nelle condizioni in cui vi è una alterazione delle membrane permettendo una valutazione del grado di demielinizzazione, è di rilevante importanza anche nelle patologie ischemico-anossiche acute e nelle neoplasie permettendo anche di valutare insieme alle alterazioni della risonanza del lattato, il grado di malignità dei tumori cerebrali; il mio-inositolo aumenta nell’Alzheimer e diminuisce nell’insufficienza epatica. La spettroscopia di risonanza magnetica (MRS), sia “in vivo” che “in vitro” è, quindi, una metodica biochimico-clinica potente ma di difficile utilizzazione in diagnostica; va specificato comunque che essa offre grandi vantaggi, soprattutto negli studi “in vivo” di alcune patologie; ad esempio, nelle Demenze degenerative in fase iniziale permette di differenziare, quando con l’esame neuroradiologico non è possibile, se la neurodegenerazione è di origine vascolare o psichiatrica, e permette quindi di indirizzare la diagnosi e la terapia; ancora, nelle sindromi extrapiramidali permette di distinguere tra il Morbo di

Parkinson ed altre Sindromi Parkinsoniane (sindrome di Steele-Richardson-Olszewski, atrofia multisistemica, degenerazione cortico-basale); grazie agli esami di spettroscopia protonica in vivo studiando le risonanze di NAA e di Cho è possibile valutare lo stato della sostanza bianca apparentemente normale all'esame classico MRI e di evidenziare meglio un possibile istaurarsi di un quadro patologico per una sclerosi multipla; ma, analogamente alla sclerosi multipla, una alterazione localizzata delle risonanze di NAA e di Cho è associabile a processi espansivi endocranici o locus epilettogeni (Epilessie del Lobo Temporale TLE, patologia a cui il nostro gruppo di lavoro al Policlinico ha dedicato un filone di ricerca data l'importanza sociale della diagnosi e cura precoce dell'epilessia [13]) o di diagnosticare malattie del metabolismo (iperglicinemia, leucodistrofie, mitocondriopatie) o miopatie metaboliche o patologie del parenchima epatico o vasculopatie o lo studio della fisiopatologia delle varie forme di Cefalea. In genere tutte queste diagnosi sono ampiamente indirizzate anche da altri esami a cui si sottopone il paziente; sono stati già citati gli esami di diagnostica per immagini con risonanza magnetica e gli esami elettroencefalografici ed elettromiografici che permettono di aumentare la specificità della diagnosi.

Il CSF è dunque si veicolo di una grande quantità di informazioni per le patologie del SNC ma per aumentare la specificità dell'esame, soprattutto per gli esami "in vitro", occorre studiare le alterazioni metaboliche di insieme non di pochi isolati metaboliti; occorre applicare le tecniche di analisi statistica multivariata essenziali nella diagnostica precoce per isolare e caratterizzare le impronte specifiche delle patologie in studio.

Il CSF è infatti una miscela complessa da trattare, composta in genere di decine e decine di metaboliti ed altri composti quali aminoacidi e proteine, generati dal funzionamento di tutte le cellule cerebrali. I neuroni sono le cellule "nobili" del cervello ma accanto ad essi si trovano, come abbiamo visto, altri tipi di cellule importanti per il sistema nervoso centrale, quali le cellule della glia ad esempio, per le quali è importante acquisire marker molecolari.

Come si è detto, tra i diversi cicli metabolici che caratterizzano il funzionamento d'insieme del SNC esistono numerose e complesse interazioni definite dai nodi e dalle reazioni comuni dei diversi percorsi metabolici. Queste interazioni complesse, variabili a volte in maniera significativa da individuo ad individuo sono difficilmente definibili dallo studio specifico delle alterazioni dei singoli metaboliti.

L'NMR dunque sebbene permetta di esplorare contemporaneamente il comportamento di decine di metaboliti del liquor, ha bisogno di opportune tecniche di elaborazione dei dati per estrarre tutto il contenuto informativo. In passato sono stati pubblicati diversi lavori sulla spettroscopia protonica del liquor per la caratterizzazione della SM.

Tipicamente gli studi classici effettuati mediante <sup>1</sup>H-MRS sul liquor e riportati finora in letteratura presentano tabelle con alcuni metaboliti selezionati mediante tecniche di analisi statistiche usuali, per cui le alterazioni dovrebbero riferirsi alle forme diverse di sclerosi multipla.

Di seguito si riportano i valori riferiti nel lavoro della Lynch pubblicato nel 1993 [14].

**Tabella 2. Elenco di alcuni dei metaboliti di importanza diagnostica in SM.**

	Literature value <sup>14-16</sup>	Control	MS	
Lactate	0.8 - 2.4 mM	2.07 $\pm$ 0.09	1.93 $\pm$ 0.07	NS
Acetate	0.05 - 4.5 mM	0.09 $\pm$ 0.01	0.12 $\pm$ 0.01	*(p = 0.032)
Glutamine	0.39 - 0.71 mM	0.55 $\pm$ 0.04	0.53 $\pm$ 0.05	NS
Citrate	0.3 mM	0.04 $\pm$ 0.02	0.37 $\pm$ 0.03	NS
Creatine	0.03 - 0.14 mM	0.14 $\pm$ 0.08	0.13 $\pm$ 0.01	NS
Glucose	3.4mM	3.14 $\pm$ 0.34	3.12 $\pm$ 0.12	NS
Formate <sup>1</sup>		3.03 $\pm$ 0.21	2.11 $\pm$ 0.41	*(p = 0.033)

Sempre nello stesso lavoro si riportano i dati comparati di questi metaboliti nel confronto tra sclerosi "short term" e sclerosi "long term".

**Tabella 3. Comparazione dei livelli metabolici in SM precoci e diagnosticate da molto tempo.**

	Longstanding MS	Early MS	
Lactate	1.96 ± 0.21	1.97 ± 0.09	NS
Acetate	0.14 ± 0.01	0.11 ± 0.01	NS
Glutamine	0.59 ± 0.11	0.49 ± 0.03	NS
Citrate	0.35 ± 0.03	0.39 ± 0.05	NS
Creatine	0.14 ± 0.02	0.12 ± 0.01	NS
Glucose	3.12 ± 0.17	3.11 ± 0.17	NS
Formate	2.01 ± 0.48	1.90 ± 0.62	NS
Unknown (2.82 ppm) <sup>†</sup>	0.18 ± 0.05	0.13 ± 0.03	NS

Le informazioni estraibili da questi dati sono altamente mascherate dalle variazioni metaboliche interpersonali. In un lavoro analogo pubblicato dal gruppo italiano della professoressa Simone nel 1996 [10], anche questo sulla spettroscopia in vitro del liquor come metodo di analisi e classificazione delle sclerosi multiple, una frase concludeva le considerazioni generali sull'alta variabilità dei valori di concentrazione dei vari metaboliti e sull'utilità dei metodi spettroscopici in diagnostica: "CSF 1H-MRS cannot be used as a stand-alone diagnostic method for the multiple sclerosis."

In realtà, come abbiamo visto, la complessità del SNC e dei cicli metabolici delle malattie e dei disordini neurologici rende bene conto della necessità di un approccio "metabonomico" al problema e delle naturali difficoltà che si incontrano in generale con le altre strategie di analisi nell'inseguire gli effetti di una alterazione di un singolo metabolita, o di un ristretto insieme di essi, su una patologia. Solo un approccio olistico può permettere una lettura di insieme utile per la comprensione della dinamica della patologia. È chiaro dunque dall'analisi dei complessi sistemi biologici in interazione coinvolti nel funzionamento basale e patologico del SNC che la possibilità di definire fenotipi patologici attraverso l'analisi dello spettro protonico NMR del liquor non può prescindere da una strategia di valutazione di tipo multivariato delle variazioni dei metaboliti nel loro insieme.

### **3.5. La strategia di analisi dei dati NMR protonici del liquido cerebrospinale.**

Le analisi dei dati sono state eseguite sia in ambiente MATLAB sia utilizzando il programma specifico SIMCA-P11 (Umetrics [www.umetrics.com](http://www.umetrics.com)).

L'ambiente di calcolo Matlab mette a disposizione un potente linguaggio di programmazione fortemente orientato all'analisi di dati sotto forma di matrici. Matlab è un ambiente di lavoro molto diffuso in quanto con esso è possibile creare dei programmi di lavoro autonomi con interfaccia grafico e ridistribuibili agli utenti sotto forma di toolbox. Sul sito ufficiale si possono trovare routine sviluppate da utenti e messe gratuitamente a disposizione.

Il programma SIMCA-P infine è un programma commerciale sviluppato di recente e fortemente orientato all'analisi metabonomica. Negli esperimenti di metabonomica basati sulla spettroscopia NMR uno dei passi fondamentali dell'analisi riguarda l'individuazione degli opportuni descrittori di spettro ovvero l'individuazione di un set di valori legati allo spettro NMR che ne descriva le proprietà [15]. È ormai prassi consolidata la riduzione della dimensionalità degli spettri NMR prima di tutto individuando, se esistono e se sono note a priori, specifiche regioni di interesse (Regions of Interest ROI's) e poi operando in queste ROI's una integrazione delle aree comprese in istogrammi di ampiezza di base opportuna definiti in letteratura "bins" o "bucket". Cruciale in questa operazione è la scelta dell'ampiezza di base di tali istogrammi. Dopo questa fase preliminare viene condotta un'analisi esplorativa sui dati prodotti, una analisi di carattere generale mediante la quale è possibile estrarre informazioni relative alla struttura dei dati, alle correlazioni tra le variabili, alla rilevanza e ridondanza con cui il problema è stato descritto, alla presenza di eventuali dati anomali o altro ancora. In particolare, la PCA (Analisi delle Componenti Principali) è oggi unanimemente riconosciuta come uno dei metodi fondamentali per l'esplorazione di dati relativi a sistemi multivariati biologici [16]. PCA è finalizzata ad estrarre la massima informazione possibile

contenuta in una struttura di dati multivariati, sintetizzandola in poche combinazioni lineari delle variabili stesse. Dal punto di vista geometrico, la PCA consiste in un processo di rotazione dei dati originali, effettuato in modo che il primo nuovo asse (che costituirà la prima componente principale) 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. L'algoritmo dispone le nuove componenti in modo che la varianza spiegata dalla componente  $n$ -esima sia sempre maggiore della varianza spiegata dalla componente  $(n+1)$ -esima. Il numero di questi nuovi assi (le componenti principali, PC) sarà quindi pari al numero di variabili originali; la matrice costruita a partire da queste combinazioni lineari viene definita matrice degli "scores". In genere si seleziona un ristretto numero di nuove componenti, talvolta due componenti, quelle che servono per esplicitare le aggregazioni dei dati in grafici bidimensionali. Un aspetto di grande rilevanza nello studio di problemi multivariati riguarda infatti la possibilità di "vedere" graficamente i dati. L'analisi delle componenti principali ci fornisce una soluzione algebrica che ci consente rappresentazioni grafiche molto efficaci per l'esplorazione iniziale delle proprietà dei dati. Ogni componente sarà una combinazione lineare delle componenti di partenza; i coefficienti di queste combinazioni lineari costituiscono la matrice dei "loadings".

Nella fase successiva, mediante l'utilizzo dei metodi di "Patterns Recognition" viene valutata la similarità relativa tra i campioni studiati (Clustering) e vengono estratte le caratteristiche specifiche delle classi di similarità riscontrate dall'analisi al fine di costruire un modello del sistema biologico.

Le tecniche di clustering utilizzate allo scopo in questo lavoro sono la HCA (hierarchical cluster analysis) e la FCM (fuzzy cluster analysis) [17]. In questa fase l'analisi viene condotta senza utilizzare informazioni sulle strutture di dati "esterne" alle metodiche. Le metodiche di analisi di questo tipo sono chiamate in letteratura "unsupervised", non supervisionate. Le due tecniche di clustering

differiscono per un aspetto molto importante: l'algoritmo HCM assegna un dato ad un cluster in maniera esclusiva (un dato appartiene ad uno ed uno solo dei clusters ipotizzati), mentre i metodi "fuzzy" assegnano un valore di appartenenza del dato ad ognuno dei clusters. Per le applicazioni di diagnosi medicale i metodi "fuzzy" rivestono in genere una importanza maggiore perché la natura degli stati patologici è difficilmente definibile in maniera univoca specialmente in un quadro complesso come quello metabolico di un sistema biologico. Entrambe queste procedure sono già implementate in ambiente Matlab ed ampiamente descritte in letteratura [15].

Nella fase successiva, dopo che le analisi non supervisionate hanno fornito un quadro iniziale della struttura intrinseca dei dati, si passa alla costruzione dei modelli mediante i metodi di classificazione supervisionata, quali DA (Discriminant Analysis), SIMCA (Soft Independent Models of Class Analogy), k-nn (k-nearest neighbours), SVM (Supported Vector Machine) e metodi di regressione supervisionata, quali PLS-DA (Partial Least Squares Discriminant Analysis)[15]. I risultati della classificazione supervisionata consistono in modelli matematici in grado di predire l'appartenenza di un nuovo campione ad una delle classi note ricavate dalla base dati; a tale scopo si utilizzano delle relazioni funzionali tra le variabili indipendenti (i descrittori molecolari) e una o più funzioni Y che possono essere di tipo qualitativo o quantitativo per caratterizzare ulteriormente i dati (Nota: in ambiente Matlab la maggior parte di queste procedure è implementata sotto forma di programmi gestibili tramite interfaccia grafico e tutte corredate da numerosi esempi applicativi).

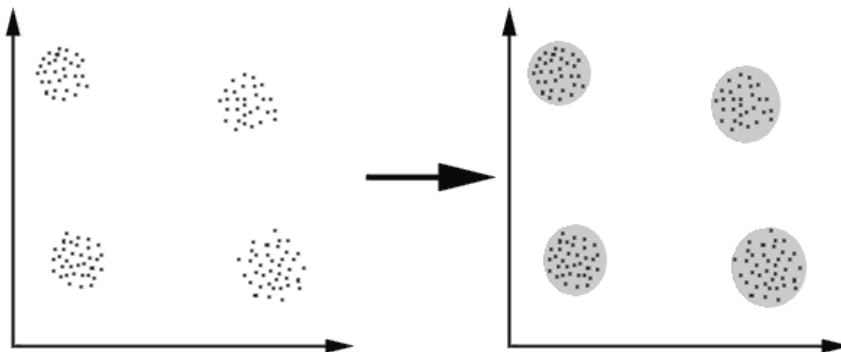
In pratica, la strategia di analisi in metabonomica si riassume quindi in quattro fasi importanti:

- acquisizione ed elaborazione (FT, correzione della fase, correzione della linea di base,) degli spettri  $^1\text{H-NMR}$ ;
- riduzione preliminare della dimensionalità del campione;

- data mining mediante analisi delle componenti primarie CA e classificazione (clustering) non supervisionata;
- classificazione (clustering) supervisionata con riconoscimento e modellizzazione delle strutture, denominate in gergo “features” o “patterns”, dei dati; tali modelli sono poi applicati per la classificazione di campioni incogniti.

Il punto di base della metodica sta proprio nell’uso delle tecniche di analisi statistica multivariata e di algoritmi di classificazione per il riconoscimento, la separazione e la definizione delle caratteristiche comuni delle classi in cui i dati si separano e che rendono conto delle similarità e delle differenze dei campioni, tecniche definite, appunto, di “clustering”.

Il clustering può essere definito come lo studio formale di algoritmi e metodi per raggruppare oggetti con caratteristiche comuni.



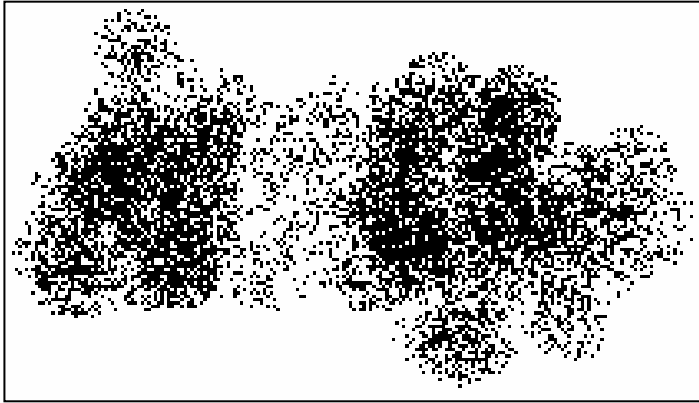
**Figura 16. Il clustering permette di determinare i parametri per descrivere le similitudini di sottogruppi di dati dell’insieme.**

Nella figura i quattro insiemi, grazie alla loro distribuzione nel piano, mostrano similitudini di proprietà e sono descrivibili mediante modelli specifici.

