

**DOTTORATO DI RICERCA
IN BIOTECNOLOGIE INDUSTRIALI-AMBIENTALI
XVIII CICLO**

**UNIVERSITA' DEGLI STUDI DI VERONA
FACOLTA' DI SCIENZE MM.FF.NN.
DIPARTIMENTO SCIENTIFICO E TECNOLOGICO**

PhD Student: NATASCIA CAMPOSTRINI

**STUDY OF NEUROBLASTOMA TUMOR USING
PROTEOME ANALYSIS**

Coordinator: Prof. HUGO LUIS MONACO

Supervisor: Prof. PIER GIORGIO RIGHETTI



ANNI ACCADEMICI 2003-2006

*Ai miei genitori
e a Damiano.*

*Nessuno conosce le proprie possibilità
finchè non le mette alla prova....*



UNIVERSITÀ DEGLI STUDI DI VERONA

DIPARTIMENTO SCIENTIFICO E TECNOLOGICO

Strada Le Grazie 15 - Ca' Vignal - 37134 Verona (Italia) - Tel. +39 045 8027925/38 - Fax. +39 045 8027929 ;

Alla cortese attenzione del Collegio dei Docenti:

Oggetto: profilo della Dott.ssa Natascia Campostrini, dottoranda di ricerca in Biotecnologie Industriali Ambientali, XVIII ciclo, Dipartimento Scientifico e Tecnologico.

La Dott.ssa Campostrini, nel corso del triennio di attività scientifica svolta nell'ambito del Dottorato di Ricerca in Biotecnologie Industriali-Ambientali presso il Dipartimento Scientifico e Tecnologico dell'Università di Verona, ha condotto un brillante ed approfondito lavoro inerente gli aspetti connessi agli approcci innovativi ed all'applicazione dell'analisi proteomica comparativa a problematiche di interesse biomedico.

Nello svolgimento del lavoro di ricerca la Dott.ssa Campostrini ha affrontato le problematiche connesse allo studio ed implementazione di metodiche e tecniche innovative per l'analisi proteomica comparativa attraverso la messa a punto di nuovi metodi statistici per l'analisi dei profili proteici.

Sono state affrontate inoltre le tematiche relative la preparazione e la separazione dei campioni proteici allo scopo di valutare i meccanismi coinvolti nella metastasi del tumore Neuroblastoma, nonché le variazioni nell'espressione proteica in seguito al trattamento con farmaci anti-angiogenesi.

I temi sono stati affrontati con rigore scientifico, portando alla messa a punto di nuove metodologie per l'analisi proteomica ad alta risoluzione e alla dettagliata spiegazione delle dinamiche di proteine coinvolte nella metastasi tumorale e nell'effetto anti-angiogenico di farmaci chemoterapici.

La produzione scientifica svolta dalla Dott.ssa Campostrini in questi tre anni, in seguito ai progetti attuati, ha portato alla pubblicazione dei seguenti articoli:

- 1) "The proteome: anno Domini 2002".
Righetti, P. G., Castagna, A., Antonucci, F., Piubelli, C., Cecconi, D., Campostrini, N., Zanusso, G., Monaco, S., *Clin. Chem. Lab. Med.* 2003, 41, 425-438.
- 2) "Proteome analysis in laboratory medicine: Close to reality?".
Righetti, P. G., Castagna, A., Antonucci, F., Cecconi, D., Campostrini, N., Gil-Augusti, M. T., Zanusso, G., Monaco, S., *LabMedica International* 2003, 20, 18-22.
- 3) "Quantitative proteomics: a review of different methodologies".
Righetti, P. G., Campostrini, N., Pascali, J., Hamdan, M., Aster, H., *Eur. J. Mass Spectrom.* 2004, 10, 335-348.

- 4) "Study of proteomic changes associated with healthy and tumoral murine samples in neuroblastoma by principal component analysis and classification methods".
Marengo, E., Robotti, E., Righetti, P. G., Campostrini, N., Pascali, J., Ponzoni, M., Hamdan, M., Aster, H., *Clin. Chim. Acta* 2004, 345, 55-67.
- 5) "Proteomic analysis of an orthotopic neuroblastoma xenograft animal model".
Campostrini, N., Pascali, J., Hamdan, M., Aster, H., Marimpietri, D., Pastorino, F., Ponzoni, M., Righetti, P. G., *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 2004, 5, 279-286.
- 6) "Critical survey of quantitative proteomics in two-dimensional electrophoretic approaches".
Righetti, P. G., Castagna, A., Antonucci, F., Piubelli, C., Cecconi, D., Campostrini, N., Antonioli, P., Aster, H., Hamdan, M., *J. Chromatogr. A* 2004, 1051, 3-17.
- 7) "Numerical approaches for quantitative analysis of two-dimensional maps: a review of commercial software and home-made systems".
Marengo, E., Robotti, E., Antonucci, F., Cecconi, D., Campostrini, N., Righetti, P. G., *Proteomics* 2005, 5, 654-666.
- 8) "Two-dimensional mapping as a tool for classification of green coffee bean species".
Gil-Agusti, M. T., Campostrini, N., Zolla, L., Ciambella, C., Invernizzi, C., Righetti, P. G., *Proteomics* 2005, 5, 710-718.
- 9) "Spot overlapping in two-dimensional maps: a serious problem ignored for much too long".
Campostrini, N., Areces, L. B., Rappsilber, J., Pietrogrande, M. C., Dondi, F., Pastorino, F., Ponzoni, M., Righetti, P. G., *Proteomics* 2005, 5, 2385-2395.
- 10) "Proteomic approaches for studying chemoresistance in cancer".
Righetti, P. G., Castagna, A., Antonioli, P., Cecconi, D., Campostrini, N., Righetti, S. C., *Expert Rev. Proteomics* 2005, 2, 215-228.
- 11) "Proteome analysis in the clinical chemistry laboratory: myth or reality?".
Righetti, P. G., Castagna, A., Antonucci, F., Piubelli, C., Cecconi, D., Campostrini, N., Rustichelli, C., Antonioli, P., Zanusso, G., Monaco, S., Lomas, L., Boschetti, E., *Clin. Chim. Acta* 2005, 357, 123-139.
- 12) "Formation of truncated proteins and high-molecular-mass aggregates upon soft illumination of photosynthetic proteins".
Rinalducci, S., Campostrini, N., Antonioli, P., Righetti, P. G., Roepstorff, P., Zolla, L., *J. Proteome Res.* 2005, 4, 2327-2337.

13) "Pre-fractionation techniques in proteome analysis".

Righetti, P. G., Cecconi, D., Campostrini, N., Antonioli, P., Accepted for publication in *Encyclopaedia of Separation Science*, Ian D. Wilson (ed), Academic Press, London, 2006 (in press). ISBN: 0122267702.

14) "Quantitative Proteomics".

Righetti, P. G., Antonioli, P., Campostrini, N., Cecconi, D., Accepted for publication in *Encyclopaedia of Separation Science*, Ian D. Wilson (ed), Academic Press, London, 2006 (in press). ISBN: 0122267702.

Il numero delle pubblicazioni prodotte certifica l'importante impegno profuso dalla Dott.ssa Campostrini nell'adempimento del suo lavoro.

L'elaborato della tesi è sicuramente collocabile al livello di ottimo.



Il responsabile Scientifico
Prof. Pier Giorgio Righetti

Visto

Il Coordinatore del Dottorato di Ricerca
in Biotecnologie Industriali ed Ambientali
Prof. Hugo L. Monaco

Verona, 28 Febbraio 2006

RIASSUNTO

STUDIO DEL TUMORE NEUROBLASTOMA ATTRAVERSO L'UTILIZZO DI UN APPROCCIO PROTEOMICO

Il progetto di dottorato è stato condotto nel laboratorio di Metodologie Biochimiche del Dipartimento Scientifico e Tecnologico dell'Università di Verona. Per la stesura di tale progetto sono state instaurate collaborazioni con alcuni laboratori: in particolare il Laboratorio di Oncologia dell'Ospedale pediatrico G. Gaslini di Genova, ha fornito tutti i campioni biologici utilizzati per le analisi proteomiche. Tutti gli esperimenti che hanno previsto l'utilizzo di animali sono stati approvati dalla Commissione etica dell'Istituto Nazionale della Ricerca sul Cancro e dal Ministero Italiano della Salute. L'identificazione delle proteine differenzialmente espresse in seguito all'analisi proteomica comparativa è stata ottenuta grazie alla collaborazione con i seguenti laboratori: il Laboratorio di Spettrometria di Massa dal Centro Ricerche GSK di Verona; l'Istituto FIRC della Fondazione di Oncologia Molecolare (IFOM) e l'Istituto Europeo di Oncologia (IEO) di Milano; i Laboratori di Espressione Genica e di Biologia Cellulare dell'Istituto Nazionale delle Malattie Infettive "Lazzaro Spalanzani" di Roma. Lo sviluppo di un nuovo metodo statistico per l'analisi dei profili proteici è stato possibile grazie alla collaborazione con il Dipartimento di Scienze e Tecnologie Avanzate dell'Università del Piemonte Orientale, mentre l'applicazione della teoria dello "Spot Overlapping" a mappe bidimensionali reali è stata ottenuta in collaborazione con il Dipartimento di Chimica dell'Università di Ferrara.

Gli obiettivi di questo progetto di dottorato sono stati:

- 1) *l'applicazione dei metodi classici della proteomica comparativa per lo studio della crescita e della progressione del tumore Neuroblastoma in due diversi modelli animali e la valutazione dell'effetto sinergistico anti-angiogenesi in seguito alla somministrazione, in maniera combinata, di due farmaci chemioterapici su una linea di cellule endoteliali;*
- 2) *lo sviluppo di un nuovo metodo statistico per l'analisi dei profili proteici e l'applicazione della teoria dello "Spot Overlapping" a mappe bidimensionali reali.*

Per quanto riguarda il primo obiettivo del progetto di dottorato, i metodi classici della proteomica comparativa sono stati utilizzati per monitorare la crescita e la progressione del tumore Neuroblastoma (NB) in sistemi *in vivo*. Il NB è il più comune tumore solido a localizzazione extracranica dei bambini, che prende origine dal tessuto simpatico, da cui fisiologicamente prendono origine la midollare del surrene ed i gangli del sistema nervoso simpatico. Sebbene la chemioterapia ad alte dosi e la irradiazione corporea totale, seguite da trapianto di midollo osseo, abbiano esteso la sopravvivenza a lungo termine in pazienti affetti da NB metastatico, la prognosi rimane a tutt'oggi generalmente infausta. Partendo da queste considerazioni c'è una grande necessità di implementare nuove strategie terapeutiche. La terapia anticancro, per essere efficace, deve essere "individualizzata", cioè essere studiata in base alle caratteristiche particolari di ogni individuo. Questo richiede una comprensione dettagliata delle proprietà di ciascun tumore rispetto alla propensione a formare metastasi e sensibilità ai farmaci, informazioni che sono difficili da acquisire mediante analisi microscopica. Devono essere ottenute informazioni dettagliate riguardanti le mutazioni a carico dei geni che controllano la crescita, la metastasi e la resistenza a farmaci. In questo contesto l'analisi proteomica può fornire un valido supporto sia diagnostico che come ricerca di base per definire marcatori utili per la diagnosi dei tumori e può anche essere utile per approfondire le conoscenze sui meccanismi coinvolti nella patogenesi del cancro. Inoltre la complessità delle interazioni tra il NB e il microambiente enfatizza l'importanza di modelli di studio *in vivo*. Proprio per questa ragione, in questo progetto, sono stati utilizzati due modelli animali di topi immunocompromessi. Per caratterizzare il processo di sviluppo e di crescita tumorale è stato utilizzato il topo nudo in cui è stato fatto un trapianto di cellule murine tumorali di NB a livello delle ghiandole surrenali. Su campioni di surrene sano e surrene con massa primaria è stata condotta un'analisi proteomica comparativa e in seguito al confronto delle mappe 2D con il software PDQuest è stato possibile individuare 84 macchie differenzialmente espresse. Grazie alla spettrometria di massa MALDI-TOF, 14 proteine sono state identificate. Tra queste, le più interessanti sono rappresentate dalla proteina 14-3-3 zeta-delta (coinvolta nel controllo apoptotico mediante l'interazione con le proteine Bad e Raf-1), dalla ciclofilina A (la cui sovraespressione correla con la tumorigenesi), dalla nucleofosmina (coinvolta nell'oncogenesi) e dalla statmina (la cui sovraespressione è stata dimostrata in molte malattie umane come le leucemie acute, i linfomi e i carcinomi al seno). Questo studio ci ha permesso di dimostrare che l'approccio proteomico rappresenta uno strumento molto utile per lo studio e la comprensione delle proteine associate allo sviluppo del tumore.

Successivamente ci siamo voluti occupare dello studio della progressione tumorale e della delucidazione dei meccanismi che portano allo sviluppo di metastasi. Per questo motivo è stato impiegato come modello di studio il topo SCID in cui sono state trapiantate cellule

umane tumorali di NB a livello delle ghiandole surrenali. E' stato scelto questo modello perchè riflette il tipico profilo di crescita del NB umano, in quanto l'iniezione delle cellule tumorali a livello delle ghiandole surrenali comporta la formazione di tumori solidi che sono localmente invasivi nei tessuti circostanti e metastatici in siti secondari. In questo caso l'analisi proteomica differenziale è stata eseguita su campioni di massa primaria (tumore primario sviluppatosi a livello delle ghiandole surrenali) e metastasi di fegato (il fegato rappresenta l'organo maggiormente colpito da metastasi nel tumore NB). Il confronto delle mappe bidimensionali con il software PDQuest ha permesso di individuare 85 proteine differenzialmente espresse, di cui 44 sono state identificate mediante analisi LC-MS/MS. Tra le proteine identificate, il gruppo più interessante è rappresentato da proteine coinvolte nella metastasi tumorale e nei processi ad essa correlati, come ad esempio la trasformazione, l'invasione, l'angiogenesi, la mobilità, l'apoptosi, la differenziazione e la proliferazione. Questo studio ci ha permesso di individuare almeno tre possibili bersagli terapeutici: la subunità p40 del Proteasoma 26S (PSMD7), il Fattore di inizio della traduzione eucariotica (eIF-4E) e l'Y Box binding protein-1 (YB1). Inoltre sono stati identificati anche due possibili bersagli molecolari in grado di controllare il processo di metastasi come Septina-2 (SEPT2) e la Proteina che inibisce la chinasi Raf (RKIP).

Si è proceduto quindi con l'analisi proteomica di cellule endoteliali trattate con due farmaci chemioterapici (Vinblastina e Rapamicina) allo scopo di individuare gli effetti molecolari di tale trattamento e di valutarne l'attività sinergistica anti-angiogenica. Uno dei passaggi fondamentali per l'attecchimento della metastasi tumorale è rappresentato dall'angiogenesi. L'angiogenesi consiste nella formazioni di nuovi vasi sanguigni, a partire da vasi pre-esistenti, e gioca un ruolo fondamentale nel processo neoplastico in quanto è essenziale per la diffusione metastatica dei tumori solidi ed ematologici. Proprio per questo motivo, la terapia anti-angiogenesi si presenta come approccio promettente nella terapia contro il cancro. Inoltre la scelta di cellule endoteliali, come modello di studio, deriva dal fatto che le cellule endoteliali sono considerate geneticamente più stabili delle cellule tumorali e quindi il problema della farmaco-resistenza, indotto dalla chemioterapia, può essere aggirato.

L'analisi proteomica è stata condotta su quattro campioni: la linea cellulare controllo, la linea trattata con Vinblastina, le cellule trattate con Rapamicina e le cellule trattate con i due farmaci in maniera combinata. Dallo studio sono risultate 113 macchie proteiche differenzialmente espresse, di cui 54 sono state identificate mediante analisi MALDI-MS e MALDI-MS/MS. Per studiare le associazioni funzionali e i bersagli comuni delle proteine identificate è stato usato un software per l'analisi dei pathways (PathwayAssist). In questa parte del progetto di dottorato è stato possibile dimostrare come il trattamento combinato con i due farmaci esibisca un effetto sinergistico sull'inibizione della crescita *in vitro* delle cellule endoteliali, inducendone l'apoptosi. Inoltre usando l'elettroforesi bidimensionale e la

spettrometria di massa, è stato possibile identificare un gruppo di proteine che potrebbero essere coinvolte nella cooperazione di Vinblastina e Rapamicina nell'inibire l'angiogenesi. Molte delle proteine identificate risultano essere modulate solo dal trattamento combinato e sembrano essere coinvolte, oltre che nel processo di angiogenesi, nella proliferazione, nella migrazione e nell'apoptosi.

Allo scopo di migliorare i metodi attualmente disponibili per l'analisi comparativa delle mappe bidimensionali si è proceduto, per quanto riguarda il secondo obiettivo del progetto di dottorato, con l'implementazione di un nuovo metodo statistico basato sull'analisi delle componenti principali (PCA) e sull'applicazione dei metodi di classificazione (SIMCA) per il confronto e l'analisi dei profili proteici. L'analisi PCA è stata applicata su un gruppo di 20 mappe (10 derivanti dai campioni di surrene sano e 10 da campioni di surrene tumorale). Il metodo della PCA ha permesso di suddividere correttamente le due classi di campioni. La separazione dei campioni nelle due classi è stata ulteriormente confermata con il modello di classificazione SIMCA. I risultati ottenuti con l'analisi PCA e con il metodo SIMCA si sono dimostrati in accordo con quelli derivanti dall'analisi PDQuest.

Infine, la teoria statistica dello "Spot Overlapping" è stata applicata per la prima volta a mappe bidimensionali reali. La teoria dello "Spot Overlapping" si basa sul fatto che la separazione di tutte le proteine presenti in un campione biologico è molto difficile da ottenere, e che la presenza di più proteine che co-migrano nella stessa macchia proteica è un fenomeno piuttosto comune. Si è riusciti a dimostrare che i dati sperimentali sull'abbondanza delle macchie sovrapposte è in buon accordo con i dati teorici calcolati sulla base di questa teoria.

Concludendo, tutti gli studi di analisi proteomica comparativa, effettuati durante questo progetto di dottorato, confermano che l'elettroforesi bidimensionale, accoppiata all'analisi tramite spettrometria di massa, è una tecnica efficace per l'analisi globale della variazione dell'espressione proteica, per lo studio e la comprensione dei meccanismi molecolari che stanno alla base dello sviluppo e della progressione di un tumore, per l'individuazione di possibili marcatori molecolari di malattie nonché di bersagli per nuove strategie terapeutiche.

PREFACE

Proteomics is defined as a scientific approach used to elucidate all protein species within a cell or tissue. It has been just over a decade since the term “proteomics” was coined, and even though this new term was used to introduce the concept of exploring changes in all proteins expressed by a genome, the tools used in proteomic analysis are borrowed from traditional biochemistry. Although this approach is commonly used for characterizing all protein species within a given cell, many researchers are also taking advantage of proteomic technology for elucidating protein changes between healthy and pathological states. There are many good reasons for tackling the proteome. Although of vexing complexity, it offers, in principle, a much richer source for the functional description of diseases and the discovery of diagnostic and therapeutic targets. An additional and unique advantage is that, in contrast with the genome, the inherently dynamic nature of the proteome allows us to monitor closely changes in the state of a cell, tissue or organism over time. Only dynamic information will permit to follow the course of a disease and track its pathogenetic mechanisms as well as its response to therapy. This may be of particular importance for diseases such as cancer, which evolve dynamically and affect many heterogeneous cell populations, either as part of the cancer or as part of the host’s reaction to the tumor.

The topics of this work are:

- 1) *the biomedical application of classical proteomic analysis for studying Neuroblastoma growth and progression in two different animal models and the synergistic anti-angiogenic effects of two chemotherapeutics in an endothelial cell line;*
- 2) *the development of a new statistical tool for comparative proteomics analysis and the application of the theory of spot overlapping to real experimental 2D-PAGE gels.*

The thesis work was conducted at the Proteomics Laboratory of the Scientific and Technologic Department of the University of Verona, in collaboration with other laboratories: the Laboratory of Oncology of the “G. Gaslini” Children’s Hospital in Genoa, which provided all the biological samples. All experiments involving animals have been reviewed and approved by the licensing and ethical committee of the National Cancer Research Institute and by the Italian Ministry of Health. The identification of proteins for all the proteomic analysis performed were possible thanks to the collaboration with the Mass Spectrometry Laboratory of Computational, Analytical & Structural Sciences, at the GlaxoSmithKline in Verona, the FIRC Institute of Molecular Oncology Foundation (IFOM) and European Institute of Oncology (IEO) in Milan as well as the Gene Expression and Cell Biology Laboratories of the National Institute of Infectious Diseases “Lazzaro Spallanzani” in Rome.

The development of a new statistical tool for the analysis of protein profiles was the result of the collaboration with the Department of Environmental and Life Sciences of the University of Eastern Piedmont. The application of the theory of spot overlapping to real 2D maps was made possible due to the collaboration with the Department of Chemistry of the University of Ferrara.

The results thus obtained are here discussed and evaluated.

CONTENTS

Riassunto	I
Preface	V
Contents	VII
List of abbreviations	XI

CHAPTER 1: PROTEOMICS IN CANCER RESEARCH

1.1	General introduction	1
1.2	Cancer Proteomics	2
1.2.1	Expression Proteomics	3
1.2.1.1	Two-dimensional electrophoresis (2D-PAGE) and protein identification by Peptide Mass Fingerprinting	3
1.2.1.1.1	Limitations of 2D-PAGE	10
1.2.1.2	Multi-dimensional protein identification technology	13
1.2.1.3	SELDI protein chip system	15
1.2.2	Functional proteomics	17
1.2.2.1	Study of protein interactions	17
1.2.2.2	Detection of post-translational modifications	21
1.3	Clinical applications of cancer proteomics	23
1.3.1	Biomarkers discovery for cancer diagnosis and prognosis	23
1.3.2	Pharmacoproteomics	25
1.3.3	Tumor immunology	25
1.4	Bibliography	26

CHAPTER 2: NEUROBLASTOMA

2.1	Introduction	31
2.2	Clinical characteristics	32
2.3	Overview of Neuroblastoma genetics	33
2.4	Diagnosis	36
2.5	Neuroblastoma treatment	37
2.5.1	Standard treatment	37
2.5.2	Other types of treatment	38
2.6	The importance of animal models in cancer	41
2.6.1	Neuroblastoma models	43
2.7	Bibliography	44

CHAPTER 3: PROTEOMIC ANALYSIS OF AN ORTHOTOPIC NEUROBLASTOMA ANIMAL MODEL

3.1	Introduction	47
3.1.1	Biological relevance of the orthotopic Neuroblastoma xenograft model	48
3.2	Experimental procedures	49
3.2.1	Orthotopic Neuroblastoma animal model	49
3.2.2	2D-PAGE	49
3.2.3	Protein pattern and statistical analysis	50
3.2.4	Protein identification by Mass Spectrometry	51
3.2.5	Immunoblot analysis	51
3.3	Results	52
3.4	Discussion	58
3.5	Conclusions	60
3.6	Bibliography	61

CHAPTER 4: STUDY OF HUMAN NEUROBLASTOMA PROGRESSION USING PROTEOME ANALYSIS

4.1	Introduction	63
4.2	Experimental procedures	65
4.2.1	Xenograft Neuroblastoma animal model	65
4.2.2	2D-PAGE	65
4.2.3	Protein identification by Mass Spectrometry	66
4.2.4	Immunoblot analysis	67
4.2.5	Biochemical function and Pathway analysis	67
4.3	Results and Discussion	68
4.3.1	Differential proteomic analysis	74
4.3.1.1	Differentially down-modulated proteins	74
4.3.1.2	Differentially up-modulated proteins	76
4.3.2	Western blot analysis	79
4.4	Concluding remarks	80
4.5	Bibliography	81