Il clustering può essere considerato il più importante problema di apprendimento non supervisionato.

Nel problema del clustering non è sempre possibile però definire in maniera precisa se un punto appartiene a un cluster oppure a un altro perché spesso questa definizione stretta non è possibile, non è consentita dalla natura del problema stesso, oppure a causa di artefatti che modificano le proprietà di alcuni spettri e rendono confusa la loro classificazione.





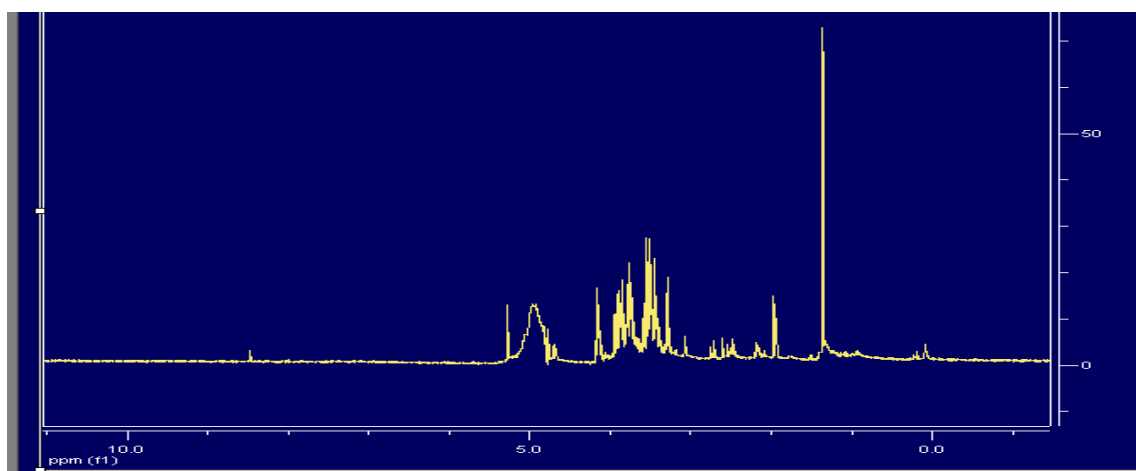
**Figura 17. Rappresentazione generica della natura non binaria ma “fuzzy” della classificazione di dati di tipo biologico.**

Le tecniche di clustering e di classificazione supervisionata devono allora saper distinguere i profili predominanti nelle classi che si vengono a formare e saperle distinguere parametricamente attraverso opportuni descrittori. Nel caso della classificazione degli stati patologici della sclerosi multipla ci siamo trovati di fronte proprio ad un problema di classificazione non univoca degli stati. Anche a causa di alcuni artefatti, di cui si parlerà più specificatamente nel capitolo 5°, si sono riscontrate alterazioni spettrali non collegati alla patologia, alterazioni che non hanno un correlato fisiologico; sono state messe a punto delle tecniche di elaborazione dei dati per rimuovere tali alterazioni. Inoltre, per classificare gli spettri nella maniera migliore si sono dovute adottare tecniche di classificazione non lineare basate sulla macchine per apprendimento a supporto vettoriale, SVM.

Le macchine a supporto vettoriale (Support Vector Machines, SVMs) sono un'insieme di algoritmi per la regressione e la classificazione di pattern, sviluppata recentemente da Vapnik ed il suo team presso i laboratori Bell AT&T. Le macchine a supporto vettoriale possono essere pensate come una tecnica per consentire l'apprendimento di classificatori polinomiali, in cui i pesi della rete neurale vengono ottenuti risolvendo un problema di programmazione quadratica con vincoli di uguaglianza o di tipo box, cioè in cui il valore di un

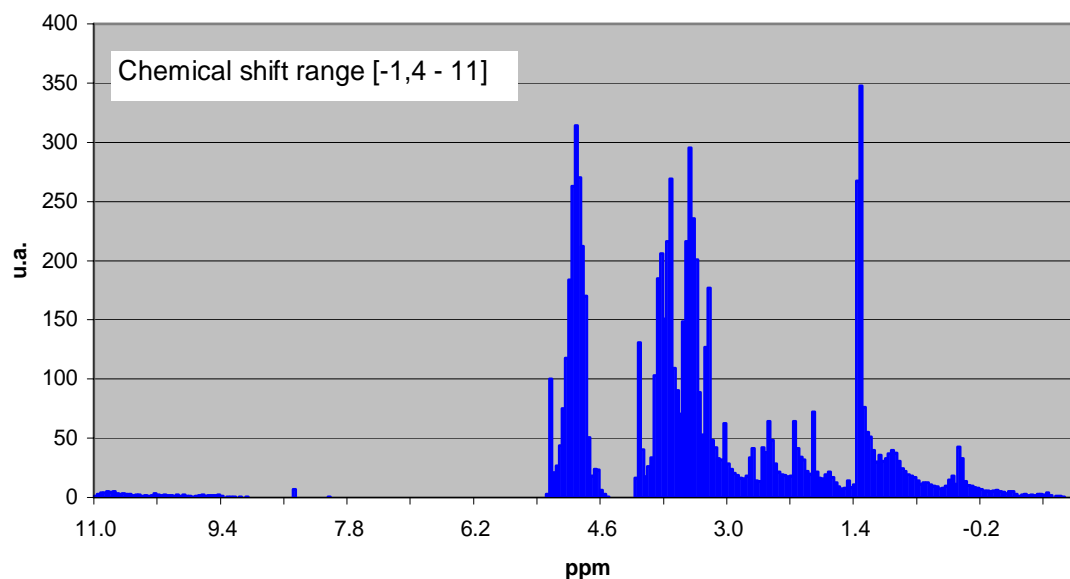
parametro deve essere mantenuto all'interno di due estremi. Alcune applicazioni per cui le SVM sono state utilizzate con successo sono il riconoscimento ottico dei caratteri, OCR (Optical Character Recognition, i programmi con cui gli scanner riportano in formato "editabile" i caratteri di un testo), l'identificazione di facce in immagini e la classificazione di testi. In ambiente Matlab è semplice costruire ed addestrare un classificatore SVM (v. procedura in Appendice 1).

Gli spettri acquisiti nei nostri esperimenti hanno una dimensione di 8096 punti distribuiti su una regione spettrale tra 11 e -1 ppm. Un'ampia regione di chemical shift attorno al picco dell'acqua è stata eliminata perché in diversi spettri è stato riscontrato un elevato picco residuo.



**Figura 18. Spettro NMR protonico di un campione di liquor cerebrospinale.**

Sulle regioni di spettro selezionate si è operata una integrazione a bucketing scegliendo una ampiezza di base dei buckets di 0,04 ppm. Dunque dagli spettri NMR si passa agli spettri integrati in "buckets".



**Figura 19.** Grafico degli integrali “bucket” per uno spettro NMR di liquor.

L’integrazione a buckets nelle regioni di chemical shift comprese tra 0,8 e 3,4 ppm e tra 6,6 ed 8,6 ppm, ha portato ad una riduzione del numero di descrittori spettrali da 8096 a 115 ottenendo così una prima riduzione di dimensionalità del campione.

Allo scopo di costruire gli elementi di base per il calcolo automatizzato delle proprietà degli spettri NMR da organizzare sotto forma di programmi autonomi ad interfaccia grafico per future applicazioni automatiche di analisi, abbiamo realizzato un funzione che esegue l’integrazione a bucket in ambiente MATLAB, l’ambiente di calcolo più diffuso nelle varie branche della scienza; in linguaggio Matlab abbiamo inoltre costruito le procedure per l’analisi dei clusters nei dati. Le procedure sono riportate nelle appendici della tesi.

Sono stati analizzati con questa procedura tutti i 45 campioni di liquor per cui la nostra matrice di dati iniziale  $X(45 \times 8096)$  diventa una matrice  $X(45 \text{ righe} \times 115 \text{ colonne})$ , 45 campioni da 115 descrittori ciascuno.

Come detto, passo successivo è l’applicazione dell’analisi PCA per un’ulteriore riduzione delle dimensioni e per il test su strutture similari nei dati.

La PCA individua in genere un set molto limitato di descrittori; le prime 4 o 5 componenti sono in genere capaci di rappresentare il 70-80 % della variabilità dei campioni. Si ottiene così il passaggio da una matrice  $X(45 \times 115)$  ad una matrice  $\hat{X}$  ( $45 \times 4$ ), ogni spettro è cioè rappresentato da un vettore a 4 componenti ciascuna espressa da una opportuna combinazione lineare dei 115 descrittori iniziali dello spettro NMR in cui i coefficienti di tale combinazione lineare rappresentano i loadings, i pesi di ciascuna variabile iniziale; per ogni spettro si otterranno dei 4-vettori

$$x_{PCA\mu}^i = (x_1, x_2, x_3, x_4)$$

esse sono le nuove coordinate, espresse nella matrice “score” dei punti sperimentali nello spazio metabonomico n-dimensionale; selezionando  $n=4$  si ottiene in genere un buon livello di variabilità spiegata dei dati.

A questa matrice si applicano le tecniche di classificazione o “clustering” per la descrizione delle similarità delle eventuali sottoclassi omogenee presenti nei dati.

Si produce quindi la matrice denominata 1H-MRS400CIS-SM\_PCA4 (matrice a 4 componenti PCA di spettri NMR protonici realizzati a 400MHz per la discriminazione tra CIS ed SM).

Tale matrice costituisce un primo tentativo di classificazione dei dati, una matrice sulle cui strutture interne abbiamo poi investigato con i metodi di clustering supervisionato SVM e di regressione di SIMCA. Il metodo SIMCA [15, 16] è stato introdotto da Wold nel 1974 e l’acronimo in letteratura è stato associato a diverse sigle. In sostanza viene calcolato un modello a componenti principali per ognuna delle classi definite a priori per l’insieme di dati; per la generica classe m-esima si ha:

$$X_m = T_m x L_m^t$$

dove  $T$  è la matrice “score” delle nuove coordinate ed  $L$  quella dei pesi definiti per la classe  $m$ .

SIMCA fornisce interessanti risultati nel trattamento dei dati “native” in presenza di artefatti. Tuttavia soffre nel calcolo di modelli se le classi sono sproporzionate nella dimensionalità. Nella classificazione in queste situazioni lavora meglio l’algoritmo di classificazione a supporto vettoriale SVM, un algoritmo di calcolo di recente sviluppo nella Teoria dell’Apprendimento [18]; c’è comunque da dire che nel caso di grosse matrici di dati l’algoritmo impostato per una analisi non lineare mette a dura prova il processore utilizzato per il calcolo.

La procedura di calcolo SVM adottata in questo lavoro è stata implementata in ambiente MATLAB; in questo lavoro per alleggerire il calcolo SVM è stata lanciata sui dati ridotti dopo PCA. Questa procedura è innovativa ed i campioni di liquor da acquisire in futuro per aumentare la dimensione della base dati concorreranno in maniera determinante alla calibrazione del metodo.

Nella Figura 20 si riporta il diagramma di flusso del processo di analisi dei dati elaborato per questi esperimenti.

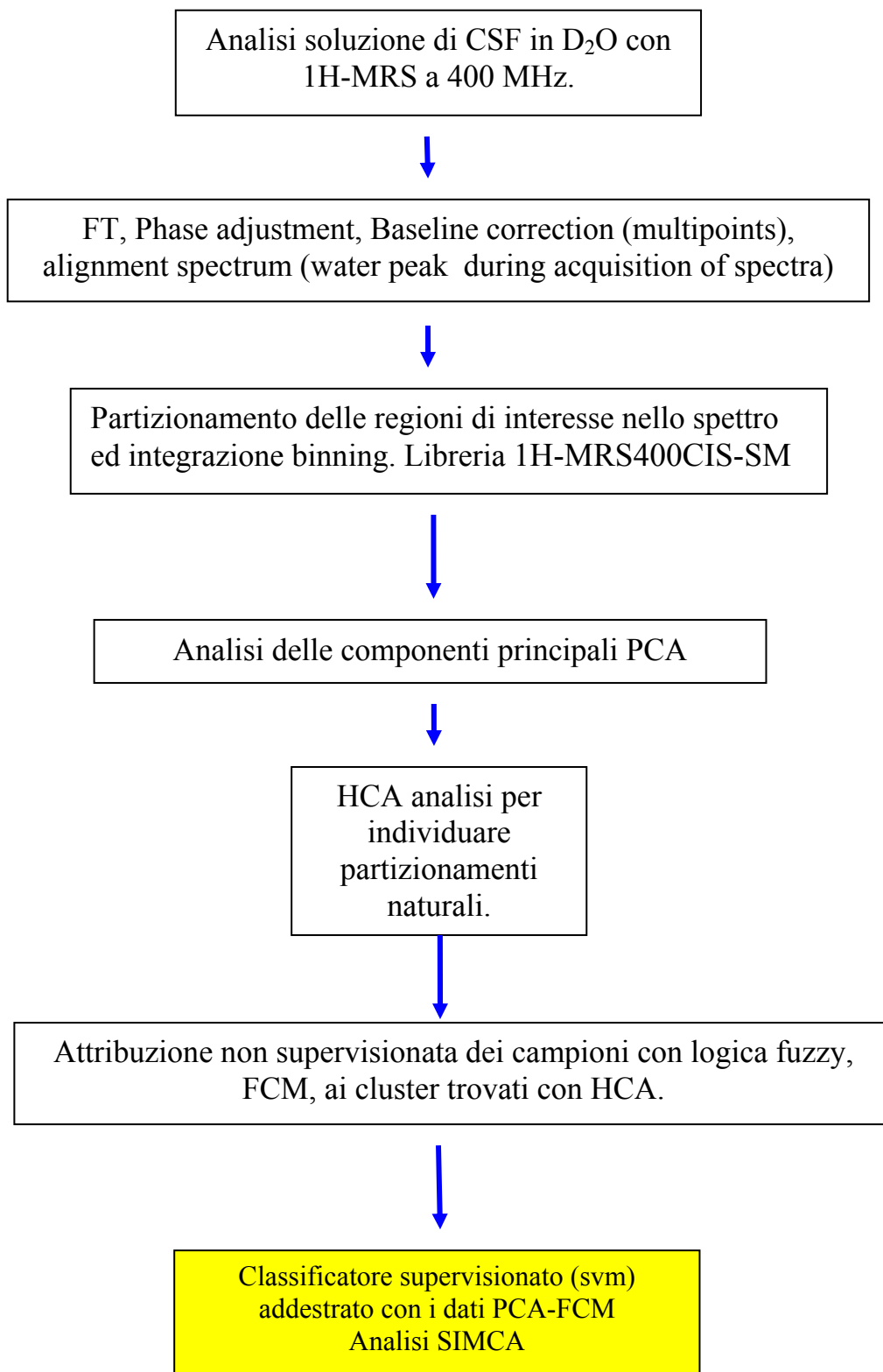


Figura 20. Diagramma di flusso dell'analisi dei dati.

Nel paragrafo successivo si presentano i risultati dell'applicazione di questa strategia di analisi dei dati.

### 3.6. Risultati.

Il database analizzato raccolto in una matrice principale  $X(45 \times 8096)$  è composto dagli spettri protonici NMR monodimensionali ottenuti da campioni di liquor Cerebrospinale.

Applicando le tecniche di analisi multivariata la matrice dati diventa una matrice  $X(45 \times 4)$ .

Il campione statistico analizzato è costituito da 10 controlli “non patologici SM” e da 35 patologici di cui 13 CIS e 22 SM. Delle 13 CIS acquisite tra il 2004 ed il 2005 11 sono divenute SM su un periodo medio di 2 anni. Attualmente rimangono dunque due casi di CIS di cui si vuole individuare in maniera precoce la possibile evoluzione.