CHAPTER 5: PROTEOMIC ANALYSIS OF ANTI-ANGIOGENIC EFFECTS BY A COMBINED TREATMENT WITH VINBLASTINE AND RAPAMYCIN IN AN ENDOTHELIAL CELL LINE

5.1	Angiogenesis	83
5.1.1	Anti angiogenesis therapy	84
5.2	Experimental procedures	87
5.2.1	Chemotherapeutics	87
5.2.2	Cell culture	87
5.2.3	Proliferation assay	87
5.2.4	Phosphatidylserine exposure	87
5.2.5	2D-PAGE	88
5.2.6	Protein pattern differential analysis	88
5.2.7	Protein identification by Mass Spectrometry	88
5.2.8	Biochemical function and pathway analysis	89
5.2.9	Statistics	89

5.3	Results	90
5.3.1	Inhibition of cell proliferation	90
5.3.2	Induction of apoptosis	91
5.3.3	Differential proteomic analysis and pathway analysis	92
5.4	Discussion	98
5.4.1	Differentially down-modulated proteins	98
5.4.2	Differentially up-modulated proteins	103
5.5	Conclusions	104
5.6	Bibliography	105

CHAPTER 6: PRINCIPAL COMPONENT ANALYSIS AND CLASSIFICATION METHODS: APPLICATIONS TO MURINE NEUROBLASTOMA SAMPLES

6.1	Introduction	107
6.2	Theory	108
6.2.1	Principal Component Analysis	108
6.2.2	Classification: SIMCA model	108
6.3	Experimental procedures	109
6.3.1	2D-PAGE	110
6.3.1.1	Sample treatment	110
6.3.1.2	Electrophoretic separation	110
6.3.1.3	Software	110
6.4	Results and discussion	111
6.4.1	Principal Component Analysis	111
6.4.2	Classification	114
6.5	Conclusions	118
6.6	Bibliography	119

CHAPTER 7: SPOT OVERLAPPING IN TWO-DIMENSIONAL MAPS

7.1	Introduction	121
7.2	Experimental procedures	124
7.2.1	Xenograft Neuroblastoma animal model	124
7.2.2	2D-PAGE, MS and data analysis	124
7.2.3	Application of the spot overlapping theory	124
7.3	Results	126
7.4	Discussion	132
7.4.1	The theory of spot overlapping at work	132
7.4.2	How to minimize spot overlap	132
7.4.3	On the ΔpI and ΔM_r spread in overlapping spots	133
7.4.4	On the significance of apparently "aberrant" spots	134
7.4.5	Protein analysis by Mass Spectrometry	136
7.5	Bibliography	137

CHAPTER 8: GENERAL CONCLUSIONS

8.1	Conclusions	139
8.2	Bibliography	142

Papers published during the PhD studies

Ringraziamenti

LIST of ABBREVIATIONS

2-D:	two-dimensional
2-DE:	two-dimensional electrophoresis
2D-PAGE:	two-dimensional polyacrylamide gel electrophoresis
AA:	acrylamide
CA:	carrier ampholytes
CHAPS:	3-[(3-cholamidopropyl)dimethylammonium]-1-propane-sulfonate
DIGE:	differential in-gel electrophoresis
DTT:	dithiothreitol
ECL:	enhanced chemiluminescence
ECs:	endothelial cells
ESI:	electrospray-ionization
ICAT:	isotope-coded affinity tag
IEF:	isoelectric focusing
IPG:	immobilized pH gradient
LC/MS-MS:	liquid chromatography tandem mass spectrometry
LC:	liquid chromatography
LCM:	laser capture microdissection
MALDI:	matrix-assisted-laser-desorption/ionization
MCE:	multicompartment electrolyzer
M _r :	relative mass
MS:	mass spectrometry
MUDPIT:	multi dimensional protein identification technology
NB:	Neuroblastoma tumor
NCBI:	national center of biotechnology information
OD:	optical density
PCA:	principal component analysis
PCs:	principal components
pI:	isoelectric point
PNA:	peptide nucleic acid
PTMs:	post-translational modifications
PVDF:	polyvinylidene difluoride membrane
RAP:	rapamycin
SC:	single component
SCID:	severe combined immune deficiency

SDO:	statistical degree of overlapping model
SDS:	sodium-dodecyl-sulphate
SDS-PAGE:	sodium-dodecyl-sulphate polyacrylamide gel electrophoresis
SELDI:	surface enhanced laser desorption ionisation
SIMCA:	soft-independent model of class analogy
SSP:	standard spot number
TBP:	tributyl phosphine
TBS:	Tris-buffered saline
TOF:	time of flight
VBL:	vinblastine

CHAPTER 1

PROTEOMICS IN CANCER RESEARCH

1.1 GENERAL INTRODUCTION

The term Proteome, as coined in 1994 by Mark Wilkins and colleagues, refers to the study of the entire asset of proteins present in a cell, or tissue, or biological fluid (the PROTEin complement expressed by a genOME) [1,2]. Although the exact definition of proteomics varies among researchers, it is widely accepted that proteomics is “the study of all the protein forms expressed within an organism as a function of time, age, state and external factors” [3]. In other words, proteomics characterizes the behaviour of the system rather than the behaviour of any single component. Proteomics is a new approach that integrates some key fundamental technologies, such as high-throughput protein separation and profiling, mass spectrometry, genomic and proteomics databases and browser-based bioinformatics tools for extracting information from these databases.

The sequencing of the human genome has confronted us with the daunting task of explaining the myriad of physiological and pathological functions of a highly complex and adaptable organism by a mere 40,000 odd genes. Clearly, a ‘one gene – one function’ relationship does not apply. Even worse, due to differential splicing and translation, each human gene may encode four to six different proteins. Taking into account the numerous PTMs (post-translational modifications), our 40,000 genes may generate around 1 million distinguishable functional entities at the protein level. Hence there are many good reasons for tackling the proteome. Although of vexing complexity, it offers, in principle, a much richer source for the functional description of diseases and the discovery of diagnostic and therapeutic targets. An additional and unique advantage is that, in contrast with the genome, the inherently dynamic nature of the proteome allows us to monitor closely changes in the state of a cell, tissue or organism over time. Only dynamic information will allow us to follow the course of a disease and track its pathogenetic mechanisms as well as its response to therapy. This may be of particular importance for diseases such as cancer, which evolve dynamically and affect many heterogeneous cell populations, either as part of the cancer or as part of the host’s reaction to the tumor [4].

1.2 CANCER PROTEOMICS

“Cancer proteomics” is the study of the molecular pathogenesis of cancer by analysing global protein expression in tumors, tumor cells or extracellular fluids from cancer patients [5]. A major justification for measuring proteins instead of gene transcripts [using nucleic acid-based methods such as cDNA arrays, SAGE (Serial Analysis of Gene Expression) and other methods] is that transcript levels do not necessarily correlate to protein levels. When the same cells or tumors have been examined by both cDNA arrays and proteome methods, the correlation between mRNA transcript profiles and corresponding protein abundances has been reported to be moderate [6-10]. Cellular structural gene products show a high correlation between mRNA and protein levels, whereas the correlation between protein and mRNA levels was found to be weaker for other gene products. It is obviously necessary to assess protein levels, and not mRNA levels, in instances where only protein expression levels correlate to disease. A major advantage of proteomics is the possibility to analyze post-translational modifications. A number of different gene products can arise from a single gene due to alternative splicing and various post-translational modifications, resulting in a substantially increased complexity of proteins compared to RNA transcripts. The complex and dynamic nature of proteins renders proteome studies quite challenging. It is often of limited value to study cellular signal transduction processes at the mRNA level. These events are characterized by protein phosphorylations, acetylations, and other modifications, or by cleavage of proteins by site-specific proteases. Changes in mRNA abundance are expected to be secondary to these modifications.

Cancer proteomics includes two categories: “expression proteomics” and “functional proteomics” [11].

“Expression proteomics” addresses the issue of protein expression levels in a given sample, such as a tumor, normal tissue or body fluid; this approach is particularly useful for protein profiling studies aimed at determining the difference between normal and tumor tissues. The protein profile (or ‘signature’) of cancer should allow investigators to identify tumor markers and, ultimately, disease-specific and/or patient-tailored molecular targets for the development of effective, non-toxic antitumoral drugs.

“Functional proteomics” includes the analysis of protein–protein interactions, protein–non-protein (e.g., DNA, RNA or lipids) interactions and protein post-translational modifications, which will help researchers to dissect the intricate network of molecular pathways underlying cellular activities in health and disease. These types of data should greatly improve our understanding of the biological process of tumor development and progression, thus fostering the identification of not only single protein-specific therapeutic targets but also pathway-specific targets.

In clinical cancer research, the techniques used in proteomics fall roughly into two groups: those used in expression proteomics, such as 2D-PAGE, multidimensional chromatography, and protein chips used for protein profiling (*e.g.*, the SELDI protein chip system), and those used in functional proteomics, such as immunoprecipitation, protein chips, array-based approaches, and yeast two-hybrid system for direct analysis of functional protein complexes, and post-translational modifications.

1.2.1 EXPRESSION PROTEOMICS

1.2.1.1 TWO-DIMENSIONAL ELECTROPHORESIS (2D-PAGE) AND PROTEIN IDENTIFICATION BY PEPTIDE MASS FINGERPRINTING

Proteins perform a variety of essential functions in mammalian organisms. Dynamic functions include catalysis of chemical transformations, transport, metabolic control and contractions. Many genetic diseases, including malignant tumors, result from altered levels of enzyme production or specific alterations to their amino acid sequence. In proteomics, proteins can be separated and characterized on the basis of their pI, the pH at which a protein is electrically neutral and in zwitterion form, and by such techniques as 2D-PAGE and chromatography (*e.g.*, ion-exchange or reverse phase). This allows hundreds to thousands of proteins to be analysed depending on protein detection methods. The different displayed proteins are converted to peptides by in-gel digestion with a site specific proteases (*e.g.*, trypsin). The mixture of digested protein is analysed by mass spectrometry, and the peptide spectrum generated by this analysis is used to search databases for protein identification (Figure 1) [12].

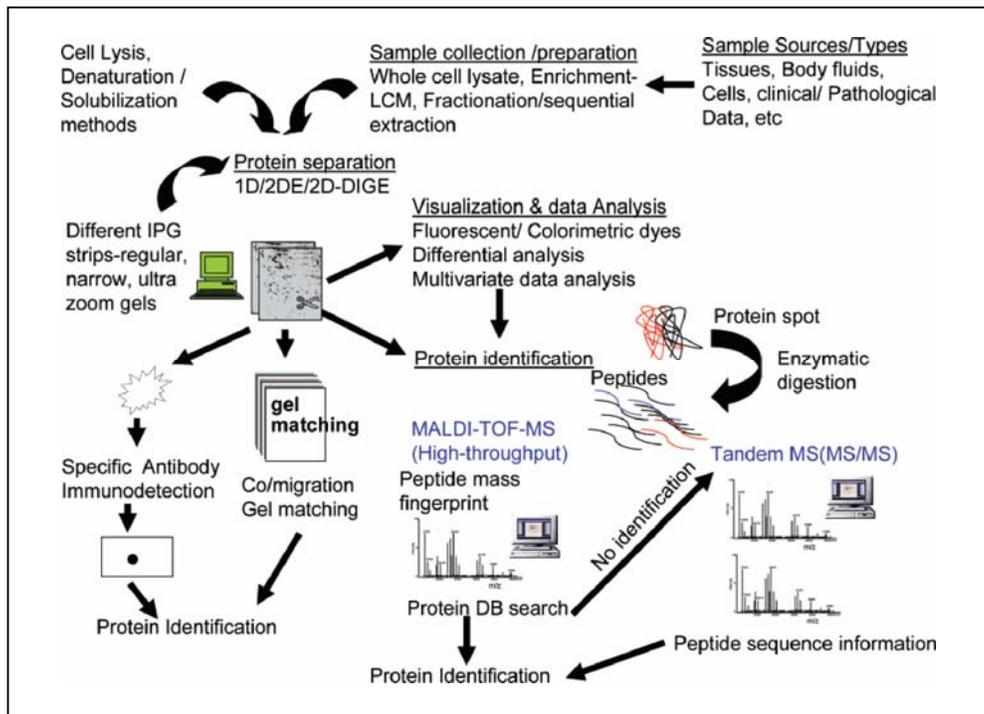


Figure 1. Overview of clinical proteomics strategies used to separate, analyze, identify and discover disease-associated proteins. Technologies include 2DE, immuno-detection, mass spectrometry (MS), and database (DB) searching. Abbreviations: 2DE, two-dimensional gel electrophoresis; DIGE, differential gel electrophoresis; MALDI-TOF MS, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry; MS/MS, Tandem mass spectrometry.

Sample preparation

Proteins can be extracted from blood, fluid, fresh tissue or cell culture by lysis buffers, which include protease inhibitors, chaotropes, detergents and reducing agents. These buffers disrupt hydrophobic interactions, hydrogen and disulfide bonds, unwanted aggregations, and the formation of secondary (and higher) structures, which affect protein mobility [13]. Factors such as protein solubility, size, charge, and pI enter into sample preparation. Sample preparation is important in reducing the complexity of the protein mixture.

Separation of proteins using 2D-PAGE

High-resolution two-dimensional gel electrophoresis remains a standard tool for the expression proteomics since its introduction 30 years ago (proteins are separated by isoelectrofocusing, IEF, with different pH gradient for the first dimension and by sodium dodecyl sulphate-polyacrylamide gel electrophoresis, SDS-PAGE, for the second dimension). The first dimension (IEF) is performed by using pH gradient strip gels with narrow or broad pH ranges. The protein will migrate at the point in the gradient equal to its pI. After IEF, the gel strips are immediately equilibrated in a buffer containing SDS, transferred, and separated using SDS-PAGE.

Differential in-gel electrophoresis (2D-DIGE)

2D-DIGE is a fairly recent improvement of the 2-DE technology. Prior to electrophoresis, the protein samples to be compared are covalently labelled with a green or red fluorescent dye respectively. A mixture of all the samples in the study is then labelled with a blue dye, serving as an internal standard. All the three samples are then mixed together and separated by 2-DE. This avoids ambiguities of spot matching as the samples are separated on the same gel, and inter-gel comparisons are greatly improved by the common internal standard. The individual samples are displayed by scanning the gel at the three distinctive wavelengths. Since the fluorescence signal is linear over a wide protein concentration range and each spot contains its own internal standard, comparative quantification is very accurate. In addition, the sensitivity is comparable with silver staining, but, in contrast with silver staining, fluorescent stains are fully MS compatible (Figure 2) [14-16].

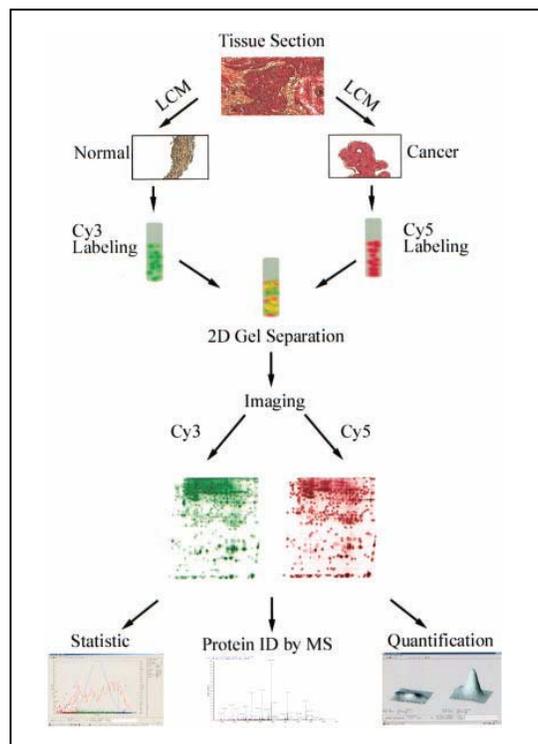


Figure 2. Flow chart of 2D-DIGE analysis of normal cells and cancer cells from the same tumor sample.

Detection of proteins in the gels

Protein detection is an important step in proteomics. Proteins are commonly visualized in gels by staining with dyes such as colloidal Coomassie Blue [17], Sypro Ruby (a fluorescent stain) [18] and Silver stain [19] that is compatible with subsequent MS analysis. The 2-D gels are then digitized and the resulting gel images are qualitatively and quantitatively analyzed with specialized software programs (such as PDQuest, Melanie, Z3 and Z4000, Phoretix and Progenesis). In this manner, proteins can be quantified and spot patterns in multiple gel can be matched and compared. Statistical analysis can be performed on groups of features (spots) in sets of gels, and variations, differences, and similarities can be evaluated. Proteins resolved by 2-DE can be identified based on unique attributes that are measured by MS.

Proteolytic digestion

Mass spectrometry can analyze intact proteins and peptides. Intact mass measurements are in general of relatively little use in protein identification and characterization. Intact proteins must be digested with enzymes before mass spectrometric analysis because the sensitivity of measurements of intact protein masses is not as accurate as that for peptides, and several post-translational modifications complicate determination based on mass. Trypsin is the most commonly used protease in proteomic analysis and displays good activity in both solution and in-gel digestion.

Protein identification

Protein identification is the most important step in expression proteomics. In proteomic research, proteins may be identified by various methods, such as peptide mass fingerprinting and tandem mass spectrometry [20]. Mass spectrometers are based on a combination of three essential components: (1) the ion source, which produces ions from the sample protein mixture; (2) the mass analyzer, which resolves ionized analytes on the basis of their mass-to-charge (m/z) ratio; and (3) the detector, which detects the ions resolved by the mass analyzer. Two types of MS are used for most proteomics studies: matrix-assisted laser-desorption ionization (MALDI) and electrospray ionization (ESI) MS instruments, which volatilize and ionize the proteins or peptides for their further mass analysis (Figure 3).

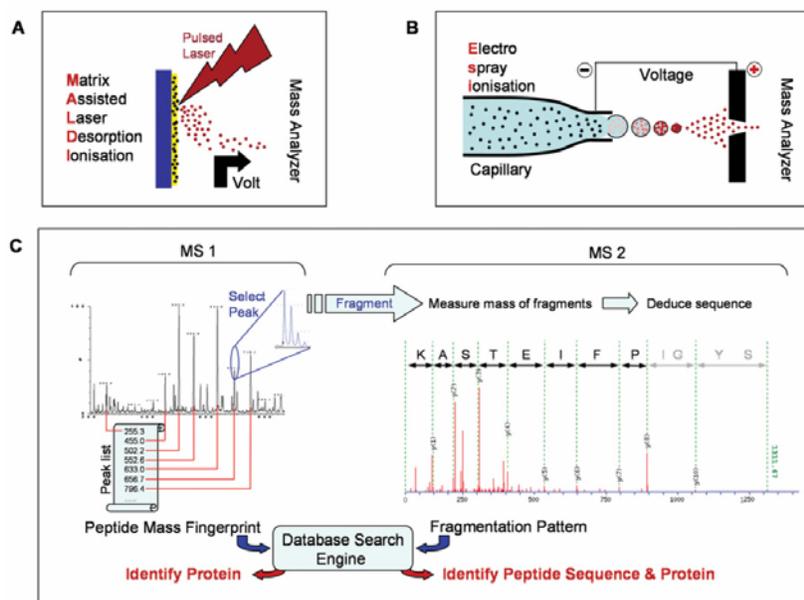


Figure 3. Some principles of MS. (A) and (B) are different ionization methods for peptides. (C) A typical MS experiment. Left-hand panel, peptide mass fingerprint. A survey scan measures the peptide masses, which are assembled into a peak list for database searches. The search engine compares the peptide masses with a theoretical peptide digest of the database and identifies proteins that match the measured masses. Several matching peptide masses are required for a statistically significant identification. Right-hand panel, determination of peptide sequences. This procedure requires a tandem MS. Using MS1, the peptide peak to be sequenced is selected from the survey scan and fragmented, usually by collision with gas. The fragment masses are determined by MS2. The peptide sequence can be deduced from the fragment masses. Note that the fragmentation is not homogenous with missing fragments precluding the deduction of the complete peptide sequence. Therefore, rather than trying to read the sequence from the mass spectrum, the fragmentation pattern is usually submitted to a database search engine which compares the observed pattern with theoretically expected peptide fragmentation patterns of the proteins in the database. Thus the peptide sequence, and in most cases the identity of the protein, can be inferred.

They are very different in design and performance, and each has its own strength and weakness. MALDI is usually coupled to time-of-flight (TOF) analyzers that measure the mass of whole proteins or peptides, whereas ESI has been mainly coupled to different MS instruments, such as ion traps or triple quadrupole instruments, to generate fragment ion (collision induced) spectra of selected precursor ions. MALDI-TOF is still widely used to study large proteins by what is known as 'peptide-mass mapping'. In this method, proteins are identified by matching a list of experimental peptide masses with a calculated list of all of the peptide masses of each entry in, for example, a comprehensive protein database. For an in-depth analysis of individual MS spectral peaks, however, a combination of multiple MS analyzers in a tandem-MS (MS/MS) instrument, such as MALDI-TOF/TOF, ESI-ion trap or ESI-triple quadrupole, is usually adopted for performing a two-stage or multi-stage mass analysis of ions; this approach is ideal for analyzing both biomacromolecules and small molecules because it yields an accurate determination (10-100 kDa) of the molecular mass

of large proteins and fine structural information directly from the biomolecules. As compared with peptide mass mapping, this approach provides information not only about peptide mass but also about peptide sequence, which is scanned against comprehensive protein sequence databases by using one of several different algorithms. Because protein identification relies on matches with sequence databases, however, high-throughput proteomics is currently restricted largely to those species for which comprehensive sequence databases are available. Several protein sequence databases are available in the public domain. An excellent annotated database is SWISSPROT, that is maintained by The Swiss Institute of Bioinformatics and The European Bioinformatics Institute [21]. The main advantages of SWISSPROT are low redundancy and a high degree of annotation. Database search programs are often included in commercial software packages that are provided with the mass spectrometers. One such example is the SEQUEST program that is used for database searching with uninterpreted product-ion spectra. A number of search engines can also be accessed free-of-charge over the Internet, for example the PeptIdent and MultIdent programs at the ExPASy Molecular Biology server, MS-fit and MS-Tag at the Protein Prospector server, or MASCOT at the Matrix Science server. These websites also provide additional proteomics software tools, technical information, and links to other resources.