Dividiamo i campioni in tre serie: i campioni acquisiti nel 2004, quelli del 2005 ed infine quelli del 2006. Un primo risultato interessante è venuto dallo studio dei campioni prelevati ed analizzati nel 2004.

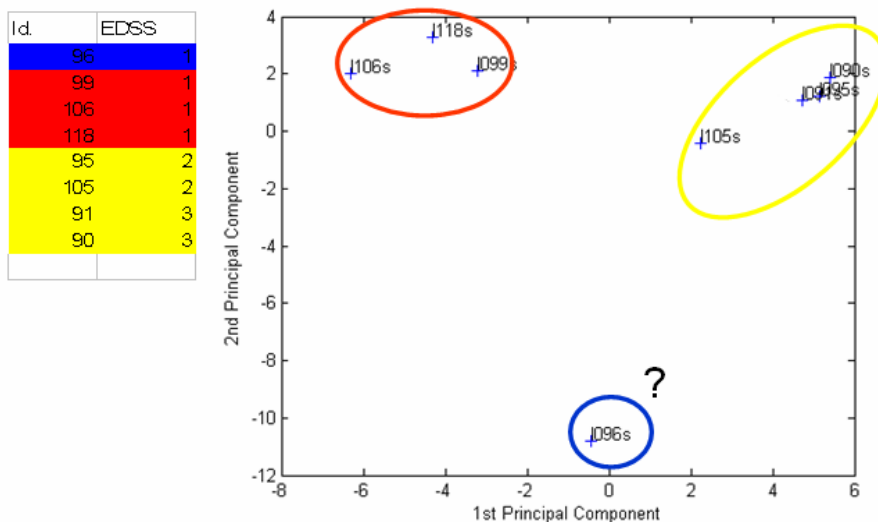
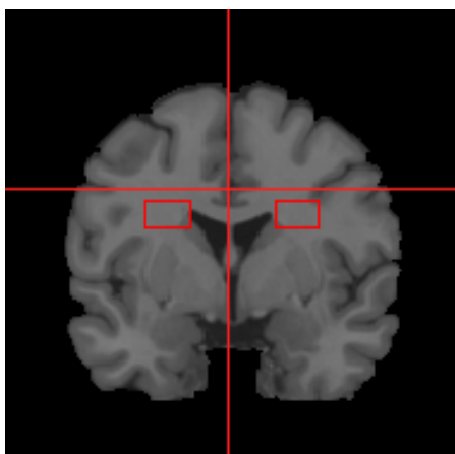


Figura 21. Analisi PCA su un primo campione di spettri. Correlazione tra i cluster e l'EDSS.

I due cluster ottenuti correlavano con il parametro di EDSS. L'EDSS ha un forte potere caratterizzante la SM; in genere l'indice di disabilità correla con un quadro di diffuse alterazioni del segnale MRI in T2 distribuite nella sostanza bianca che si trova nelle zone del cervello attorno alle strutture denominate “ventricoli”, che sono le strutture deputate alla produzione del liquor, e del tratto spinale. Queste alterazioni sono anche rilevate mediante la spettroscopia in vivo effettuate in tali zone.

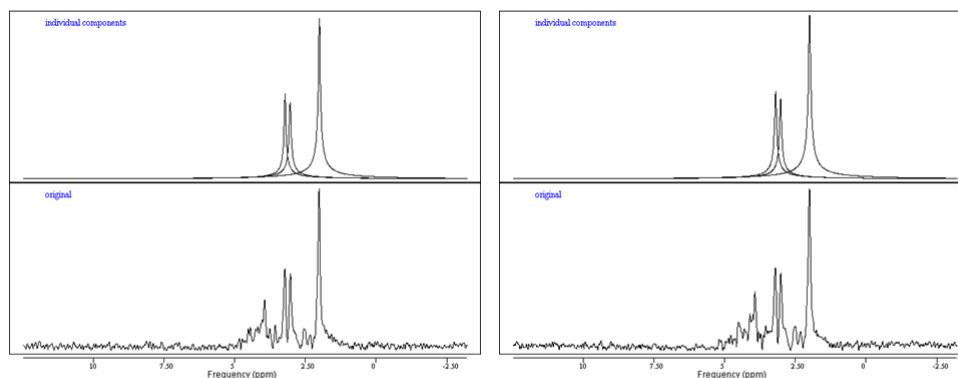
Di seguito riportiamo una misura effettuata dal nostro gruppo presso il Policlinico Universitario di Monserrato.



**Figura 22. Localizzazione dei voxel per l'esame di spettroscopia nella regione periventricolare.**

In genere si calcola l'asimmetria del picco del N-Acetyl Aspartato tra i due emisferi. Questo valore ci aiuta ad esprimere l'alterazione della sostanza bianca espressa tramite le alterazioni del segnale T2, le lesioni o placche di demielinizzazione, e l'indice EDSS di disabilità. Gli spettri acquisiti in spettroscopia in vivo nei due voxel periventricolari di Figura 22 hanno la forma riportata nella figura seguente e dalla quale si ricavano le concentrazioni dei metaboliti di interesse.





Individual components and real spectra of the WM in Pz1 dx.

Individual components and real spectra of the WM in Pz1 sx.

**Figura 23. Postprocessing degli spettri 1H-MRS relativi alle zone di sostanza bianca periventricolare.**

Questi risultati preliminari ci hanno spinto ad aumentare il numero di spettri acquisiti con l'intenzione di effettuare sui pazienti anche la spettroscopia in vivo. Purtroppo in questa fase preliminare del progetto non è stato possibile proseguire con questa strategia di ricerca con dati acquisiti in parallelo “in vivo” “in vitro” per diverse ragioni sia tecniche che etiche.

È oramai assodato che la composizione chimica del CSF riflette direttamente o indirettamente lo stato dei processi biochimici che occorrono nel parenchima cerebrale; inoltre le informazioni disponibili attualmente dalla routine clinica sul CSF risultano incomplete per una diagnosi differenziale in SM, e quindi le informazioni ricavate dalla spettroscopia NMR del CSF si candidano come una promettente integrazione delle informazioni diagnostiche estraibili dal CSF. Tuttavia, problematiche di tipo tecnico, medico e di etica limitano ancora molto lo sviluppo delle ricerche sulla MRS del CSF e limitano molto l'esplorazione delle potenzialità diagnostiche della metodica e della sua diffusione a livello di routine clinica. Per questo risulta molto problematico attuare un protocollo di ricerca innovativo sul liquor; occorre prima definire in maniera analitica le potenzialità di nuovi sistemi di analisi.

Rimanendo comunque l'esigenza di incrementare il numero di campioni di liquor analizzati abbiamo allora utilizzato i liquor conservati congelati a  $-20\text{ }^{\circ}\text{C}$  presso il centro di SM; questi campioni sono quanto resta dei prelievi effettuati

sui pazienti per le analisi cliniche di routine. Non tutti i campioni conservati possono essere utilizzati per le analisi soprattutto perché le quantità minime di cui si deve disporre per l'analisi NMR è di 2 ml. Infatti quantità minori darebbero grosse difficoltà per acquisire un segnale sufficiente in tempi idonei. Si conosce bene il fenomeno di degradazione del liquor quando viene manipolato a temperatura ambiente per molto tempo [11]. Inoltre rimane sempre il problema dei controlli. È difficile ottenere campioni di liquor di controlli sani; si possono utilizzare i controlli patologici di altre malattie ma anche in questo caso ci sono notevoli problemi dovuti alla possibilità di disporre di campioni quanto più omogenei possibile.

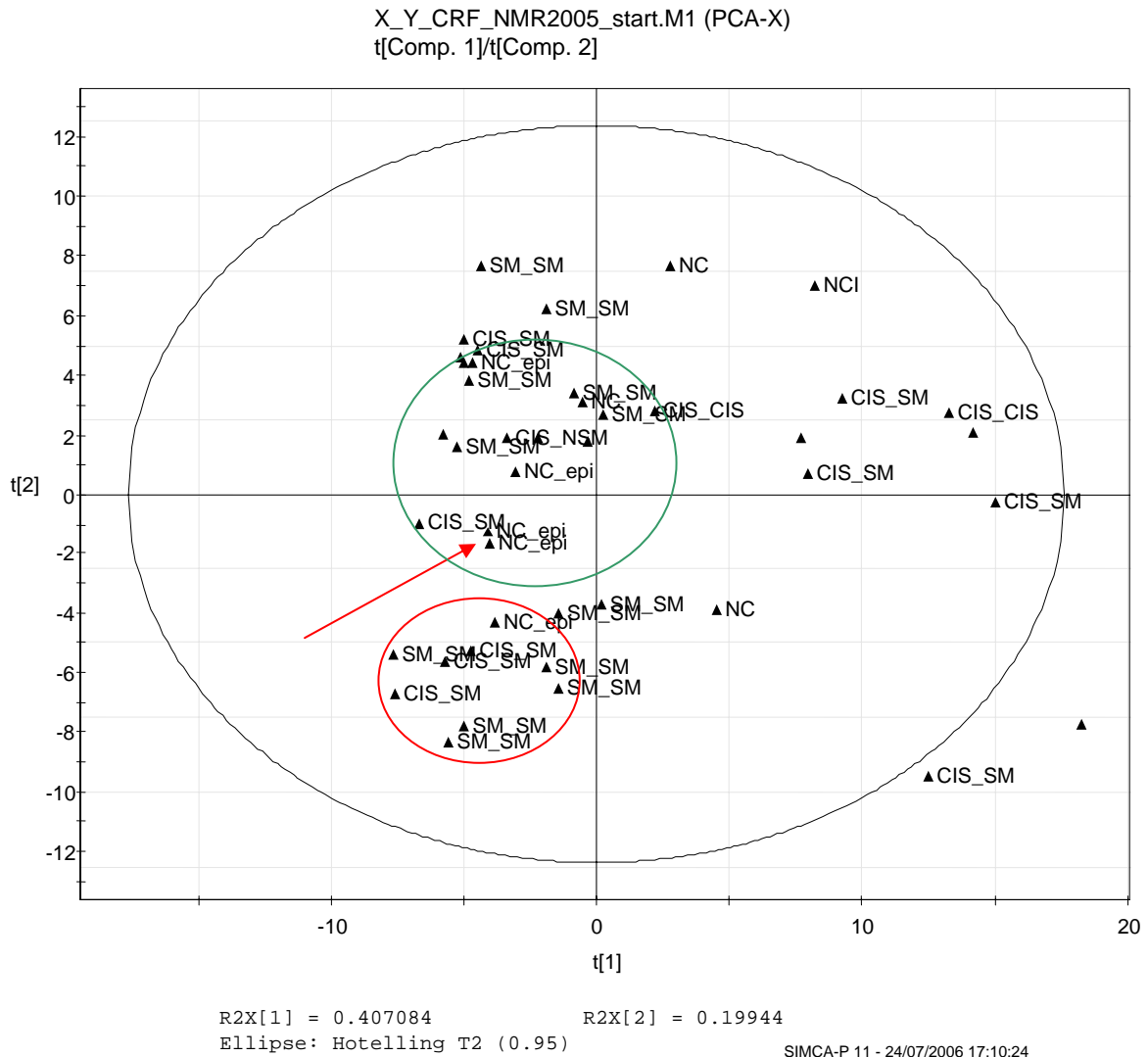
Si è quindi deciso di selezionare ed analizzare i campioni conservati per aumentare soprattutto la popolazione dei patologici ed arrivare ad una caratterizzazione dei fenotipi patologici di sclerosi multipla.

Durante il 2005 sono stati acquisiti ed analizzati altri 20 spettri patologici SM ed altri 5 controlli patologici non SM di diversa natura (Epilessia farmaco-resistente, Malattia di Charcot-Mariè-Thoot).

Dallo studio di questi campioni sono nate le osservazioni sulla degradazione dei campioni di liquor.

A queste problematiche probabilmente è possibile ricollegare la situazione di mixing degli stati riscontrabile nel nostro database che si esprimono graficamente nella PCA riportata di seguito nella Figura 24.

Analizzando l'intera matrice di dati acquisiti, 42 campioni in totale considerando che alcuni spettri sono stati eliminati per ragioni tecniche, non si è potuto rilevare una evidente ripartizione in classi degli spettri che potesse correlare con le diagnosi a posteriori e con i parametri clinici forniti. I dati viceversa correlavano con una ripartizione temporale secondo cui i campioni del 2005 risultavano ben distinti dai campioni del 2004.



**Figura 24.** PCA sulla matrice native dei dati. I cluster non sono ben separati ed alcuni degli spettri hanno similarità non specifiche

A questo punto si è deciso di rivedere il processamento dei dati abbandonando tutte le procedure di trattamento automatiche e lavorando sui singoli spettri operando “fasatura”, apodizzazione e correzione della linea di base manualmente e per ciascun spettro. Si è evidenziato la presenza di cluster con configurazioni miste, in cui si mescolavano cioè stati patologici e stati di controllo. Utilizzando le tecniche di cluster analisi basate sulla logica “fuzzy” sono state ricavate le popolazioni di due cluster uno definito benigno, per l’alto contenuto di spettri di controllo rispetto ai patologici (in particolare 5 controlli e 2 CIS-MS) e l’altro definito patologico SM per la sovrabbondanza di spettri SM

in esso contenuti (6 SM, 5 CIS-MS ed 1 controllo neurologico).

**Tabella 4. “Clusterizzazione” degli spettri NMR dopo PCA.**

<b>Cluster1</b> C1(-0,1 -1,12) u.a.	<b>Cluster2</b> C2(-0,09 -3,01) u.a.
'L090_04' SM SM	'L040_05' CIS SM
'L095_04' SM SM	'L004_05' CIS SM
'L096_04' SM SM	'L009_05' CNI
'L099_04' SM SM	'E_001' CN_epi
'L106_04' SM SM	'E_002' CN_epi
'L113_04' SM SM	'L066_05' CNI
'L098_05' CIS SM	'E_005' CN_epi
'L114_05' CIS SM	
'L014_05' CIS SM	
'L114_04' CIS SM	
'L013_05' CIS SM	
'L011_05' CNI	

I centroidi ed i raggi di tali clusters dipendono dalle caratteristiche di similitudine degli spettri assimilati e dell'entità delle differenze. Le classi sono costituite da campioni alquanto disomogenei e questo rende la clusterizzazione ottenuta poco affidabile in diagnosi.

L'ipotesi testata in questa fase dell'analisi è relativa alla definizione delle caratteristiche di insieme dei due cluster. La presenza di spettri con artefatti modifica sia il raggio dei cluster che la posizione del centro, parametri entrambe definibili come caratteristiche geometriche nello spazio metabonomico, lo spazio n-dimensionale in cui si proiettano i dati dopo la decomposizione in componenti principali. Utilizzando la logica “fuzzy” ad ogni spettro si associa una funzione di appartenenza ad ognuno dei cluster di cui si ipotizza la presenza. Per esempio i due controlli epilettici indicati nella figura seguente hanno mostrato una funzione di appartenenza del 45% al cluster “benigno” ed un 23% al cluster dei patologici per SM (la somma è diversa da uno perché si è testata l'ipotesi di partizione dei dati a 4 classi).

Sulla base di queste ipotesi si sono selezionati due cluster di riferimento, definiti benigno e patologico, contenenti le decomposizioni PCA a 4 componenti degli spettri con una funzione di appartenenza maggiore del 40% per la classe considerata.

La soglia così bassa e la presenza di miscele di configurazioni rende la ripartizione poco significativa ma ci ha permesso di realizzare una procedura di

analisi molto promettente che basandosi su una semplice ipotesi di incremento della popolazione statistica potesse darci le basi per costruire un “discriminatore” basato sulle macchine a supporto vettoriale, SVM.

In pratica gli spettri dei cluster hanno permesso di “addestrare” un classificatore non lineare; usando i 4-vettori delle PCA degli spettri selezionati si è costruita una matrice di training per il classificatore SVM. Il classificatore è stato poi lanciato su un sottoinsieme di spettri campione usati per l’analisi. In questo campione di test c’erano anche i due spettri CIS che sono stati classificati uno come CIS\_CIS, cioè come singolo episodio di demielinizzazione, e l’altro come CIS\_SM come cioè episodio di demielinizzazione destinato ad evolvere come SM.

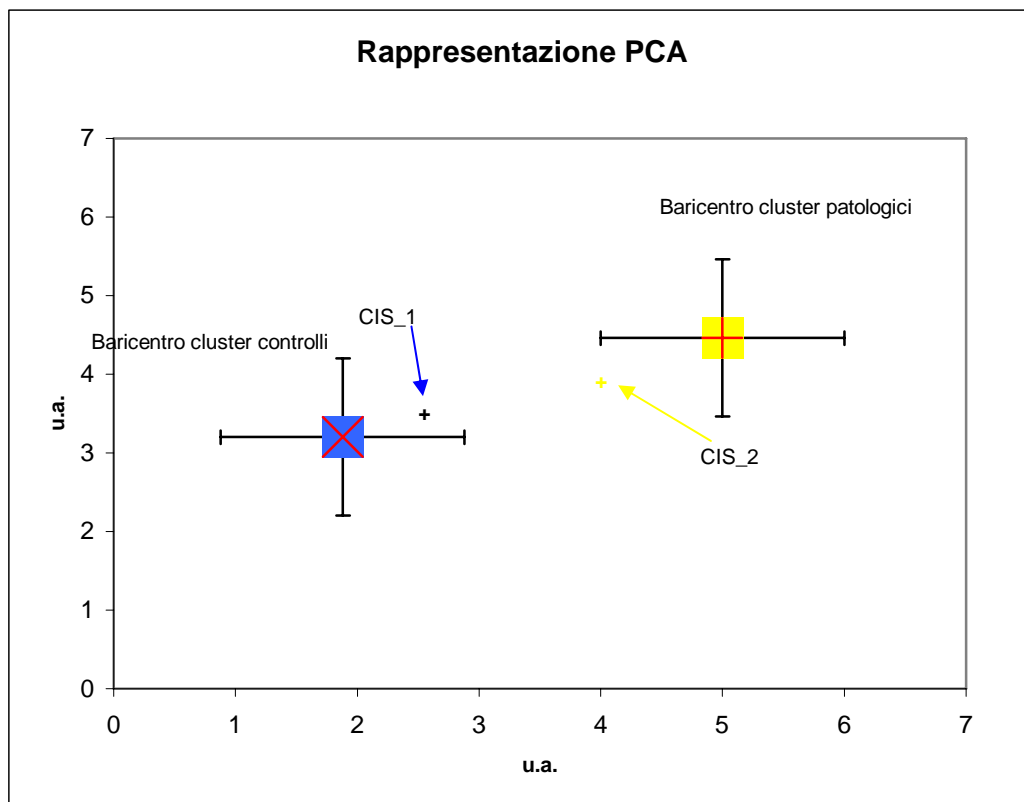


Figura 25. Rappresentazione grafica della classificazione operata con SVM su dati incogniti di CIS.

I risultati di questo lavoro sono stati presentati al congresso Nazionale della SIN 2006 [20]. Essi hanno un forte valore dal punto di vista metodologico ma chiaramente la ridotta dimensione del campione rende statisticamente poco significativo il potere predittivo del classificatore.

Aumentando il numero degli spettri il contributo degli spettri con artefatti è destinato a diminuire e del resto l'ipotesi dei grandi numeri è proprio l'idea di base della metabonomica.

Tuttavia la matrice di dati disponibile non poteva prevedere un incremento significativo a breve termine per cui ci siamo concentrati sulle possibili soluzioni via software del problema degli artefatti.

Come detto l'analisi delle componenti principali non individuava cluster correlabili ai cluster patologici.

Gli artefatti di spettro, che caratterizzano principalmente i campioni del 2005, generano, anche grazie al basso numero di campioni acquisiti, la difficoltà nell'individuazione di proprietà similari dello stato dei campioni.

La rimozione degli artefatti può essere eseguita attraverso diverse procedure. Il metodo adottato e presentato in questo lavoro si basa sull'algoritmo di classificazione supervisionata SIMCA, Soft Independent modelling of Class Analogy, con filtro di correzione ortogonale del segnale, definita in letteratura OSC, Orthogonal Signals Correction [21].

In pratica si adotta una funzione  $Y$  che viene costruita come una combinazione lineare dei parametri di disabilità EDSS, dell'Indice di Link  $Ilink$  e del numero di bande oligoclonali  $BO$ ; questa ipotesi di lavoro è una ipotesi preliminare, basata in parte sui risultati ottenuti sui dati del 2004 ed in parte dai dati generali sulla patologia noti dalla letteratura.

In futuro si prevede di testare una serie di funzioni  $Y$  da legare alla matrice  $X$  dei dati e testare la funzione migliore .

Il vettore  $Y$  considerato è una combinazione lineare dei valori clinici espressi dalla usuale routine cliniche sul CSF:

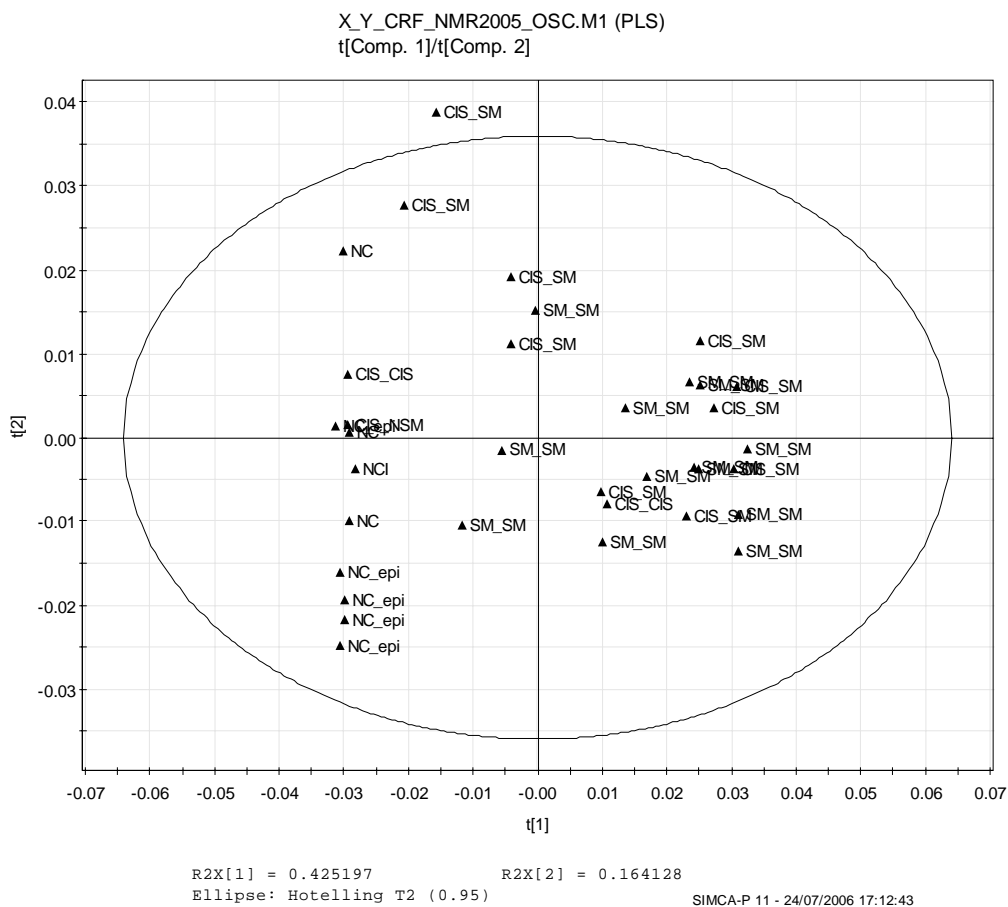
$$Y = a \times EDSS + b \times Ilink + c \times BO$$

La matrice dei descrittori di stato  $X$  si può pensare allora come composta da due componenti, una linearmente dipendente dalla  $Y$ , l'altra, "ortogonale",

statisticamente indipendente dalla Y.

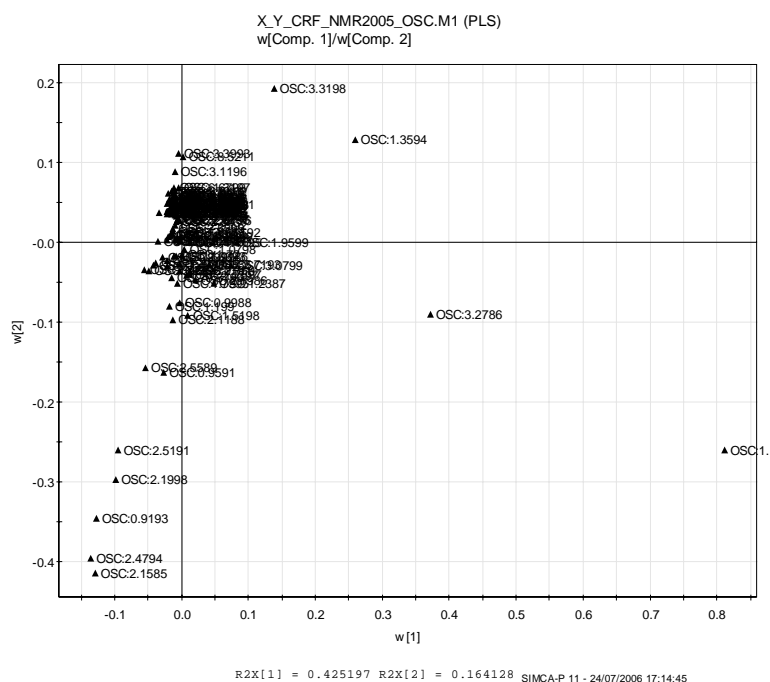
Il filtro OSC rimuove tale componente, che tipicamente può essere pensata come quella legata alle variazioni da artefatti sullo spettro e quindi non collegata a variazioni dei parametri clinici ed ortogonali alla classificazione “patologici-controlli sani”.

L'applicazione del Filtro OSC e relativa matrice di PCA regressiva (PLS-DA Partial Least Square Discriminant Analysis) a due componenti sono mostrate nella seguente figura.



**Figura 26. Decomposizione a componenti principali dopo il filtraggio delle componenti di artefatto.**

La matrice dei nuovi loadings esprime il seguente grafico relativo alle prime due componenti:



**Figura 27. Loadings della analisi delle componenti principali dopo il filtraggio degli artefatti.**

Il modello PLS-DA scelto può essere variato ed adattato operando l'esclusione degli outliers e ricalcolando le componenti principali.

Un primo modello PLS-DA senza alcuna manipolazione dei dati da i seguenti "loadings" responsabili della partizione:

**Tabella 5. Matrice de metaboliti responsabili della separazione controlli-patologici SM acquisita con analisi PLS-DA.**

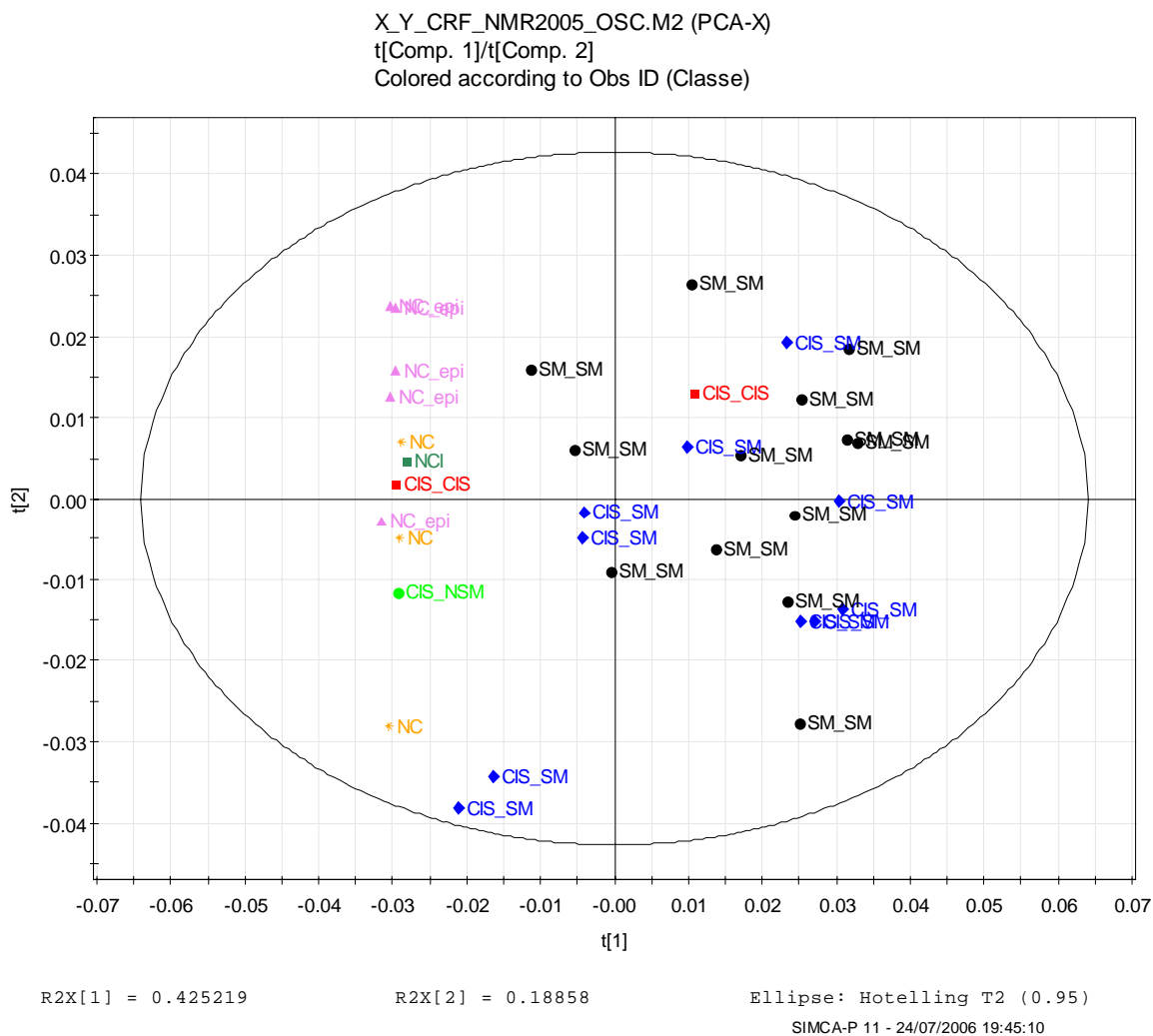
<b>Var ID (Primary)</b>	<b>M1.VIP[2]</b>	<b>Metaboliti di riferimento della regione spettrale</b>
<b>OSC:1.3991</b>	8.70397	A. lattico, A. 2-idrossibutirrico
<b>OSC:3.2786</b>	3.99463	Mio-Inositolo
<b>OSC:1.3594</b>	2.78236	A. Lattico
<b>OSC:3.3198</b>	1.48542	Glucosio
<b>OSC:2.4794</b>	1.45865	A. 4-idrossibutirrico
<b>OSC:2.1585</b>	1.38415	Glutammina
<b>OSC:0.9193</b>	1.36103	Isoleucina
<b>OSC:2.1998</b>	1.04618	Glutammina
<b>OSC:2.5191</b>	1.01069	Nota1
<b>OSC:1.9599</b>	0.880928	N-acetyl tirosina
<b>OSC:3.0799</b>	0.787978	Creatina



<b>OSC:1.4388</b>	0.590033	A. 2-idrossibutirico
<b>OSC:2.5589</b>	0.57744	A. 4-idrossibutirico
<b>OSC:1.2387</b>	0.526903	A. 3-idrossibutirico
<b>OSC:2.2395</b>	0.513341	Acetone
<b>OSC:3.2395</b>	0.502836	Colina

Nota1=picco del 4-idrossibutiric acid oppure picco a 2,5 del 3-idrossibutiric acid in colleg. con il picco a circa 1,33.

Dopo il filtro e la rimozione di alcuni “outliers” il database viene modificato si ottiene una matrice X' a cui è possibile riapplicare l'analisi PCA. Si possono produrre e confrontare così diversi modelli statistici.