Bioinformatics

Bioinformatic tools are integral components of proteomic analysis because the development of proteomic technologies incorporating high-throughput methods relies on powerful data analysis tools. The handling and analysis of the type of data to be collected in proteomic investigations are forcing new collaborations among computer scientists, biostatisticians, and biologists. Bioinformatic tools are necessary for multiple steps of proteomic investigations, including analysis, storage, management, search, and retrieval. Bioinformatic tools are being developed around current proteomic platforms of two-dimensional gel electrophoresis, MS, and arrays. Tools developed to analyze two-dimensional electrophoresis protein patterns include software applications possessing user-friendly interfaces for linearization and merging of scanned images, segmentation and detection of protein spots on the images, matching, and editing [22]. Commercial two-dimensional image analysis software available to the proteomics community include ImageMaster™ (Amersham Biosciences), Melanie 3 (GeneBio), Progenesis™ (Nonlinear Dynamics), PDQuest™ 2-D Analysis Software (Bio-Rad Laboratories), ProteinMine™ (Scimagix), and the Z3 2D-Gel Analysis System (Compugen Limited) [23]. Some of the software packages can interact with automatic robotic systems to excise spots of interest from the gel for subsequent MS analysis. Interfacing the image analysis software with database tools for storing images is of critical importance in proteomic research, but is a shortfall of many commercial products [24]. Software tools such as

RADARS (Rapid Automated Data Archiving and Retrieval Software) are beginning to address this issue. The RADARS system automatically initiates database searches for protein identification from raw data files in addition to storing the processed data and search results in a relational database [25]. Current methods of analysis of MS data include peptide mass fingerprinting or peptide mass analysis [26-27], peptide sequence tag query [28], and MS/MS-ion search analysis [29]. Protein samples processed through MALDI-TOF MS can be identified in real time by a simultaneous search of a sequence database. Free software tools for database searches for protein identification are available through the internet. These include Mowse (United Kingdom Human Genome Mapping Project Resource Center), MS-FIT (University of California at San Francisco), PeptideSearch (European Molecular Biology Laboratory), and MS-TAG (University of California at San Francisco). Algorithms are developed to help in the determination of amino acid sequences of peptide fragments de novo from MS/MS data [30]. Protein data derived from two-dimensional electrophoretic analysis have been used to develop artificial learning models for the classification of tumors into benign, borderline, and malignant [31]. Statistical algorithms such as partial least squares and hierarchical clustering have been used to that effect. Algorithms are being developed that can cluster and distinguish cancers from healthy tissue samples through the use of training sets that contain protein profile spectra derived from the technologies described above. Bioinformatic tools are in development to handle, process, and meaningfully interpret the large body of data that is emerging from advances in proteomic technologies. These tools are essential in the discovery of sensitive and specific biomarkers in cancer research. In addition to this, other software packages for interpreting the biological significance of proteomic data are, for example, FatiGO and PathwayAssist. FatiGO is a web tool used to extract relevant Genes Ontology (GO) terms for a group of genes with respect to a set of genes of reference (typically the rest of genes). FatiGO can deal with thousands of genes from different organisms (currently human, mouse, *Drosophila*, worm, yeast, as well as genes whose proteins are included in Swissprot database), and can be queried using different gene identifiers (GenBank ID, Unigene, ENSEMBL, systematic name, Swissprot/TrEMBL). FatiGO addresses another common problem: the multiple ways in which genes are annotated. Different manufacturers of genomic platforms use distinct gene IDs. The most common gene IDs can be used as input for the application [32]. PathwayAssist is a software application developed for navigation and analysis of biological pathways, gene regulation networks and protein interaction maps. It comes with a database of molecular networks automatically assembled from scientific abstracts. It contains more than 100,000 events of regulation, interaction and modification between proteins, cell processes and small molecules. The database has been compiled by the application of the text-mining tool MedScan to the whole PubMed. MedScan preprocesses input text to extract relevant

sentences, which are subjected to natural language processing. The preprocessing step uses a manually curated dictionary of synonyms to recognize biological terms. Sentences that do not contain at least one matched term are filtered out. The natural language processing kernel deduces the syntactic structure of a sentence and establishes logical relationships between concepts. Finally, the results are matched against the functional ontology to produce the biological interpretation. PathwayAssist enables researchers to create their own pathways and produce publication quality pathway diagrams. For visualization purposes, pathways are represented as a graph with two types of nodes. The nodes of the first type are reserved for proteins, small molecules and cellular processes. The nodes of the second type (controls) represent events of functional regulation, chemical reactions and protein–protein interactions [33].

Confirmation and validation of protein of interest

Reliable protein identification depends on several variables, including the accuracy of fragment mass determination, the number of mass peaks submitted to databases query, the mass distribution of the query mass, the number of masses matching between samples and database proteins, and the number of modifications. In some cases, the data collected by peptide mass fingerprint are not sufficient for reliable identification of a protein, therefore confirmation and validation are needed. There are several ways for confirming identified proteins: generating an additional fingerprint with a protease of a different specificity, amino acid sequencing by Edman's method, amino acid composition, Western blot analysis, and immunohistochemical analysis.

1.2.1.1.1 LIMITATIONS OF 2D-PAGE

Despite various improvements of the 2-DE technique, the main inherent limitation of the method still remains: the fact that many proteins are expressed at such low levels that they will escape detection. For some applications, this problem can partly be overcome by increasing the resolving power of the first dimension separation by the use of narrow pH-range IPG strips (covering no more than 1 pH unit). Hoving *et al.* [34] and Westbrook *et al.* [35] note that “zoom” and “ultrazoom” gels are quite important for avoiding or at least minimizing the problem of spot overlapping in 2-DE, an ever present hazard in 2-DE maps [36-37].

Some highly abundant proteins are not detectable by 2-DE since they are not soluble in the detergent/urea buffer used in the first dimension, other proteins co-migrate with structural proteins and are therefore not detected. It is possible to identify low-abundance proteins by combining 2-DE with Western blotting.

Antibody cocktails can be used for simultaneously visualizing proteins that are of particular interest in a specific project (e.g., signal transduction components) [38]. Other tools are available to find low-abundance proteins and the two most representative are prefractionation techniques and “equalizer bead” technology.

Prefractionation techniques

Prefractionation techniques allow to discriminate between high-abundance and low-abundance components of the proteic mixture [39]. At present two major approaches have been described: chromatographic and electrophoretic. Fountoulakis' group has extensively developed the chromatographic prefractionation methodology; in a first procedure, Fountoulakis *et al.* [40] and Fountoulakis and Takács [41] adopted affinity chromatography on heparin gels as a pre-fractionation step for enriching certain protein fractions in the bacterium *Haemophilus influenzae*. The same group applied chromatographic prefractionation also by chromatofocusing on Polybuffer Exchanger, hydrophobic interaction chromatography (HIC) on a TSK Phenyl column [42] and hydroxyapatite chromatography [43], always on bacterial samples, with interesting results. A similar approach was also applied to the study of human foetal brain soluble proteins.

Alternative prefractionation approaches are based on electrophoretic techniques exploiting the same isoelectric focusing principle adopted in the first dimension of a 2-D map. Such approaches include the use of Multicompartment Electrolyzer and other devices for protein fractionation [44] and allow very high resolution.

Another possible prefractionation method deals with purification of specific clusters of proteins *via* subcellular fractionation. However, running multiple 2-DE gels is usually not feasible when performing expression proteomics of cancer tissues: many clinical samples do not contain sufficient amounts of protein to be analyzed on multiple 2-DE gels. Furthermore, subcellular fractionation is difficult to standardize, leading to reproducibility problems. The combination of subcellular organelle purification to enrich rare polypeptides and 2-DE separation has been widely described [45].

Equalizer beads

A novel approach for mining the “unseen proteome” that exploits solid phase ligand libraries was recently described and was termed “Equalizer Bead technology” [46-47]. Equalizer Beads simultaneously lower the concentration of high abundance proteins and enrich the concentration of low trace proteins without depletion of any specific proteins, in a high throughput manner. Each bead binds a specific binding partner and a wide diversity of ligands is available (typically millions). The population of beads has such diversity that a binding partner should exist for most, if not all, proteins present in a sample.

Importantly each bead has an equivalent binding capacity. Therefore the capacity for high abundance proteins is equal to that of low abundance proteins. When a complex sample is incubated with this bead library, each protein component of the mixture will find its binding partner. High abundance proteins saturate their binding partner and excess protein is washed away. In contrast low abundance proteins are “concentrated” on their specific affinity ligand. In this way the abundance of trace proteins is increased relative to the highly abundant proteins (see figure 4). After incubation the sample is eluted with each of the original components still represented in the eluant, but with a reduced range of concentrations. Unlike the use of depletion methods no fraction is discarded and proteins that may be bound to high abundance species like albumin are not lost.

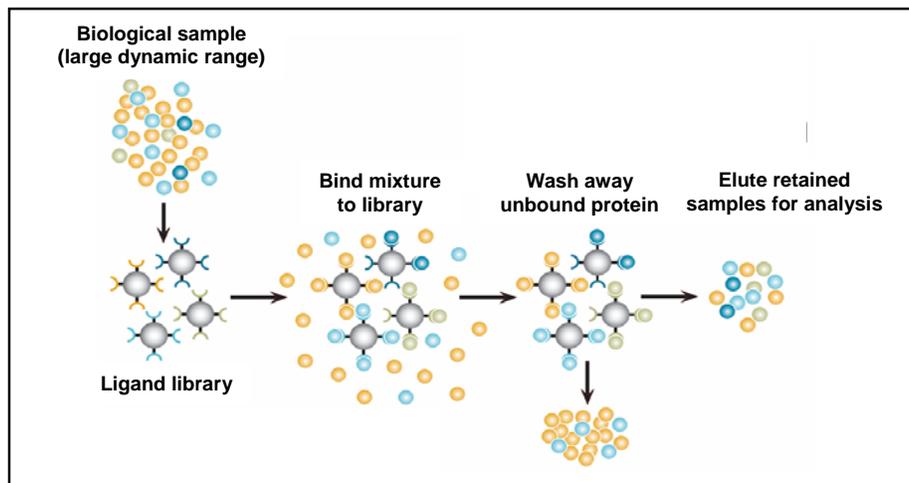


Figure 4. Equalizer bead technology: each bead has a different ligand, that has an affinity for specific proteins in the sample.

1.2.1.2 MULTI-DIMENSIONAL PROTEIN IDENTIFICATION TECHNOLOGY

In recent years, alternative approaches to the classical 2D-PAGE methodology, for protein separation and analysis have come to the limelight. For example, Yates and coworkers [48] demonstrated the use of shotgun identification of protein mixtures by proteolytic digestion followed by multidimensional liquid chromatography. The technique, dubbed MudPIT (multidimensional protein identification technology) consists in strong cation exchange chromatography coupled with reversed phase chromatography and thus, although called “multidimensional”, is just as two-dimensional as 2D-PAGE. There is a fundamental distinction, though, between the two approaches: whereas 2D-PAGE gives a display, in the charge/mass plane, of intact polypeptide chains, just as present in a cell or tissue, MudPIT offers a display of their tryptic digests, *i.e.* of the constituent peptides in which an intact protein is broken down prior to analysis. There are benefits and disadvantages claimed on both sides: whereas MudPIT allows analysis also of hydrophobic and very large proteins, 2D-PAGE, in turns, permits easy identification of all post-translational modifications, by which a single gene product gives rise to a number of different proteins with different functions. However, it should be mentioned that in contrast to the gel-derived proteome, for which quantitative data are available both intragel (*i.e.*, giving relative abundance data within a cell) and intergel (giving quantitative variations between different conditions), the MudPIT-derived proteome cannot accurately quantify proteins, but can only determine whether they are present or absent.

A sophisticated version of the MudPIT technology uses isotope coded affinity tags (ICAT) for facilitating comparison of the proteomes of two samples simultaneously [49]. ICAT is based on the reaction of protein mixtures with reagents with specificity toward sulfhydryl groups (Figure 5). These reagents contain a linker of variable size and a biotin affinity tag. Samples are separately reacted with reagents with linkers of two distinct sizes. The same peptides present in the two samples will differ slightly in molecular mass and can be separated by MS (Figure 6).

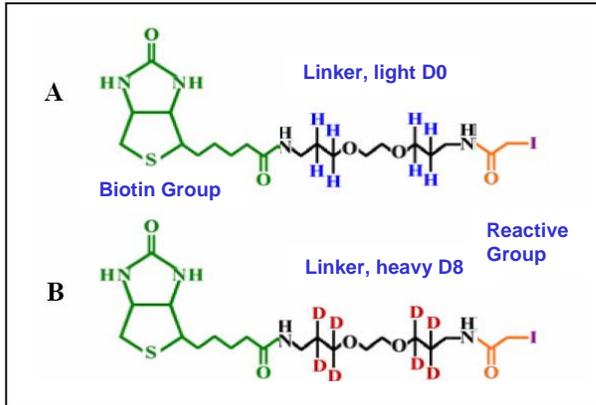
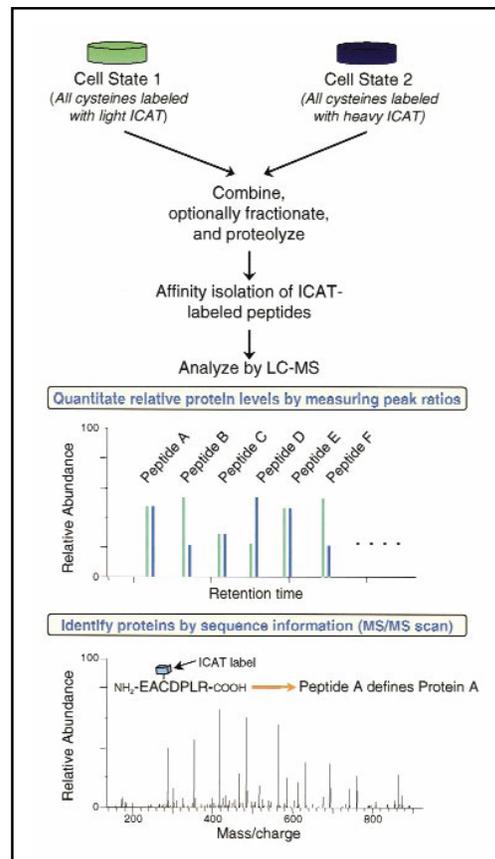


Figure 5. Structure of the ICAT reagent. The reagent consists of three elements: an affinity tag (biotin), which is used for isolating ICAT-labeled peptides; a linker that can incorporate stable isotopes; and a reactive group with specificity toward thiol groups (cysteines). The reagent exists in two forms, (A) light (contains no deuteriums) which is used for labelling the proteins in normal tissue and (B) heavy (contains eight deuteriums) which is used for labelling the proteins in disease tissue.

Figure 6. The ICAT strategy for quantifying differential protein expression. Two protein mixtures representing two different cell states have been treated with the isotopically light and heavy ICAT reagents, respectively; an ICAT reagent is covalently attached to each cysteinyl residue in every protein. Proteins from cell state 1 are shown in green, and proteins from cell state 2 are shown in blue. The protein mixtures are combined and proteolyzed to peptides, and ICAT-labeled peptides are isolated utilizing the biotin tag. These peptides are separated by microcapillary high-performance liquid chromatography. A pair of ICAT-labeled peptides are chemically identical and are easily visualized because they essentially coelute, and there is an 8 Da mass difference measured in a scanning mass spectrometer (four m/z units difference for a doubly charged ion). The ratios of the original amounts of proteins from the two cell states are strictly maintained in the peptide fragments. The relative quantification is determined by the ratio of the peptide pairs. Every other scan is devoted to fragmenting and then recording sequence information about an eluting peptide (tandem mass spectrum). The protein is identified by computer searching the recorded sequence information against large protein databases.



1.2.1.3 SELDI PROTEIN CHIP SYSTEM

Recent advances in proteomics offer opportunities for rapidly identifying new biomarkers or pattern of markers for the early detection, and diagnosis of cancer, and for monitoring the therapeutic efficacy and toxicity of treatments used to improve long-term survival of patients. The surfaced-enhanced laser desorption ionization/time-of-flight (SELDI-TOF) mass spectrometry is certainly one of the most promising approaches for achieving these goals. SELDI-TOF mass spectrometry differs from other types of mass spectrometry in that the proteins in a sample can be selectively retained on various chemical surfaces. Different protein chips are available (Figure 7), which allow the analysis of sub-populations of proteins according to their biochemical properties. The surface chemistries of the arrays include a series of classic chromatographic chemistries and specialized affinity capture surfaces. Classic chromatographic surfaces include normal phases for generic protein binding; hydrophobic surfaces for reversed-phase capture; cation- and anion-exchange surfaces; and immobilized metal affinity capture (IMAC) for metal-binding proteins. Specific proteins of interest can be covalently immobilized on pre-activated surface arrays, enabling customized experiments to investigate antibody-antigen, DNA-protein, receptor-ligand, and other molecular interactions.

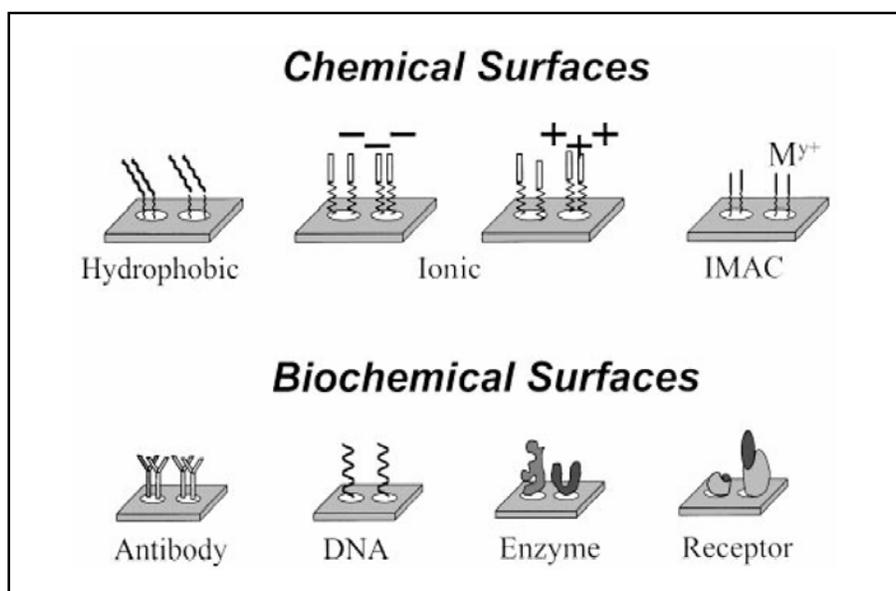


Figure 7. Activated surfaces for SELDI analysis

The basic steps in expression difference mapping by SELDI starts with fractionation, followed by sample binding to ProteinChip arrays. After the samples are bound to the arrays, the chips are washed and an energy absorbing molecule (EAM, also called matrix) is overlaid on the spots. After drying the matrix, the arrays are scanned on the ProteinChip reader using SELDI technology in which a nitrogen laser desorbs/ionizes the proteins and their molecular masses measured by TOF MS. Finally, the different spectra generated are analyzed using statistical algorithms such as Biomarker Wizard or Biomarker Patterns Softwares (Figure 8). Such instrumentation is rapidly been adopted in many clinical chemistry laboratories around the world and might soon become part of the standard instrumentation of such laboratories. The major applications of the ProteinChip System are: (1) expression profiling or proteomics; (2) purification development and monitoring; (3) protein characterization and identification; and (4) protein interaction assays (biochemical arrays). Numerous reports on its use for biomarker discovery in a variety of different tumors have been published [50]. In particular, SELDI-TOF has been applied to the study of several kind of cancers such as prostate [51], ovary [52], breast [53], lung [54], kidney [55], bladder [56], pancreas [57], liver [58], colon [59], endometrium [60], cervix [61], melanoma [62] and lymphoma [63].

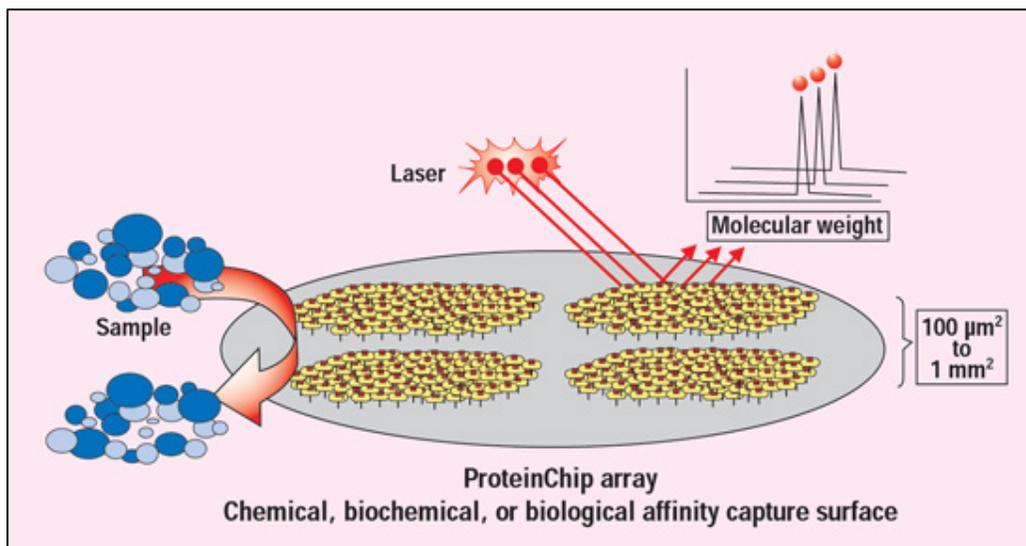


Figure 8. An illustration of the SELDI process and ProteinChip array. The sample goes directly onto the array, and proteins are captured, retained, and purified directly on the chip (affinity capture). The protein mass profile is read by SELDI, and retained proteins can be processed directly on the chip.

1.2.2 FUNCTIONAL PROTEOMICS

Proteins assemble themselves into networks through a variety of protein-protein interactions and post-translational modifications. In a cancer cell, the protein network is disrupted, deranged, or locally hyperactive compared with a normal cell. The cause of the dysregulated process may be at the genetic level (*e.g.*, activation of Ras or inactivation of p53 through mutations) or at the epigenetic level (*e.g.*, DNA hypermethylation at the promoter region). The central role of protein networks in a disease results in many opportunities and strategies for therapeutic intervention. Discovering proteins that play a role in cancer development and elucidating the exact protein linkages of the networks before, during, or after treatment will be of critical importance for drug development, correct target selection, and evaluation and monitoring of clinical therapy.

1.2.2.1 STUDY OF PROTEIN INTERACTIONS

One of the most effective approaches for proteomic investigation of protein interactions is direct analysis of multiprotein complexes using a combination of traditional techniques from biochemistry, molecular biology, and innovative proteomic methodology. Use of biochemical approaches followed by proteomic analysis can be applied not only to identification of the proteins involved in one-to-one interactions, but also to analysis of large complexes consisting of many proteins. Three main techniques are used for studying protein interactions: immunoprecipitation-based methods, the yeast two-hybrid system and protein microarrays.

Immunoprecipitation-based methods

With Immunoprecipitation (IP), the protein of interest is isolated from a complex mixture, such as a cell lysate, using an analyte-specific reagent (ASR; *e.g.*, antibody) along with its interacting partners (Figure 9). Alternatively, the protein of interest can be fused to a high affinity tag and isolated from the complex mixture by using the appropriate capture reagent. The use of IP allows endogenous proteins to be isolated without the need for cDNAs encoding the protein of interest or the need to express the fusion construct; however, for high-throughput applications this would require a specific ASR for every protein of interest, whereas tag-based affinity purifications using a single isolation chemistry can be applied to all proteins. To identify the binding partners, the proteins that associate with the index protein are often separated by gel electrophoresis and can be probed either with specific reagents on Western blots (if their identities are suspected and the corresponding reagents are available), or they can be digested before or after separation by specific proteases followed by analysis on a mass spectrometer [64].

This approach has the potential of capturing natural protein complexes and does not require previous knowledge on the interacting proteins or cloning of their genes. However, because all the proteins co-purify together, this method cannot determine which proteins are in direct contact with one another.

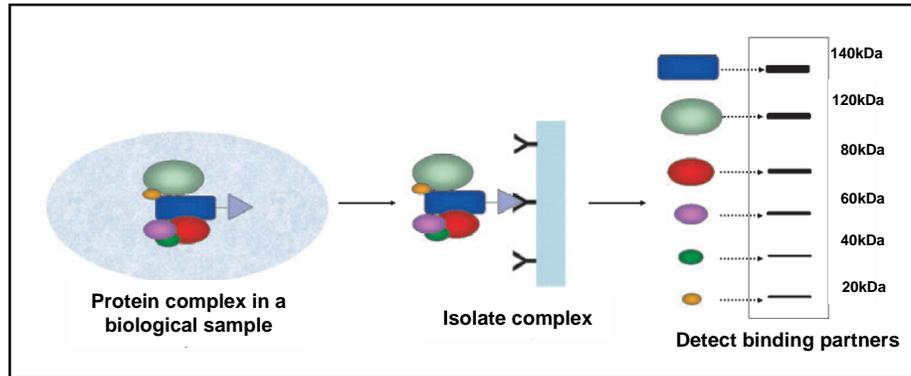


Figure 9. Immunoprecipitation. Proteins of interest (rectangle) can be isolated from complex biological sample by using antibodies specific to the protein or by modifying the protein with a tag (triangle). The protein of interest and its binding partners can then be separated based on charge, size or isoelectric point, and detected using antibodies specific to the interacting partners or by using mass spectrometry.