**Figura 28. Modello ottenuto mediante analisi PCA sui dati filtrati. Le colorazioni evidenziano la collocazione degli spettri raccolti nelle varie classi (Controlli, CIS\_SM, SM\_SM, CIS\_CIS)**

Nel modello M2 riportato in Figura 28 migliora la significatività della seconda componente dopo l'eliminazione di alcuni "outliers".

I dati di "scores" precedenti sono ancora accompagnanti dal grafico dei "loadings":

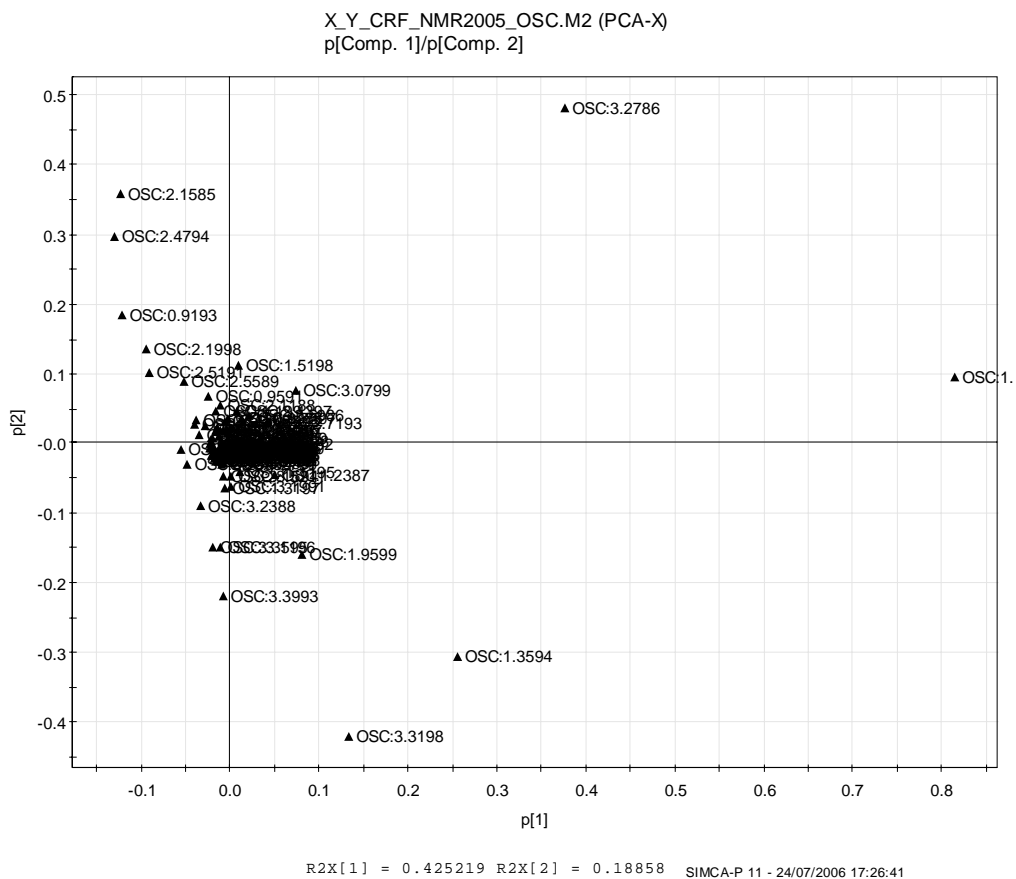


Figura 29. Grafico dei loadings del modello M2.

che in forma matriciale è:

Tabella 6. Matrice dei loadings del modello M2

Var ID (Primary)	M2.p[1]	M2.p[2]
OSC:1.3991	0.815151	0.0965744
OSC:3.2786	0.375688	0.480764
OSC:1.3594	0.256707	-0.304928
OSC:3.3198	0.134134	-0.419595
OSC:1.9599	0.0814412	-0.160361
OSC:3.0799	0.0741704	0.0777578
OSC:1.2387	0.0495656	-0.0452594
OSC:2.7193	0.0415855	0.0299669
OSC:2.5986	0.0211218	0.0416075

<b>OSC:1.7597</b>	0.0128951	0.0362735
<b>OSC:2.759</b>	0.0125065	0.00652257
<b>OSC:1.1195</b>	0.0118775	-0.0420223
<b>OSC:1.5198</b>	0.0103466	0.112863

L'analisi ha mostrato che la presenza di artefatti negli spettri NMR produce una errata classificazione di alcuni di essi ed è stata applicata una procedura di “filtraggio” dei dati; l'algoritmo di filtro prevede la separazione della matrice dei descrittori X in due componenti indipendenti, la prima componente calcolata è quella ortogonale alla funzione di riferimento Y ed è quella rimossa dal database. Rimane una matrice X' da cui dipende la Y.

Gli spettri sono stati colorati sulla base di una partizione in classi: SM\_SM, CIS\_SM, NC, NCI, NC\_epi ed infine CIS\_CIS, i due casi di riferimento.

Il modello ha bisogno di calibrazioni che derivino dal riequilibrarsi delle popolazioni nelle varie classi.

In questo senso i miglioramenti operati selezionando diversi modelli con eliminazione di dati lontani dal modello calcolato non sono stati significativi.

Ci limitiamo allora a considerare il primo Modello M1. Il set di variabili metaboliche covarianti responsabile nel modello M1 della suddivisione è riportato in tabella 1.

Essi rappresentano degli intervalli di chemical shift; considerando la prima variabile descritta come OSC1:1.3991 essa va letta come l'intervallo relativo alla regione spettroscopica da 1,39 ad 1,35. In questa regione cade uno dei due picchi del doppietto di risonanza del lattato a 1,34 ppm.

La regione spettrale indica appunto tali intervalli ed accanto la colonna “Metaboliti di riferimento della regione spettrale” indica i metaboliti noti in letteratura che ricadono in questo intervalli.

La colonna M1.VIP[2] esprime il livello di importanza della regione spettroscopica nella differenziazione dei dati: di solito si considerano come significativi i valori >1.

**Tabella 7.** Intervalli di chemical shift e metaboliti di riferimento per il modello della SM.

<b>Var ID (Primary)</b>	<b>M1.VIP[2]</b>	<b>Regione spettrale</b>	<b>Metaboliti di riferimento della regione spettrale</b>
<b>OSC:1.3991</b>	8.70397	1,39-1,35	A. lattico, A. 2-idrossibutirrico
<b>OSC:3.2786</b>	3.99463	3,28-3,24	Mio-Inositolo, Glucosio
<b>OSC:1.3594</b>	2.78236	1,35-1,31	A. Lattico
<b>OSC:3.3198</b>	1.48542	3,32-3,28	Glucosio
<b>OSC:2.4794</b>	1.45865	2,47-2,43	Acido 4-idrossibutirico
<b>OSC:2.1585</b>	1.38415	2,15-2,11	Glutammina
<b>OSC:0.9193</b>	1.36103	0,92-0,88	Acido 2-idrossisovalerico, Isoleucina
<b>OSC:2.1998</b>	1.04618	2,2-1,98	Glutammina
<b>OSC:2.5191</b>	1.01069	2,52-2,48	Nota 1
<b>OSC:1.9599</b>	0.880928	1,96-1,92	N-acetil tirosina, Picco incognito
<b>OSC:3.0799</b>	0.787978	3,08-3,04	Creatina
<b>OSC:1.4388</b>	0.590033	1,44-1,40	Acido 2-idrossibutirico
<b>OSC:2.5589</b>	0.57744	2,56-2,52	Acido 4-idrossibutirico
<b>OSC:1.2387</b>	0.526903	1,24-1,20	Acido 3-idrossibutirico
<b>OSC:2.2395</b>	0.513341	2,24-2,20	Acetone

Nota1 = picco del 4-idrossibutiric acid oppure picco a 2,5 del 3-idrossibutiric acid in correlazione. con il picco ad 1,33ppm.

Questa è la matrice che definisce l'impronta, il modello di variazione dei metaboliti responsabile secondo l'analisi prodotta delle differenze fenotipiche tra controlli e patologici per SM.

I modelli M1 ed M2 messi a punto con queste tecniche di elaborazione prevedono anch'essi la possibilità di etichettare una parte degli spettri come test del modello. Ancora una volta si può individuare il comportamento dei due

spettri CIS. Questa volta la similarità dei due spettri con i cluster benigni e patologici è più marcata; i due spettri rappresentati in rosso sono ben visibili e separati come mostrato nella precedente figura. Ripetiamo che in metabonomica non sono tanto importanti i singoli metaboliti quanto il loro insieme ed i pathway metabolici in cui sono coinvolti d'insieme. Tuttavia dal punto di vista medicale la presenza del lattato come risultato finale di questa analisi ha una grande importanza nell'ambito della classificazione della patologia della SM. È interessante partire dai pathways metabolici in cui è coinvolto il lattato (Cap. 3.4); partendo da quello relativo alla glicolisi/gluconeogenesi si ritrovano nodi importanti in cui avvengono reazioni che interessano altri cataboliti. È interessante notare come la variazione di glucosio sia collegata alla variazione di lattato e come l'analisi effettuata riporti queste due variabili in forte correlazione.

Il set di metaboliti in Tabella 7 individua quindi l'impronta specifica di differenziazione degli stati patologici di sclerosi multipla dai controlli.

#### ***4. Analisi metabonomica 1H-NMR su urine di neonati “pretermine” .***

Le metodiche messe a punto e presentate in questo lavoro hanno comunque carattere generale e possono essere applicate a molti problemi di diagnostica medica. Attualmente ad esempio, il gruppo di lavoro ha in essere una serie di studi analoghi a quello presentato nei capitoli precedenti; in questo capitolo presentiamo, come esempio, lo studio mediante spettroscopia protonica di campioni di urine di neonati nati “pretermine”. In questa condizione i neonati possono sviluppare diverse patologie e, anche a scopo preventivo, sono sottoposti comunque a trattamenti farmacologici protettivi. Questi bambini, ad esempio, che in genere hanno un'insufficienza renale alla nascita dovuta proprio alla nascita prematura, vanno incontro spesso a quadri di tossicità renale, dovuti alle terapie a cui sono sottoposti e spesso in contemporanea presenza di altri disordini (insufficienza polmonare, epatica, squilibri elettrolitici). Risulta molto importante individuare se esistono classi di neonati pretermine, per natura predisposti a sviluppare questi quadri patologici più di altri. Allo scopo di verificare la veridicità di queste ipotesi si stanno studiando con le tecniche metabonomiche gli spettri NMR di campioni di urine prelevati a tempi regolari e dalla dinamica metabolica risultante, oltre alla capacità di risposta alle terapie si vede anche se esistono stati metabolici iniziali comuni che rappresentino fenotipi caratteristici, in analogia ad altri studi su neonati presentati in letteratura [22].

L'urina, come è noto, è il liquido prodotto dai reni che filtrano il sangue per depurarlo dalle scorie prodotte dal metabolismo. Tramite l'urina quindi si eliminano dall'organismo i prodotti di scarto e l'eccesso di acqua o di sostanze che vi sono disciolte.

Il campo di indagine offerto dall'analisi delle urine è vastissimo e complesso. Molte delle analisi usuali portano a diagnosticare le disfunzioni dei reni, ma le

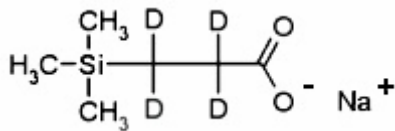
indagini possono mettere in risalto anche tanti altri problemi (es. diabete, infezioni, patologie a carico del fegato etc.).

Lo studio all'NMR delle urine per certi versi risulta complicato quanto quello del Liquor sia a causa dell'alta densità delle risonanze soprattutto nella regione di chemical shift degli zuccheri sia a causa dell'elevata variabilità interpersonale caratteristica di questo biofluido. Del resto la facilità di raccolta e la totale assenza di disturbi per il paziente fanno sì che si cerchi, per quanto possibile, di esplorare le potenzialità diagnostiche dei profili metabolici ottenibili dalle urine per la caratterizzazione di alcune patologie. Inoltre la preparazione dei campioni in genere è semplificata rispetto al liquor e anche il lavorare a temperatura ambiente non degrada il campione; i campioni ad esempio, possono essere analizzati senza bisogno di liofilizzazione; questo riduce le operazioni di manipolazione e quindi anche di possibile contaminazione dei campioni.

Il database è costituito attualmente da 28 campioni di urine provenienti dalla Clinica di Patologia Neonatale e Terapia Intensiva Neonatale dell'Università di Cagliari; 10 campioni di urine patologici e 4 campioni di urine di controllo sono stati prelevati ad un tempo iniziale  $t=0$ ; questi campioni sono utilizzati per differenziare gli stati iniziali patologici da quelli di controllo. Gli altri 14 spettri riguardano prelievi di urine effettuati ad intervalli di tempi regolari su 5 neonati di cui si è monitorato anche terapia, peso ed altri parametri clinici che concorrono a formare il quadro completo del paziente.

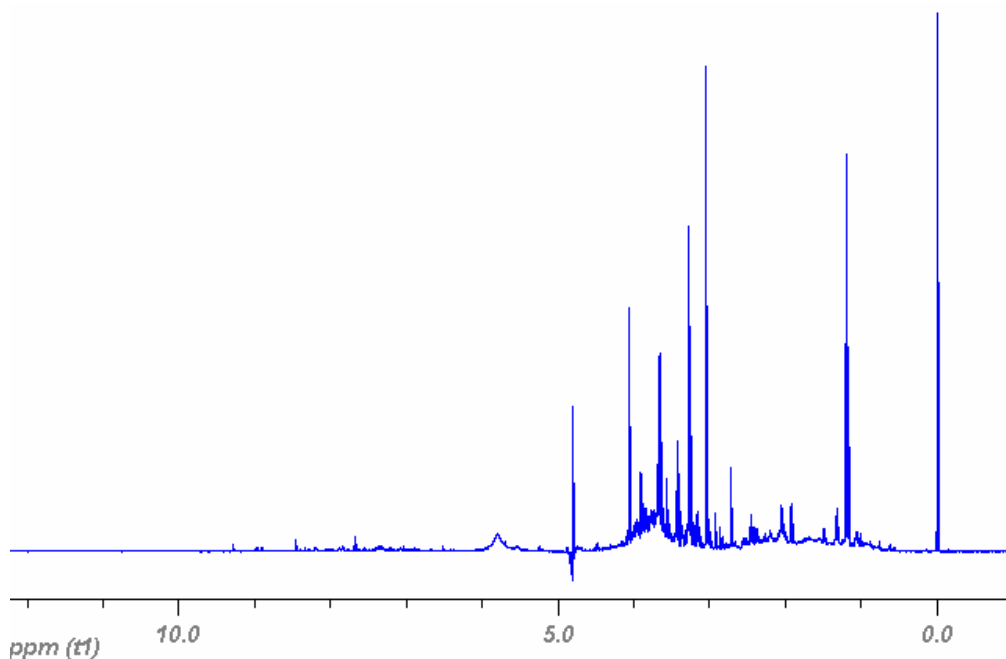
I campioni di urine sono stati centrifugati a 10000 rpm per 1 minuto, allo scopo di rimuovere componenti solidi dal campione, e sono stati prelevati 400 uL di surnatante; la stabilizzazione del pH è stata ottenuta aggiungendo al campione 200 ul di tampone fosfato in  $H_2O$  0,2M con  $pH=7,45$ .

Infine al campione così preparato è stato aggiunto uno standard interno, 50ul di TSP (3-(Trimethylsilyl)- Propionic acid-D4, sodium salt) ad una concentrazione di 10.3mM in  $D_2O$ , per una concentrazione finale nel campione di circa 0.8mM.



**Figura 30.** Standard di riferimento TSP usato per gli esperimenti NMR protonici sulle urine.

Le urine trasportano un elevato contenuto informativo sugli stati metabolici; data l'origine e la dislocazione sistemica di questo biofluido si prestano molto bene agli studi tossicologici e di patologie del metabolismo in generale (Diabete). Nella Figura 31 è rappresentato lo spettro NMR protonico di un campione di urine di neonato pretermine.



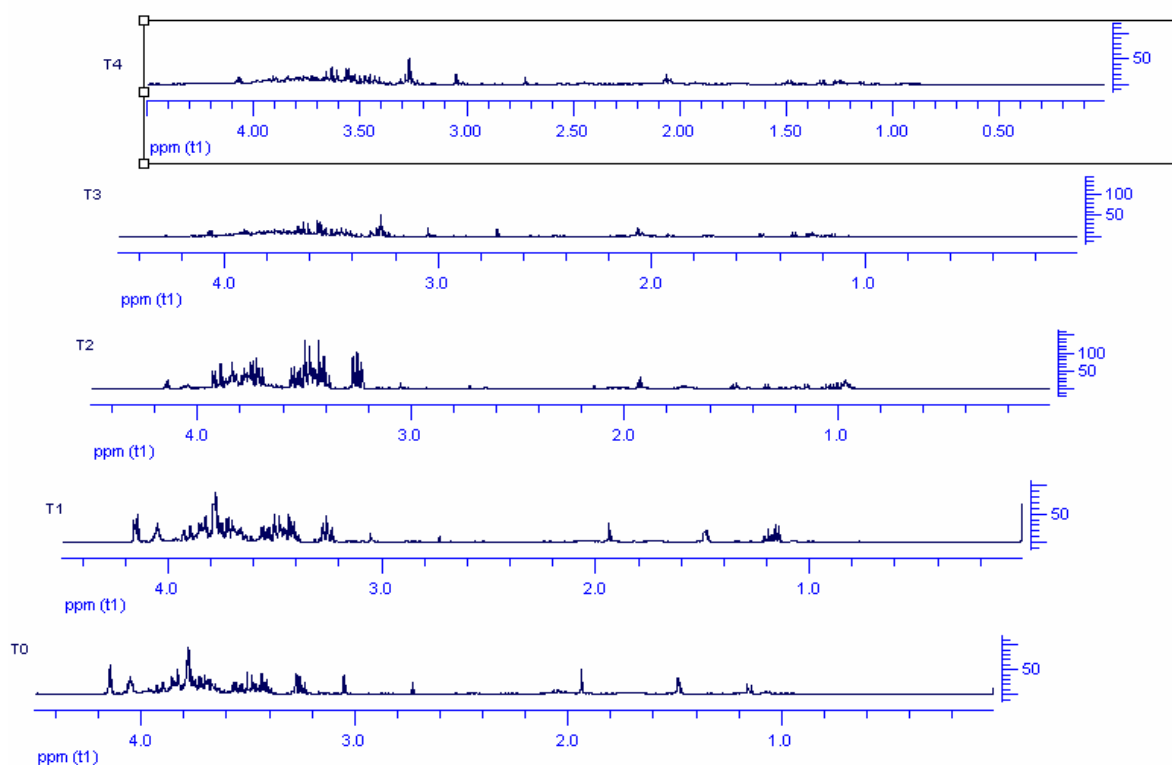
**Figura 31.** Spettro NMR protonico di urine di neonati pretermine.

Anche in questo caso risulta indispensabile programmare strategie di calcolo ed analisi di tipo metabonomico. Nel caso di campioni prelevati ad intervalli regolari per lo studio dell'evoluzione globale del metabolismo del paziente in funzione del tempo è poco produttivo analizzare metabolita per metabolita le variazioni di concentrazione in funzione del tempo, ma è più proficuo realizzare, ad esempio, un studio sulla dinamica degli stati metabolici costruiti a partire da vettori di stato che siano, come nel caso presentato in questo lavoro, una espressione delle proprietà di insieme del sistema biologico in studio. Una



tecnica di calcolo utile per ricavare queste proprietà di insieme è ancora l'analisi delle componenti principali.

Nella Figura 32 sono riportati gli spettri NMR protonici a tempi diversi acquisiti dalle urine di un neonato con disfunzione renale a cui è stata applicata una terapia protettiva, a base di Glazidim ed Idroplurivit, somministrata a tempi diversi (lo spettro iniziale  $T_0$  è quello in basso).



**Figura 32. Analisi temporale delle urine di un neonato “pretermine” patologico.**

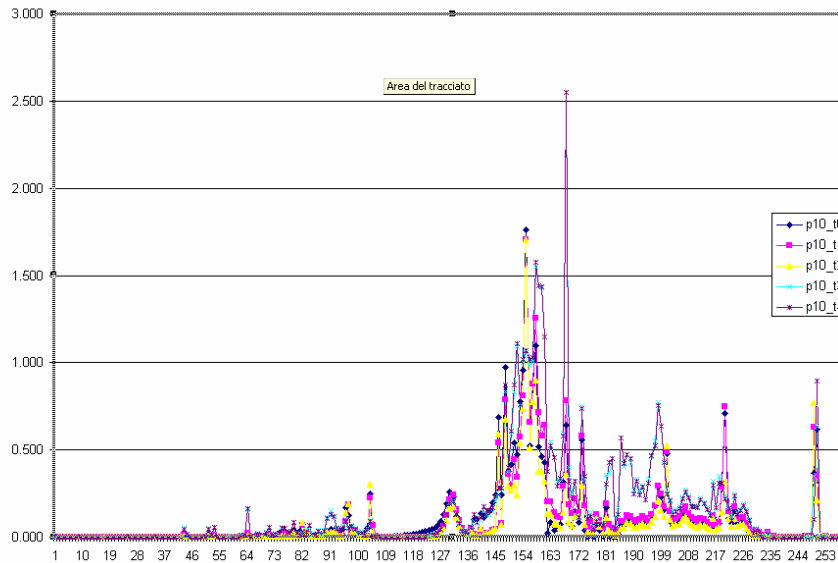
Di questi spettri sono state selezionate per l'analisi le seguenti regioni:

$r1=11-5,5$  ppm

$r2=4,5 -0,5$ ppm

Per la valutazione delle aree integrate dei segnali è stata applicata di nuovo la tecnica di integrazione a bucket su bin di ampiezza di base di 0,04 ppm e si produce quindi la riduzione della dimensionalità dei descrittori degli spettri che vengono rappresentati, in questo caso, da 261 valori.

I cinque spettri per singolo paziente ottenuti dopo queste procedure sono stati sovrapposti per una prima analisi “visiva” delle alterazioni.

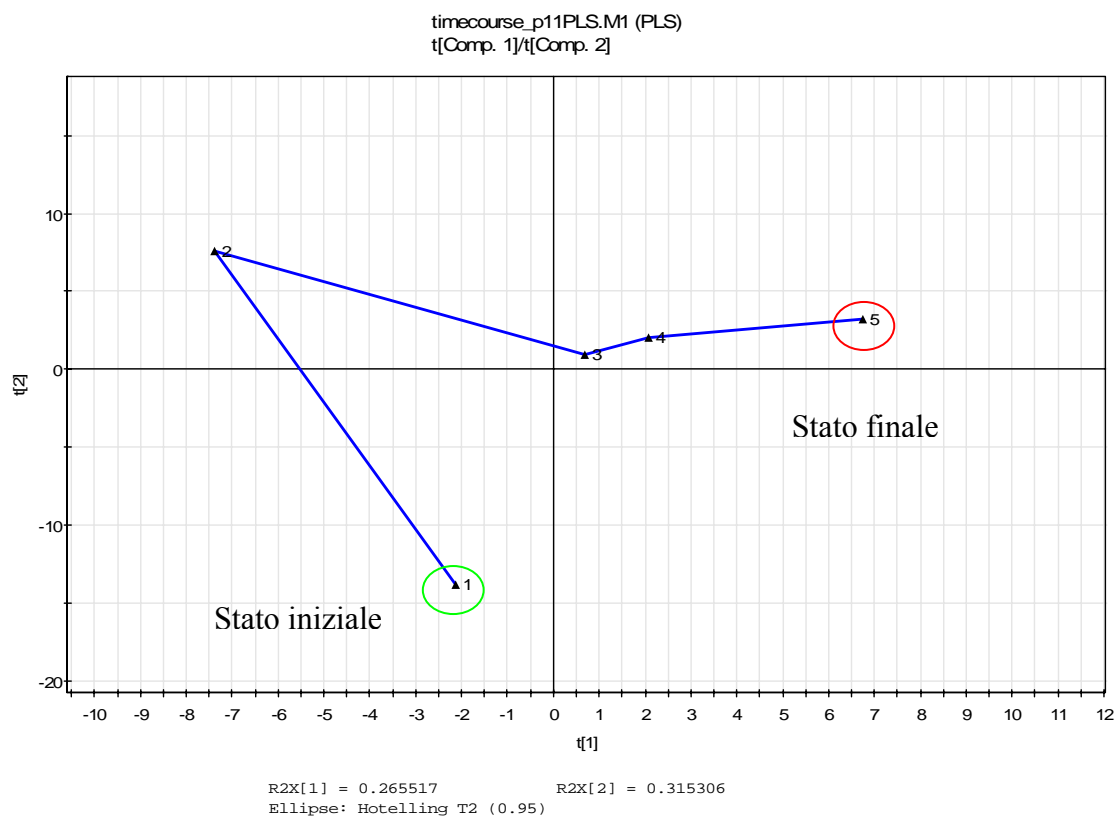


**Figura 33. Confronto delle aree dei segnali NMR a tempi diversi, per un singolo paziente.**

Il grafico delle aree dei bins, normalizzate al picco del TSP, mostra per il campione P10 l'aumento marcato dei metaboliti in diverse regioni spettrali. Si vede chiaramente dalla sovrapposizione diretta degli spettri una evoluzione temporale del profilo metabolico, ma collegare questa dinamica a singoli metaboliti e cicli vitali non fornisce informazioni utili per la razionalizzazione del problema. Occorre l'approccio metabonomico.

L'evoluzione temporale del metabolismo del singolo soggetto si può studiare attraverso le tecniche di calcolo della dinamica delle serie temporali adattate alla metabonomica; questa dinamica è rappresentata, nel nostro caso, da 5 punti (prelievi a 5 periodi diversi in genere distanziati di una settimana) che rappresentano 5 stati metabolici differenti nello spazio metabonomico n-dimensionale definito dall'analisi PCA dei campioni acquisiti (e confrontati con una opportuna base dati di patologici e di controlli). La tecnica permette di evidenziare l'evoluzione temporale del metabolismo del soggetto in modo comparativo rispetto agli altri patologici e rispetto ai controlli selezionati. Questi stati hanno una dipendenza esplicita dal tempo caratteristica non solo dello stato metabolico patologico di base del paziente ma anche della risposta metabolica al trattamento farmacologico a cui è sottoposto.

Come tutte le serie temporali, le serie temporali metaboliche possono essere caratterizzate mediante tecniche di dinamica non lineare. Con queste tecniche è possibile evidenziare la natura complessa e non lineare dei sistemi biologici [4, 5]; la facilità e la ripetibilità dei prelievi di urine senza danni al paziente permettono di operare studi di questo tipo, al contrario di quanto succede per il liquor in neurologia. Nella Figura 34 viene rappresentata una delle traiettorie metaboliche acquisite in maniera comparata per uno dei casi patologici.



**Figura 34.** Traiettoria metabolica di un paziente patologico durante trattamento farmacologico.

Il confronto d'insieme di queste “traiettorie metaboliche”, operato su una ampio campione statistico, può fornire le caratteristiche dei fenotipi patologici anche in relazione ai trattamenti farmacologici.

Oltre alle dinamiche dei singoli soggetti si stanno studiando anche i differenti profili metabolici relativi agli stati patologici iniziali, alla nascita, stati puramente fisiologici, in cui non sono ancora stati forniti trattamenti ai piccoli pazienti.

L'analisi PLS-DA eseguita sui primi campioni acquisiti separa bene controlli e patologici ed individua anche due “outliers” del modello sulla cui natura non è ancora possibile fare considerazioni in virtù del limitato campione statistico.

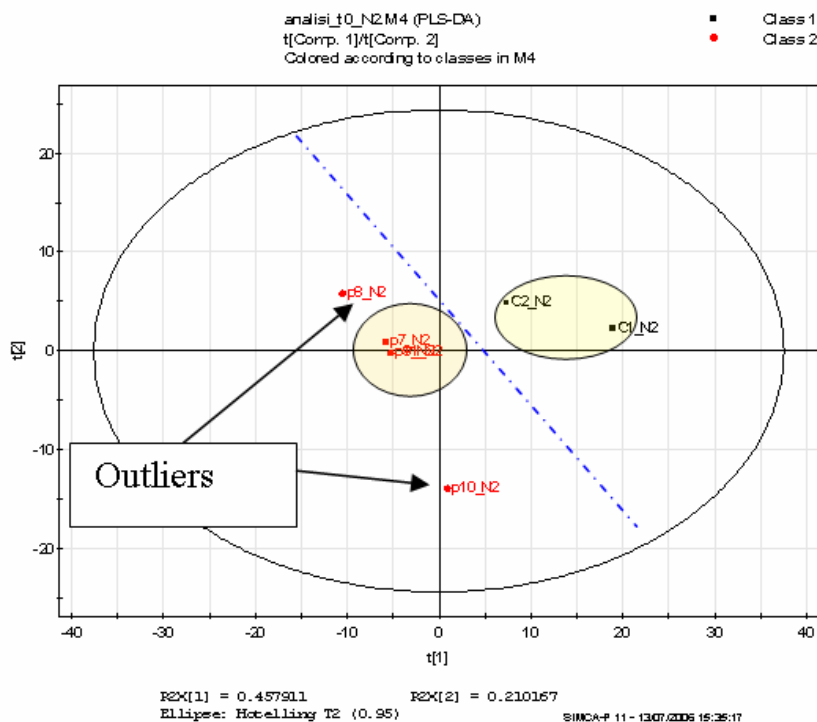


Figura 35. PCA sui dati di spettroscopia protonica di urine di neonati pretermine.

I loadings dell'analisi PCA riportata in Figura 35 e riportati in Tabella 8 indicano gli intervalli di Chemical Shift individuati, finora, come responsabili di queste differenze.

Tabella 8. Metaboliti responsabili delle differenti dinamiche metaboliche di neonati pretermine patologici.

Regione spettrale	Metaboliti
0,84-0,92	Alfa-OH-isovalerato Beta-OH-butirrato
1,24-1,32; 1,36-1,40	Lactate
2,04-2,36	Acetone, Acetoacetato, beta-d-idrossibutirrato, Glutamina
2,40-2,44	Succinato, beta-d-idrossibutirrato
2,56-2,64	Citrato
2,76-2,80	Citrato, DMA (DIMETILAMMINA)
2,88-2,92	DMG (DIMETILGLICINA)
3,04-3,08	Creatina, Creatinina
3,16-3,20	Creatinina
5,64-5,80	Urea

In giallo sono evidenziati due componenti metaboliche di interesse per le patologie renali di cui soffrono i piccoli pazienti.

Questi dati sono stati utilizzati per addestrare un classificatore SVM per riconoscere i profili metabolici patologici di campioni di urine incogniti dei neonati. Nella tabella seguente si riporta la classificazione ottenuta su tutti i campioni a disposizione che evidenzia la maggiore similarità degli spettri di tutti i pazienti analizzati finora con lo stato patologico piuttosto che con lo stato di controllo sano.

**Tabella 9. Classificazione SVM degli stati metabolici di neonati pretermine.**

Campioni da classificare	SVM Classificazione	Training set usato 5 spettri t <sub>0</sub> : C1_t0, C2_t0, P8_t0, P9_t0
'p11_t0'	0	
'p11_t1'	0	
'p11_t2'	0	
'p11_t3'	0	
'p11_t4'	0	
'p10_t0'	0	
'p10_t1'	0	
'p10_t2'	0	
'p10_t3'	0	
'p10_t4'	0	
'p9_t0'	0	
'p9_t1'	0	
'p9_t2'	0	
'p8_t1'	0	
'p8_t2'	0	
'p7_t1'	0	
'p7_t2'	0	
0=patologico 1=controllo sano Legenda esempio campioni p10_t0 = prelievo a tempo t= 0 per il campione patologico 10		

Il dato presentato è molto interessante. I pazienti sono sottoposti a trattamento farmacologico e quello che si vede da questa analisi è che nessuno degli stati metabolici dinamici, anche dopo il trattamento, si caratterizza come stato non

patologico nonostante si riscontrino sul piano clinico, per alcuni dei piccoli pazienti, miglioramenti generalizzati.