Yeast two-hybrid system

In contrast to IP-based method, which is primarily an *in vitro* biochemical method, the Yeast two-hybrid system (YTH) detects interactions *in vivo* in yeast cells. This is accomplished by measuring the signal from reporter genes, whose transcription is induced when their cognate transcription factors are reconstituted, by bringing together two functional halves through the interaction of the linked proteins [65]. In the two-hybrid assay (Figure 10), two fusion proteins are created: the protein of interest (X), which is constructed to have a DNA binding domain (BD) attached to its N-terminus, and its potential binding partner (Y), which is fused to an activation domain (AD). If protein X interacts with protein Y, the binding of these two will form an intact and functional transcriptional activator. This newly formed transcriptional activator will then go on to transcribe a reporter gene, which is simply a gene whose protein product can be easily detected and measured. In this way, the amount of the reporter produced can be used as a measure of interaction between our protein of interest and its potential partner. A variety of reporter systems have been adapted to the YTH system, usually expressing enzymes that either produce metabolites to support growth or induce color changes in specific substrates. The YTH systems specifically measure binary interactions, although the interactions must occur in yeast, and specifically in the context of the yeast nucleus, which may lead to false negatives, especially for some mammalian proteins. To address this, similar approaches, called mammalian two-hybrid systems, have been developed that

reconstitute active domains of reporter proteins, such as functional enzymes, ubiquitin, fluorescent proteins, and others to demonstrate the presence of an interaction [66-67]. Mammalian two-hybrid systems have been successfully used for monitoring protein interactions in the cytosol as well as among membrane-bound proteins, but challenges associated with the high-throughput introduction of DNA into mammalian cells, and the need for high-quality libraries of genes to test, have limited their widespread adoption [68].

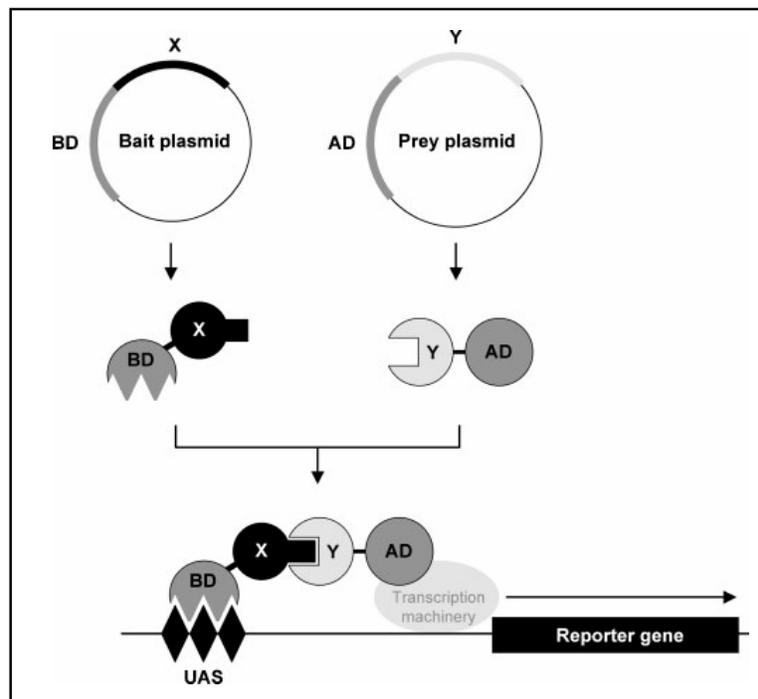


Figure 10. The principle of the yeast two-hybrid system. Two plasmids are constructed, the bait-encoding protein X fused to the C-terminus of a transcription factor DNA-binding domain (BD) and the prey-encoding protein Y fused to an activation domain (AD). Alternatively, the prey can consist of proteins encoded by an expression library. Each plasmid is introduced into an appropriate yeast strain either by co-transformation, sequential transformation, or by yeast mating. Only if proteins X and Y physically interact with one another are the BD and AD brought together to reconstitute a functionally active transcription factor that binds to upstream specific activation sequences (UAS) in the promoters of the reporter genes, and to activate their expression.

Protein microarrays

The idea of protein microarray is not new. In fact, the basics and theoretical considerations of protein microarrays were introduced in the 1980's by Ekins and colleagues [69-70]. There is considerable benefit to be gained from microarray technology [71]. First, in principle thousands of proteins can be spotted on a single slide or similar support, enabling researchers to interrogate simultaneously the function of many different proteins with minimal sample consumption. Second, hundreds or even thousands of copies of an array can be fabricated in parallel, enabling the same proteins to be probed repeatedly with many different molecules under numerous conditions [72]. Protein microarrays are finding their way into

both expression and functional proteomics [73]. In expression proteomics, protein-detecting microarrays comprise several different affinity reagents, such as antibodies, arrayed at high spatial density on a solid support. Each agent (or 'probe') captures its target protein from a complex mixture (e.g., serum or cell lysate), and the captured proteins are subsequently detected and quantified. Protein microarrays also provide a well-controlled, *in vitro*, way in which to study protein function on a genome-wide basis. For this application, sample proteins (or peptides), rather than affinity reagents, are arrayed on a solid support and then probed with compounds of interest (Figure 11).

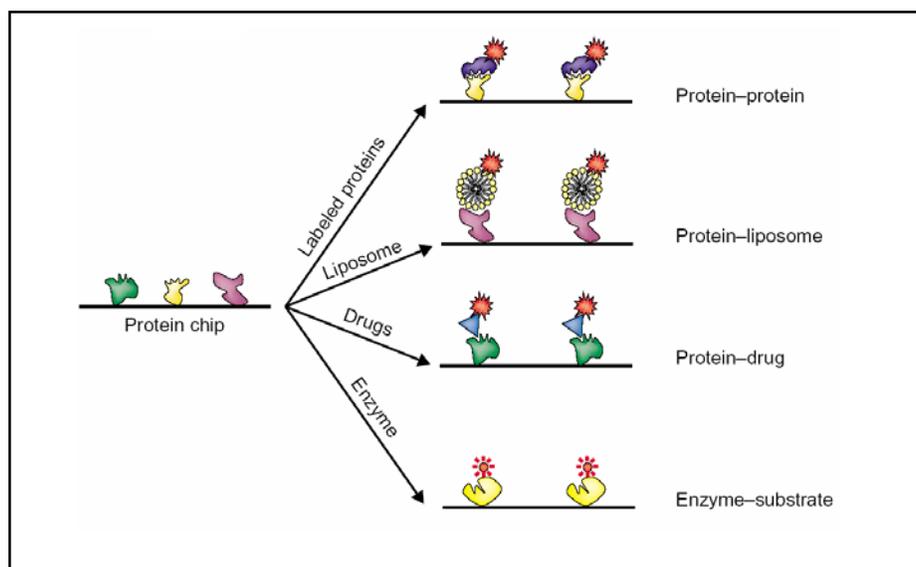


Figure 11. Functional protein chips are constructed by immobilizing large numbers of purified proteins on a solid surface. Protein chips have an enormous potential in assaying for a wide range of biochemical activities (e.g., protein-protein, protein-nucleic acid and enzyme-substrate interactions), as well as drug and drug target identification.

With this strategy, new substrates or ligands of known enzymes (or receptors or drugs), as well as disease related post-translational modifications, can be identified. An advantage of studying protein function in an array format is that the investigator can control the conditions of the experiment. This includes factors such as pH, temperature, ionic strength and the presence or absence of cofactors, as well as the modification states of the proteins under investigation. Furthermore, protein microarrays can be used to study the interaction of proteins with non-protein molecules, including nucleic acids, lipids and small organic compounds.

An extension of these methodologies is tissue microarray technology (TMA) – a high-throughput technique that allows the rapid visualization of one or a few molecular targets (proteins or nucleic acids) in hundreds or thousands of tissue specimens at a time. Arrays are generated by robotically spotting small cylinders (~0.6 mm³) of tissue derived from

individual paraffin-embedded specimens onto a slide. Proteins and nucleic acids are usually identified by immunohistochemistry and fluorescence in situ hybridization, respectively. The TMA technical principle is therefore similar to that of the direct assay version of antibody microarrays, with the fundamental difference that TMA provides investigators with precious information on target distribution at both the cell level (*e.g.*, cytoplasm vs nucleus) and the tissue level (*e.g.*, stromal vs tumor cells). Most applications of TMA can be found in the field of cancer research [74].

1.2.2.2 DETECTION OF POST-TRANSLATIONAL MODIFICATIONS

Post-translational modifications (PTMs) are covalent processing events that change the properties of a protein by proteolytic cleavage or by addition of a modifying group to one or more amino acids. Far from being mere “decorations”, PTMs of a protein can determine its activity state, localization, turnover, and interactions with other proteins. In signaling, for example, kinase cascades are turned on and off by the reversible addition and removal of phosphate groups, and in the cell cycle ubiquitination marks cyclins for destruction at defined time points [75].

Others examples of the biological effects of protein modifications are attachment of fatty acids for membrane anchoring and association, glycosylation for protein half-life, targeting, “cell-cell” and “cell-matrix” interactions. Consequently, the analysis of proteins and their post-translational modifications is particularly important for the study of diseases such as heart disease, cancer, neurodegenerative diseases and diabetes [76]. Presently, phosphorylation of proteins is the most studied and best understood PTM. However, the analysis of phosphoproteins and phosphopeptides is still one of the most challenging tasks in contemporary proteome research. Since not every phosphoprotein is accessible by a certain method and identification of the phosphorylated amino acid residue is required in the majority of cases, various strategies for the detection and localization of phosphorylations have been developed. Identification and localization of protein phosphorylations is mostly done by MS nowadays but phospho-proteins and -peptides are often suppressed in comparison to the unphosphorylated species if measured in complex mixtures. Thus, the isolation of pure phosphopeptide samples is a main task [77]. As PTMs cause a characteristic shift in mass, MS is, in principle, ideally suited for the identification of PTMs. However, sequence coverage of a protein by MS analysis is rarely complete, hence PTMs occurring outside the covered sequence will be missed. In addition, some PTMs reduce ionization efficiency and hence sensitivity, or are labile and are lost during fragmentation. The former disadvantage applies to phosphorylation and the latter to serine/threonine phosphorylation in particular, although this can be turned to advantage using precursor-ion scanning MS experiments [78]. Unfortunately, phosphorylation is one of the most common and important PTMs, and

improvements in the MS analysis of phosphorylation sites are being pursued vigorously [79]. The ability of phosphate groups to bind to chelated iron or gallium can be used to enrich phosphopeptides by IMAC. However, poor and non specific binding, and inefficient elution of the phosphopeptides warranted further improvements. Methylation of the peptide mixture before IMAC reduces the non-specific retention of acidic peptides [80], while improvements in the MS methodology have enhanced the sensitivity of detection of phosphorylated species [81-82]. Interesting variations of this kind of studies are the chemically replacement of the phosphate group with a biotin affinity tag in order to permit the selective isolation of the derivatized peptide [83], or to mark the phosphorylation sites by chemically converting phosphoserine and phosphothreonine residues into lysine analogues (aminoethylcysteine and β -methylaminoethylcysteine respectively), which can then be cleaved with a lysine-specific protease and detected by MS [84]. Although elegant, the chemical methods currently suffer from technical problems, such as the complexity and incompleteness of the chemical reactions, and the requirement of large amounts of starting material. Stains have been developed for visualizing phosphoproteins in gels or on blots [85], which in conjunction with glycosylation-specific staining can even be used to monitor changes in phosphorylation and glycosylation simultaneously [86]. These kind of stains are MS compatible and potentially very useful, but have not yet been widely tested.