Ulteriori considerazioni al momento non sono possibili perché non disponiamo di dati sufficienti sui diversi trattamenti farmacologici adottati ma sarà importante, da questi dati, valutare le correlazioni metaboliche che individuano la popolazione di neonati che maggiormente soffre lo specifico trattamento. Data la disponibilità di diversi piani di trattamento per queste patologie risulta importante valutare a priori le classi di pazienti e le loro risposte caratteristiche ai trattamenti per valutare anticipatamente benefici ed effetti collaterali per ottimizzare la somministrazione dei farmaci ed evitare al massimo le complicanze.

## **5. Conclusioni e sviluppi futuri.**

Sulla base dei risultati acquisiti in questo lavoro di tesi sia per il lavoro sul liquor e la SM sia per quello sulle urine dei neonati pretermine è possibile fare alcune considerazioni conclusive.

I dati NMR di alcuni degli spettri acquisiti per lo studio della SM presentano un artefatto dovuto a degradazione del campione di liquido cerebrospinale indotta probabilmente dalla conservazione di alcuni degli spettri a  $-20^{\circ}\text{C}$  per un periodo troppo lungo e dall'incompatibilità di questa condizione con l'esigenza di preservare la natura metabolica del campione; le aree delle risonanze di alcuni metaboliti, quali ad esempio il piruvato e la glutammina, sono alterate per effetto della degradazione spontanea o per attività enzimatica nel CSF. Tali alterazioni inducono un artefatto che maschera i reali stati metabolici della patologia. Operando attraverso un filtro sui dati NMR è stata eliminata la componente da artefatto e la successiva analisi delle componenti principali rivela una significativa separazione dei campioni di controllo, indicati con CN, controlli neurologici, dai patologici SM\_SM e CIS\_SM e cioè i patologici definiti fin dalla prima diagnosi come sclerosi multiple per il grave quadro clinico ed i patologici definiti inizialmente CIS, sindrome demielinizzante isolata ma diventate nel tempo sclerosi multiple conclamate. Il modello ricavato permette di classificare i due campioni incogniti a disposizione nel database acquisito definiti CIS\_CIS, sindromi clinicamente isolate stabili, i quali si dispongono diversamente: uno nel cluster dei controlli e l'altro in prossimità del cluster dei patologici. Questo modello fornisce un primo strumento predittivo del possibile stato di evoluzione della patologia che esordisce come episodio clinicamente isolato, appunto CIS. Il follow-up dei due pazienti permetterà di valutare ed eventualmente calibrare il modello predittivo.

Altro risultato ottenuto è quello di una prima caratterizzazione relativa degli stati patologici SM rispetto ai controlli. La disposizione dei campioni patologici risulta avere una dispersione maggiore e nell'analisi a cluster di questi dati è

possibile individuare una sotto partizione in due classi con caratteristiche differenti, possibile indice di una “dinamica” degli stati patologici di SM che avviene attorno ad almeno due stati stazionari su cui sarà opportuno investigare in modo più approfondito in futuro.

C’è da dire che la limitatezza del campione statistico non ha permesso di effettuare quelle suddivisioni nei campioni in generale fortemente consigliate nell’analisi metabonomica volta alla rilevazione di particolari fenotipi delle patologie (separazione dei campioni provenienti da maschi da quelli di femmine, separazioni in fasce di età etc.). Questo potrebbe produrre altri effetti di mascheramento dei dati.

Sarà bene effettuare tali suddivisioni appena le rispettive popolazioni dei campioni lo consentiranno. La numerosità statistica delle classi non è stimabile a priori. Ma è certo che maggiore è il numero di campioni appartenenti alla stessa popolazione maggiori saranno le possibilità di definire descrittori di tali classi ad alta specificità per la patologia in studio.

Il set di metaboliti che definisce il modello che differenzia i cluster di spettri analizzati è riportato di seguito e definisce il pattern di riferimento, il profilo metabolico, per la differenziazione tra sani e patologici per SM che abbiamo ricavato dai nostri esperimenti:

**Tabella 10. Profilo metabolico per la differenziazione controlli patologici non SM e patologici SM.**

<b>Regione spettrale [ppm]</b>	<b>Metaboliti di riferimento della regione spettrale</b>
0,88-0,92	Acido 2-idrossisovalerico, Isoleucina
1,20-1,24	Acido 3-idrossibutirico
1,32-1,40	A. Lattico A. 2-idrossibutirico
1,92-2,24	N-acetil-tirosina, Glutamina, Acetone
2,42-2,56	Acido 4-idrossibutirico
3,04-3,08	Creatina
3,24-3,32	Mio-Inositolo, Glucosio



Come detto uno dei risultati più importanti ottenuti da questo modello a livello speculativo riguarda il lattato ed i metaboliti coinvolti nei cicli metabolici correlati.

Il lattato, prodotto dal metabolismo ossidativo anaerobico, è un metabolita fondamentale nella caratterizzazione delle malattie mitocondriali; questo lavoro getta le basi biochimiche per le analisi volte all'integrazione dei dati genetici e funzionali dello stato dei mitocondri cellulari nella SM. L'alterazione del metabolismo ossidativo ricollegabile al lattato ed ai metaboliti ad esso correlato e rilevati nell'analisi può avere un sufficiente grado di specificità e potrebbe rendere conto dell'insorgenza o dell'evoluzione di malattia di sclerosi multipla.

Il modello proposto in questo lavoro ha sicuramente dei limiti. Ad esempio non si trova traccia delle alterazioni metaboliche dei cicli in cui sono coinvolte la colina e la glicina, il cui contributo alla caratterizzazione della SM è stato proposto da diversi lavori in letteratura (10, 14). I motivi possono essere molti: innanzitutto i range di chemical shift considerati in questa fase preliminare del lavoro escludono la regione dove si trovano i segnali della glicina. In uno dei modelli iniziali il contributo della colina a 3,19 ppm viene evidenziato ma con un basso peso statistico. Questo può derivare sia dal fatto che la metabonomica evidenzia i contributi covarianti dei metaboliti e quindi mancando una regione di chemical shift potenzialmente importante potrebbe essersi persa la correlazione colina-glicina. L'altro motivo potrebbe essere l'effetto del filtro adottato. L'algoritmo usato considera una funzione di riferimento costruita come combinazione lineare di soli tre parametri clinici forniti. La costruzione di questa funzione passa attraverso ipotesi conservative preliminari e potrebbero essere importanti anche altri parametri per ottimizzare l'azione del filtro. Del resto i valori di  $R^2X$ , la significatività statistica dei dati ottenuti, sono relativamente bassi ed i diversi modelli aumentano e di poco il contributo della seconda componente trovata. La conclusione è che sarebbe meglio trattare dati senza artefatti; questo eliminerebbe l'obbligo del filtraggio ed i dati potrebbero fornire altre informazioni. Lo standard proposto per il futuro prevede che i

campioni da analizzare siano prelevati ed analizzati contestualmente limitando i periodi di conservazione a  $-20^{\circ}\text{C}$  ai soli periodi di trasporto e pretrattamento. Il nostro lavoro ha mostrato la fattibilità di studi su campioni conservati per lunghi periodi ma ha anche evidenziato i possibili limiti di questo tipo di studio. Certamente la conservazione a  $-80^{\circ}\text{C}$  adottata di recente come standard nei laboratori dei Centri SM aumenterà l'importanza di questo tipo di studi.

Il modello che è stato realizzato ha prevalentemente una valenza teorica e qualitativa, dato che devono essere ancora fatte le opportune calibrazioni attraverso un incremento delle popolazioni statistiche delle varie classi patologiche riscontrate. Si è comunque ritenuta importante nell'ambito di questo lavoro di tesi la sua presentazione come impostazione iniziale della risoluzione di un problema complesso. Si ritiene inoltre che un approccio simile possa diventare nel futuro uno strumento semplice, di facile impiego, per lo screening degli spettri di campioni già acquisiti o dei composti che devono essere ancora raccolti.

Con le stesse metodiche si stanno valutando anche le potenzialità diagnostiche dell'analisi metabonomica delle urine di neonati pretermine con insufficienza renale (Cap. 4) e di pazienti a rischio di tumore alla mammella (Si consideri la possibilità di effettuare lo screening di questa patologia attraverso esami NMR su campioni di urine con una sensibilità e specificità simili a quelli degli usuali esami radiologici permetterebbe una sensibile riduzione della dose di radiazione alla popolazione, con enormi benefici esprimibili anche in termini di costi sociali. Anche in questo caso solo i metodi di analisi metabonomici permettono di realizzare un quadro metabolico utile per una discriminazione precoce).

Nel caso dei neonati pretermine stiamo cominciando a costruire le banche di dati per l'addestramento di classificatori automatici degli stati metabolici alla nascita per individuare quei pazienti potenzialmente a rischio di intossicazione da trattamento farmacologico. Anche in questo caso possediamo una prima mappatura metabolica che differenzia gli stati iniziali patologici da quelli di

controllo:

**Tabella 11. Mappatura metabolica iniziale per la differenziazione degli stati patologici in neonati pretermine.**

Regione spettrale	Metaboliti
0,84-0,92	Alfa-OH-isovalerato; Beta-OH-butirrato
1,24-1,32; 1,36-1,40	Acido Lattico
2,04-2,36	Acetone, Acetoacetato, beta-idrossi-butirrato, Glutamina
2,40-2,44	Succinato, beta-idrossi-butirrato
2,56-2,64	Citrato
2,76-2,80	Citrato, DMA (DIMETILAMMINA)
2,88-2,92	DMG (DIMETILGLICINA)
3,04-3,08; 3,16-3,20	Creatina, Creatinina
5,64-5,80	Urea

Proseguendo nell'acquisizione di campioni di urine si potrà caratterizzare al meglio tutti i possibili stati metabolici e derivare preliminarmente i probabili effetti dei piani di trattamento terapeutico.

Nell'immediato futuro il gruppo di lavoro sta pianificando di sviluppare le basi di dati relativi alle patologie in studio e presentati in questo lavoro; inoltre stiamo mettendo a punto anche altre metodiche di analisi quali la spettrometria di massa e l'analisi HPLC.

La metabonomica comincia ad avere un ruolo molto importante in campo medico; questa metodica si affianca certamente alla analisi genica e proteomica per fornire quella completezza del quadro biologico di un paziente che è la base per l'ottimizzazione della terapia. Grazie a questa disciplina i metodi chemiometrici assumono oggi un valore specifico anche in campo medico applicato producendo un elevato beneficio sociale oltre che importanti passi in avanti nella ricerca scientifica. È facile intuire le potenziali applicazioni di un sistema completo di caratterizzazione metabolica e le ricadute che queste linee di ricerca possono avere in termini di sviluppo e benessere sociale. In un periodo in cui si sperimentano sempre più gli effetti dei cambiamenti ambientali sulla vita e sulla salute si capisce quanto sia importante disporre di metodiche di

indagine scientifica che partendo dalla cellula e dal suo contenuto informativo essenziale, il DNA, sappia caratterizzare i sistemi biologici nella loro complessità e nella loro risposta a malattie ed influenze ambientali.

## **Glossario**

**Atassia cerebellare:** interessa una parte dell'encefalo chiamata cervelletto, deputata a regolare la giusta distribuzione del tono muscolare tra i gruppi di muscoli interessati nel mantenere l'equilibrio o nel determinare un movimento; l'atassia c. inizia in modo più o meno improvviso con la comparsa di disturbi dell'equilibrio e della coordinazione motoria, perdita nella misura dei movimenti e deambulazione a gambe divaricate con ampie oscillazioni del tronco e con tendenza a cadere all'indietro o di lato. Questa deambulazione viene appunto definita atassica.

**Astrociti.** Il tipo più abbondante di cellule della glia, gli astrociti sono forniti di numerose estroflessioni che ancorano i neuroni al loro rifornimento di sangue. Regolano l'ambiente chimico esterno dei neuroni rimuovendo gli ioni, in particolare il potassio, e riciclano i neurotrasmettitori rilasciati durante la trasmissione sinaptica. La teoria corrente sostiene che gli astrociti siano i "blocchi di costruzione" della barriera emato-encefalica. Gli astrociti dovrebbero essere inoltre in grado di regolare la vasocostrizione e la vasodilatazione producendo sostanze come l'acido arachidonico i cui metaboliti sono vasoattivi. È noto che gli astrociti comunicano uno con l'altro usando il calcio e possiedono riserve di glicogeno.

**Assoni.** prolungamenti neuronali attraverso i quali viaggiano correnti bioelettriche in senso centrifugo. Normalmente, gli unici punti di dispersione del segnale sono costituite dai nodi di Ranvier, soluzioni di continuo poste fra la guaina mielinica di un'oligodendrocita e quella adiacente.

**Bande oligoclonali.** Sono un'espressione dell'infiammazione cerebrale e sono immunoglobuline o anticorpi prodotte dai linfociti nei focolai infiammatori.

**Mielina.** specializzazione della membrana plasmatica degli oligodendrociti avvolta in spire concentriche attorno agli assoni. Forma una guaina isolante che si comporta come un circuito RC altamente performante, in grado di ottimizzare la trasmissione degli impulsi elettrici a basso voltaggio attraverso la rete neuronale, minimizzando le dispersioni.

**Liquor Cerebrospinale.** liquido con caratteristiche osmolari analoghe alla componente liquida del sangue, circolante in spazi compresi fra il rivestimento meningeo più intimo (pia madre) e l'aracnoide. Forma una falda liquida che protegge le strutture molli (cervello e midollo) dalle sollecitazioni meccaniche derivanti dalle strutture inestensibili della teca cranica e del canale rachideo. Va incontro ad equilibrati processi di produzione e riassorbimento da parte di cellule specializzate, la cui perfetta funzionalità è determinante per il mantenimento di una composizione costante. Eventuali scostamenti dalla normale composizione del liquor sono una spia importante di varie patologie a carico del SNC, indagata con un prelievo mirato (rachicentesi). Nella SM, l'evidenza di particolari bande elettroforetiche oligoclonali è segno considerato predittivo dello stato di malattia, e costituisce un riconosciuto criterio diagnostico.

**Glia.** Le cellule della glia sono cellule che, assieme ai neuroni, costituiscono il sistema nervoso. Hanno funzione nutritiva e di sostegno per i neuroni, assicurano l'isolamento dei tessuti nervosi e la protezione da corpi estranei in caso di lesioni. Tra i tipi di cellule della glia più importanti si trovano gli astrociti, forniti di numerose estroflessioni che

ancorano i neuroni al loro rifornimento di sangue, e gli oligodendrociti che hanno il compito di rivestire gli assoni del sistema nervoso centrale con una sostanza grassa chiamata mielina producendo la cosiddetta guaina mielinica. La guaina isola l'assone permettendo quindi una migliore propagazione dei segnali elettrici.

**Indice EDSS.** Indice espresso su una scala numerica tra zero e nove che indica il livello di disabilità del paziente di SM e quindi il relativo danno del sistema motorio, circolatorio e nel cognitivo.

**Indice di Link.** Deriva dalle concentrazioni di albumina (la maggiore proteina della parte non cellulare del sangue) e di immunoglobuline sia nel sangue sia nel liquor.

**Mitocondri.** Sono organi microscopici presenti in grandi quantità all'interno delle cellule. Sono considerati le centrali energetiche degli organismi, infatti al loro interno avvengono quei processi biochimici (respirazione mitocondriale) che forniscono alle cellule l'energia di cui hanno bisogno per tutte le loro funzioni vitali. Il più importante tra questi processi è la fosforilazione ossidativa.

**Neuroni.** cellule "nobili" del SNC, fittamente connesse fra di loro a formare una rete neurale, complessivamente sede dell'attività nervosa. I neuroni comunicano fra loro attraverso una serie di impulsi bioelettrici a basso voltaggio, capaci di eccitare o inibire altre popolazioni neuronali sortendo effetti complessi. Nella parte terminale dei loro assoni, a livello dei prolungamenti sinaptici, la trasmissione diventa neurochimica, mediata dalla sintesi e dal rilascio controllato di pacchetti quantici di neurotrasmettitori nell'interfaccia che divide una cellula dall'altra.

**Nuclear Magnetic Resonance NMR.** Fenomeno fisico in cui è possibile scambiare energia nella condizione di risonanza con un sistema di spin nucleari immersi in un campo magnetico statico. Questo fenomeno ha permesso lo sviluppo di metodi di indagine molecolare e di diagnostica per immagine molto importanti nella medicina moderna.

**Principal Component Analysis PCA.** Analisi statistica volta ad individuare precise combinazioni lineari dei dati che possono individuare un ridotto numero di variabili per descrivere sistemi complessi. Si basa sull'applicazione di trasformazioni geometriche di corpo rigido operate sulle strutture dei dati che derivano da operazioni di diagonalizzazione della matrice originaria.

**Pattern Recognition PR.** Tecniche di analisi dei dati volte alla identificazione e determinazione di profili specifici o modelli o leggi di similarità presenti nelle matrici di dati multivarianti.

**Sostanza bianca.** Termine di anatomia macroscopica con cui si designa l'aspetto dei tessuti encefalici mielinizzati, opposti alla sostanza grigia, tipica della corteccia encefalica. Nel contesto della sostanza bianca corrono le vie di trasmissione che connettono la corteccia (sede dei corpi neuronali organizzati in aree e deputati alla elaborazione cosciente e alle funzioni superiori) con i nuclei di relè sottostanti.

**Sclerosi Multipla (SM).** malattia cronica degenerativa del SNC, caratterizzata da un quadro clinico proteiforme e variabile, con un'evoluzione difficilmente prevedibile, che esita in uno stato di disabilità di grado severo. La lesione caratteristica della malattia è la placca, ovvero



un focus di demielinizzazione della sostanza bianca attribuibile ad un processo di autoaggressione da parte del sistema immunitario.

**Spettroscopia di Risonanza Magnetica MRS.** spettroscopia di assorbimento in Radiofrequenza che permette studi delle caratteristiche dei sistemi attraverso la risonanza magnetica

### ***Sigle usate nella Tesi.***

**CIS.** Sindrome infiammatoria clinicamente isolata. È l'esordio della malattia demielinizzante in cui i parametri clinici ancora non consentono una diagnosi differenziata.

**CI.** Sigla utilizzata in questo lavoro per definire controllo neurologico da malattia di natura infiammatoria

**CIN.** Sigla utilizzata in questo lavoro per definire controllo neurologico da malattia di natura non infiammatoria

**CN\_EPI.** Sigla utilizzata in questo lavoro per definire controllo neurologico da malattia di epilessia farmaco-resistente.

**CIS\_SM.** Sigla utilizzata in questo lavoro per definire i casi di esordio di malattia demielinizzante a singolo episodio che è evoluta in SM.

**SM\_SM.** Sigla utilizzata in questo lavoro per indicare i casi in cui la malattia SM è esordita subito con i caratteri della patologia grave ed ha mantenuto tale carattere.

## **Appendice 1. Routines per l'ambiente di calcolo Matlab**

1) Gli input di questa routine sono il file ascii contenente i dati, il numero di bin a cui si vuole ridurre lo spettro ed il numero di campioni contenuti nello spettro. La procedura restituisce un vettore contenente le aree integrate sui bins o buckets.

```
%function binning
%clear;
%load % fare il load del WS prima di lanciare la procedura
file=input('file ');
h=input('Numero dei Bin ');
l=input('lunghezza del file ');
a=zeros(1,h);
c=1; b=fix((l/h));
a(1)=sum(file(c:b));
if a(1)<0
    a(1)=0;
end
for n=1:(h-1)
    c=1+(n*b);
    x=(n+1)*b;
    a(n+1)=sum(file(c:x));
    if a(n+1)<0
        a(n+1)=0;
    end
end
end
```

2) Gli input di questa routine sono il file ascii contenente i dati per l'addestramento, le caratteristiche ed i dati da classificare.

```
%function svmclassify%
%clear;
%load % fare il load del WS prima di lanciare la procedura
tr=input('tr '); %carica il vettore di training%
gr=input('gr '); %carica il vettore delle caratteristiche%
data=('data'); %carica il vettore di dati da classificare%
svmstruc=svmtrain(tr,gr); %si crea la struttura di addestramento svmstruct%
classes = svmclassify(svmstruc,data); %si usa svmstruct per classificare il vettore data%
end
```

## **Bibliografia**

1. Beale et al. "Plant Metabolomics by NMR spectroscopy". Atti del Congresso First Maga Circe Conference On Metabolic Systems Analysis, Sabaudia(La) 2006.
2. Lindon J. "Metabolomics technologies and pharmaceutical" Atti del Congresso First Maga Circe Conference On Metabolic Systems Analysis, Sabaudia(La) 2006.
3. Benno H. et al. "Transcriptome meets metabolome: hierarchical and metabolic regulation of the glycolytic pathway". FEBS Letters 500 (2001) 169-171.
4. Arecchi F.T. "Caos e complessità nel vivente" Corso SUS 2003
5. Giuliani A., Zbilut J, Conti F., Manetti C., Miccheli A. "Invariant features of metabolic networks: a data analysis application on scaling properties of biochemical pathways". *Physica A* 337 (2004) 157 – 170.
6. Marrosu M.G. et al, "Human Molecular Genetics, 2001, Vol. 10, No. 252907-2916-Oxford University Press
7. McDonald WI, Compston A, Edan G, et al. "Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the Diagnosis of Multiple Sclerosis". *Ann Neurol* 2001;50(1):121-7.
8. J.K. Nicholson, I.D. Wilson, "High resolution proton magnetic resonance spectroscopy of biological fluids" *Progress in Nuclear Magnetic Resonance Spectroscopy*, Vol. 21, Pergamon, Oxford, 1989, pp. 449–501.
9. Burlin A., Aureli T., Bracco F., Conti F., Battistin L. "MR Spectroscopy: a powerfull tool for Investigating Brain Function

- and Neurological Disease”. *Neurochemical Research* Vol. 25 Nov. 9/10. 2000. 1365-1372.
10. Simone I. et al. “High resolution proton MR spectroscopy of cerebrospinal fluid in MS patients. Comparison of biochemical changes in demyelinating plaques.” *Journal of neurological Sciences* 144 (1996) 182-190.
  11. Maillet J. et al, “Experimental protocol for clinical analysis of cerebrospinal fluid by high resolution proton magnetic resonance spectroscopy”, *Brain Research Protocols* 3 (1998 )123–134.
  12. Puligheddu M., Chessa L., Balestrieri A., Lisai M., Barberini L., Marrosu F. “Orthotopic liver transplantation In Neurological Wilson disease”. *Atti del Congresso Nazionale Società Italiana di Neurologia* 2006, Bari.
  13. European Congress of Radiology ECR, Vienna 4-8 March 2005. Role of MR spectroscopy in patients with temporal lobe epilepsy (TLE). A. Sias, A. Balestrieri, L. Barberini, M. Puligheddu, M. Fraschini, F. Marrosu, G. Mallarini
  14. J. Lynch et al, “Nuclear Magnetic Resonance Study of Cerebrospinal Fluid From Patients With Multiple Sclerosis”. 1993 *The Canadian Journal of Neurological Sciences*.
  15. J.C. Lindon et al “Patterns Recognition methods in biomedical magnetic resonance”, *Progress in Magnetic Resonance Spectroscopy* 39 2001.
  16. Griffin J. “Metabonomic: NMR spectroscopy and pattern recognition analysis of body fluids and tissues for characterization of xenobiotic toxicity and disease diagnosis” – *Current opinion in Chemical Biology* 2003, (7:648-654).

17. el-Deredy W. "Pattern recognition approaches in biomedical and clinical magnetic resonance spectroscopy: a review.". *NMR Biomed.* 1997 May; 10(3):99-124.
18. Mitchell T.M. "Machine learning", McGraw-Hill Companies Inc. 1997.
19. Spraul M. et al. "Automatic reduction of NMR spectroscopic data for statistical and pattern recognition classification of samples". *J. Pharm. Biomed. Anal.* (1994) 12, 1215–1225.
20. Flore M.A. et al. "Early Discrimination Of Clinically Isolated Syndrome(CIS) Individuals At High Risk For Multiple Sclerosis By 1H-MRS Based CSF Analysis: Preliminary Results". *Atti del Congresso Nazionale Società Italiana di Neurologia 2006, Bari.*
21. Wold S, Antti H, Lingren F, Ohman J: "Orthogonal signal correction of near-infrared spectra".
22. Messana I. et al, "Proton nuclear magnetic resonance spectral profiles of urine in type II diabetic patients". *Clinical Chemistry* 44:7 1529–1534 (1998).

### **Siti web di interesse per la Metabonomica**

Il sito dell'Associazione Italiana di Metabonomica, ASSIMET

<http://www.assimet.it>

Il sito web di Consortium for METabonomic Toxicology, COMET

<http://bc-comet.sk.med.ic.ac.uk/>

Il sito web di Standard Metabolic Reporting Structure, SMRS

<http://www.smrsgroup.org/>

Il sito del progetto metaboloma umano per una lista molto estesa dei metaboliti e delle loro proprietà chimico-fisiche

<http://www.metabolomics.ca/>

## ***Indice delle Tabelle***

Tabella 1. Assegnazione delle risonanza 1H- NMR del CSF.....	32
Tabella 2. Elenco di alcuni dei metaboliti di importanza diagnostica in SM.....	43
Tabella 3. Comparazione dei livelli metabolici in SM precoci e diagnosticate da molto tempo. .....	44
Tabella 4. “Clusterizzazione” degli spettri NMR dopo PCA. ....	60
Tabella 5. Matrice de metaboliti responsabili della separazione controlli-patologici SM acquisita con analisi PLS-DA.....	64
Tabella 6. Matrice dei loadings del modello M2.....	66
Tabella 7. Intervalli di chemical shift e metaboliti di riferimento per il modello della SM.....	68
Tabella 8. Metaboliti responsabili delle differenti dinamiche metaboliche di neonati pretermine patologici.....	76
Tabella 9. Classificazione SVM degli stati metabolici di neonati pretermine. ....	77
Tabella 10. Profilo metabolico per la differenziazione controlli patologici non SM e patologici SM. ....	80
Tabella 11. Mappatura metabolica iniziale per la differenziazione degli stati patologici in neonati pretermine. ....	83

## **Indice delle Figure**

Figura 1. La piramide dei processi biologici .....	9
Figura 2 Complessità dei sistemi biologici. ....	10
Figura 3. Il rivestimento mielinico dell'assone, sede dell'attacco della malattia.....	14
Figura 4 Le fasi dell'infiammazione. ....	15
Figura 5. Circuiti di flusso cerebrale del liquor.....	19
Figura 6. Rachicentesi o prelievo lombare di liquido cerebrospinale. ....	20
Figura 7. Disordini neurometaboliti e percorsi delle alterazioni geniche e di complessi proteici che inducono alterazioni funzionali e generano i disordini neuronali.....	22
Figura 8. Sequenza con presaturazione del segnale dell'acqua ed acquisizione del FID.....	24
Figura 9. Spettro NMR del H-1 di liquido cerebrospinale. ....	25
Figura 10. Spettro 1H-NMR del CSF nella regione tra 0 e 1,5 ppm. ....	28
Figura 11. Spettro 1H-NMR del CSF nella regione tra 1,75 e 3,35 ppm. ....	29
Figura 12. Spettro 1H-NMR del CSF nella regione tra 3 e 4,75 ppm. ....	30
Figura 13. Spettro 1H-NMR del CSF nella regione tra 5,2 ed 8,75 ppm. ....	31
Figura 14. Ciclo glicolisi/gluconeogenesi. ....	38
Figura 15. Ciclo metabolico della Tirosina e metaboliti collegati (v. citrato e piruvato). ....	40
Figura 16. Il clustering permette di determinare i parametri per descrivere le similitudini di sottogruppi di dati dell'insieme. ....	48
Figura 17. Rappresentazione generica della natura non binaria ma "fuzzy" della classificazione di dati di tipo biologico. ....	49
Figura 18. Spettro NMR protonico di un campione di liquor cerebrospinale. ....	50
Figura 19. Grafico degli integrali "bucket" per uno spettro NMR di liquor. ....	51
Figura 20. Diagramma di flusso dell'analisi dei dati.....	54
Figura 21. Analisi PCA su un primo campione di spettri. Correlazione tra i cluster e l'EDSS. ....	55
Figura 22. Localizzazione dei voxel per l'esame di spettroscopia nella regione periventricolare.....	56
Figura 23. Postprocessing degli spettri 1H-MRS relativi alle zone di sostanza bianca periventricolare.....	57
Figura 24. PCA sulla matrice native dei dati. I cluster non sono ben separati ed alcuni degli spettri hanno similarità non specifiche .....	59
Figura 25. Rappresentazione grafica della classificazione operata con SVM su dati incogniti di CIS.....	61
Figura 26. Decomposizione a componenti principali dopo il filtraggio delle componenti di artefatto.....	63
Figura 27. Loadings della analisi delle componenti principali dopo il filtraggio degli artefatti. ....	64
Figura 28. Modello ottenuto mediante analisi PCA sui dati filtrati. Le colorazioni evidenziano la collocazione degli spettri raccolti nelle varie classi (Controlli, CIS_SM, SM_SM, CIS_CIS) .....	65
Figura 29. Grafico dei loadings del modello M2. ....	66
Figura 30. Standard di riferimento TSP usato per gli esperimenti NMR protonici sulle urine. ....	72
Figura 31. Spettro NMR protonico di urine di neonati pretermine. ....	72

Figura 32. Analisi temporale delle urine di un neonato “pretermine” patologico.....	73
Figura 33. Confronto delle aree dei segnali NMR a tempi diversi, per un singolo paziente..	74
Figura 34. Traiettoria metabolica di un paziente patologico durante trattamento farmacologico. ....	75
Figura 35. PCA sui dati di spettroscopia protonica di urine di neonati pretermine. ....	76





## ***Ringraziamenti***

Le difficoltà incontrate nella realizzazione di questo progetto di ricerca sono state tante, anche perché il quesito medico affrontato è ancora poco conosciuto; esse sono state in gran parte superate grazie al contributo di molte persone come deve accadere in un progetto di natura multidisciplinare come quello in questione.

Ringrazio il mio supervisore prof. Adolfo Lai per il suo aiuto che si è spinto ben oltre i compiti istituzionali soprattutto nelle fasi di forte disagio affrontate quando i risultati tardavano ad arrivare; ringrazio il responsabile del mio assegno di ricerca, prof. Francesco Marrosu per la grande disponibilità mostrata nei confronti di questo progetto.

Un grazie a tutti membri del gruppo di lavoro: la dott.ssa Paola Scano, la dott.ssa Flaminia Cesare Marincola, e la dott.ssa Emanuela Locci, del Laboratorio di NMR del prof Lai, un grazie a loro sia per il contributo al lavoro scientifico sia per quello di orientamento per la mia formazione di Chimico-Fisico; ringrazio ancora la prof.ssa Marisa Marrosu, la dott.ssa Eleonora Cocco e la dott.ssa Elena Mamusa del Centro Studi Sclerosi Multipla di Cagliari per il supporto alla ricerca e per la pazienza mostrata nei numerosi tentativi di farmi capire i concetti neurologici di base che riguardano la SM.

Ringrazio il prof. L. Atzori, del Dipartimento di Tossicologia per il supporto alla ricerca sulle patologie neonatali.

I miei ringraziamenti a tutto il Collegio dei Docenti del Corso di Dottorato in Chimica ed al suo Coordinatore, prof Italo Ferino.

Infine mi sembra questa la giusta sede per esprimere un sentito ringraziamento alla dott.sa Daria Chissalè della Segreteria del Corso di Dottorato in Scienze Chimica dell'Università di Cagliari: il suo lavoro preciso e puntuale è stato essenziale. Auguro a tutti i dottorandi dei futuri di corsi in Scienze Chimiche di avere la collaborazione della dott.sa Chissale nei loro percorsi formativi.