1.3 CLINICAL APPLICATIONS OF CANCER PROTEOMICS

Clinical proteomics can have important applications that may directly change clinical practice by affecting critical elements of care and management. Proteomics may hold promise for early detection of disease using proteomic patterns of body fluid samples, diagnosis based on proteomic signatures as a complement to histopathology, individualized selection of therapeutic combinations that best target the entire disease-specific protein network, real-time assessment of therapeutic efficacy and toxicity, and rational modulation of therapy based on changes in the diseased protein network associated with drug resistance [87].

1.3.1 BIOMARKERS DISCOVERY FOR CANCER DIAGNOSIS AND PROGNOSIS

Early detection is critical in cancer control, prevention, and treatment. Cancer proteomics deals with the identification and quantitative analysis of differentially expressed proteins from normal tissues, premalignant and malignant tissues.

Clinical chemistry routinely makes very extensive and successful use of biomarkers (which is defined as “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathological processes, or pharmacological responses to a therapeutic intervention”), such as cholesterol, glucose levels etc., to help determine and monitor a patient’s state of health. Their interpretation is usually embedded in a diagnostic framework, mainly because very few of them are specific for a particular disease. The arrival of the genome age seems to have promised that we will be able to refine biomarker analysis to a level where it can, on its own, diagnose disease, predict the best form of therapy for an individual and monitor its effects. The challenge is formidable, and proteomics plays a key role in taking it on. Genomic changes are relatively static, whereas the proteome is dynamic, reflecting physiological and pathological changes much more acutely and accurately. Proteomic biomarker research focuses on the following main areas.

- (i) The identification of new targets for therapeutic intervention. This approach examines the tumor cells directly.
- (ii) The identification of markers that permit early detection of disease, better stratification and are of prognostic value. These molecules could be produced by the tumor itself, or reflect the host’s reaction to the tumor. This approach mostly examines body fluids. A crucial question is how large a tumor mass is detectable, and how does it compare with imaging methods, whose current detection limit is around one cubic centimetre of tumor volume. Studies with mice bearing human tumor xenografts suggest that the host response to the tumor predominates the

biomarker pattern, thus giving hope that even very small or disseminated tumors may be amplified and detectable through the host response.

- (iii) Markers for the monitoring of response to therapy. This aspect is of rapidly growing importance with the advent of non-cytotoxic drugs that cannot be monitored by clinical parameters alone, *e.g.*, angiogenesis or cell cycle inhibitors.

Discovery of low abundance proteins in various proteomes constitutes a major problem in biology research and biomarker discovery. A novel approach was recently investigated by Castagna *et al.* [88]. With the equalizer beads technology it was possible to dilute the most abundant species, with a concurrent concentration of the dilute and rare ones. Here, the concept was applied to the analysis of very dilute urinary proteins that normally escape to regular detection methods. By summing up the data with those already existing in the literature, it would appear that in urine, Castagna *et al.* could identify ca. 800 gene products, definitely a huge improvement from the early papers published in the 1980s, when just about a dozen proteins could be identified in 2-D maps, mostly of serum origin, and essentially only *via* immunoblots [89].

Biomarker research faces several big challenges. One is the biological variability of patients' samples, which reflect the compound picture of variables stemming from differences in age, sex, exercise, diet, circadian rhythm, etc. Furthermore, just as the 'single gene – single function' relationship does not exist, a 'single marker – single disease' correlation is very unlikely. Thus we have to look for diagnostic patterns, which are mass peaks in the case of MS-based approaches. This task requires the development of new bioinformatics tools that can extract such patterns from data full of biological noise and individual variability, reliably recognize them in different samples where the patterns may be slightly changed and eventually cross-correlate them with other parameters related to the patients' health and treatment status. Such efforts are underway [90-91], but more development work will be needed for generating robust software that can handle such data. In fact, a recent controversy concerning the validity of a highly acclaimed study on biomarkers for the early detection of ovarian cancer [52] is in great part a debate about data handling and statistical evaluation [92-93].

1.3.2 PHARMACOPROTEOMICS

Pharmacoproteomics is the use of proteomic technologies in drug discovery and development. The majority of drug targets are proteins. Proteomics has become an important tool for the discovery of new targets, and the further characterization of known drug targets with unknown function. Proteomic approaches for the analysis of protein expression and PTMs are also very useful for the investigation of the cellular responses to drugs, and for studying their mechanism of action and the basis of resistance [94]. Resistance to drug therapy is one of the major problems in cancer treatment and probably one of the main causes of limited long-term survival. Resistance may be acquired in response to a selection pressure by the drug treatment (acquired resistance) or expressed by cells already resistant and which will never respond to the drug treatment (intrinsic resistance). Understanding of the molecular mechanisms involved in drug resistance should lead to the improvement of the efficiency of existing drugs and to the identification of new therapeutic targets for overcoming therapeutic escape. Moreover, the identification of sets of markers of resistance could help clinicians to predict the response to chemotherapeutic agents. Here, the proteomic approach could be very helpful because of its capacity to analyse global protein changes in drug-sensitive and drug-resistant cells [95-96].

1.3.3 TUMOR IMMUNOLOGY

Many cancer patients mount an immune response, producing antibodies against their tumor. The targets of these antibodies could be useful for diagnostic and therapeutic purposes, and hence both genetic [97] and proteomic [98] approaches have been developed. A typical proteomics approach uses 2-DE to produce Western blots which are probed with patients' sera. Proteins selectively detected by cancer patients', but not control, sera are subsequently identified by MS. This method has identified antigens recognized by the immune system in cancers of the pancreas [99], kidney [100], lung [101], breast [102] and liver [103]. All studies succeeded in identifying potentially useful antigens, but their number was very small, and they represented abundant proteins, suggesting that the method could benefit from an enhancement of sensitivity. This may be achieved by using patients' sera for immunoprecipitating antigens before 2D-PAGE and MS analysis [104], resulting in a larger number of identified tumor-associated antigens (TAAs), although these were still mainly from highly abundant proteins.

1.4 BIBLIOGRAPHY

- [1] Wilkins, M. R., Sanchez, J. C., Gooley, A. A., Appel, R. D., Humphery-Smith, I., Hochstrasser, D. F., Williams, K. L., *Biotechnol. Genet. Eng. Rev.* 1996, **13**, 19-50.
- [2] Wilkins, M. R., Sanchez, J. C., Williams, K. L., Hochstrasser, D. F., *Electrophoresis* 1996, **17**, 830-838.
- [3] Reynolds, T., *J. Natl. Cancer Inst.* 2002, **94**, 1664-1666.
- [4] Kolch, W., Mischak, H., Pitt, A. R., *Clin. Sci. (Lond)* 2005, **108**, 369-383.
- [5] Alaiya, A. A., Franzen, B., Auer, G., Linder, S., *Electrophoresis* 2000, **21**, 1210-1217.
- [6] Anderson, L., Seilhamer, J., *Electrophoresis* 1997, **18**, 533-537.
- [7] Alaiya, A., Roblick, U., Egevad, L., Carlsson, A., Franzen, B., Volz, D., Huwendiek, S., Linder, S., Auer, G., *Anal. Cell Pathol.* 2000, **21**, 1-9.
- [8] Chen, G., Gharib, T. G., Wang, H., Huang, C. C., Kuick, R., Thomas, D. G., Shedden, K. A., Misek, D. E., Taylor, J. M., Giordano, T. J., Kardia, S. L., Iannettoni, M. D., Yee, J., Hogg, P. J., Orringer, M. B., Hanash, S. M., Beer, D. G., *Proc. Natl. Acad. Sci. U.S.A.* 2003, **100**, 13537-13542.
- [9] Nishizuka, S., Charboneau, L., Young, L., Major, S., Reinhold, W. C., Waltham, M., Kouros-Mehr, H., Bussey, K. J., Lee, J. K., Espina, V., Munson, P. J., Petricoin, E., 3rd, Liotta, L. A., Weinstein, J. N., *Proc. Natl. Acad. Sci. U.S.A.* 2003, **100**, 14229-14234.
- [10] Izzotti, A., Bagnasco, M., Cartiglia, C., Longobardi, M., De Flora, S., *Int. J. Oncol.* 2004, **24**, 1513-1522.
- [11] Pandey, A., Mann, M., *Nature* 2000, **405**, 837-846.
- [12] Wu, W., Hu, W., Kavanagh, J. J., *Int. J. Gynecol. Cancer.* 2002, **12**, 409-423.
- [13] Harry, J. L., Wilkins, M. R., Herbert, B. R., Packer, N. H., Gooley, A. A., Williams, K. L., *Electrophoresis* 2000, **21**, 1071-1081.
- [14] Von Eggeling, F., Gawriljuk, A., Fiedler, W., Ernst, G., Claussen, U., Klose, J., Romer, I., *Int. J. Mol. Med.* 2001, **8**, 373-377.
- [15] Tonge, R., Shaw, J., Middleton, B., Rowlinson, R., Rayner, S., Young, J., Pognan, F., Hawkins, E., Currie, I., Davison, M., *Proteomics* 2001, **1**, 377-396.
- [16] Zhou, G., Li, H., DeCamp, D., Chen, S., Shu, H., Gong, Y., Flaig, M., Gillespie, J. W., Hu, N., Taylor, P. R., Emmert-Buck, M. R., Liotta, L. A., Petricoin, E. F., 3rd, Zhao, Y., *Mol. Cell. Proteomics* 2002, **1**, 117-124.
- [17] Neuhoff, V., Arold, N., Taube, D., Ehrhardt, W., *Electrophoresis* 1988, **9**, 255-262.
- [18] Yan, J. X., Harry, R. A., Spibey, C., Dunn, M. J., *Electrophoresis* 2000, **21**, 3657-3665.
- [19] Shevchenko, A., Wilm, M., Vorm, O., Mann, M., *Anal. Chem.* 1996, **68**, 850-858.
- [20] Aebersold, R., Mann, M., *Nature* 2003, **422**, 198-207.
- [21] Bairoch, A., Apweiler, R., *Nucleic Acids Res.* 2000, **28**, 45-48.

- [22] Monardo, P. J., Boutell, T., Garrels, J. I., Latter, G. I., *Comput. Appl. Biosci.* 1994, **10**, 137-143.
- [23] Rosengren, A. T., Salmi, J. M., Aittokallio, T., Westerholm, J., Lahesmaa, R., Nyman, T. A., Nevalainen, O. S., *Proteomics* 2003, **3**, 1936-1946.
- [24] Chakravarti, D. N., Fiske, M. J., Fletcher, L. D., Zagursky, R. J., *Dev. Biol.* 2000, **103**, 81-90.
- [25] Field, H. I., Fenyo, D., Beavis, R. C., *Proteomics* 2002, **2**, 36-47.
- [26] Zhang, W., Chait, B. T., *Anal. Chem.* 2000, **72**, 2482-2489.
- [27] Perkins, D. N., Pappin, D. J., Creasy, D. M., Cottrell, J. S., *Electrophoresis* 1999, **20**, 3551-3567.
- [28] Fenyo, D., Qin, J., Chait, B. T., *Electrophoresis* 1998, **19**, 998-1005.
- [29] Eng, J. K., McCormack, A. L., Yates, JR III., *J. Am. Soc. Mass Spectrom.* 1994, **5**, 976-989.
- [30] Dancik, V., Addona, T. A., Clauser, K. R., Vath, J. E., Pevzner, P. A., *J. Comput. Biol.* 1999, **6**, 327-342.
- [31] Alaiya, A. A., Franzen, B., Hagman, A., Silfversward, C., Moberger, B., Linder, S., Auer, G., *Int. J. Cancer* 2000, **86**, 731-736.
- [32] Al-Shahrour, F., Diaz-Uriarte, R., Dopazo, J., *Bioinformatics* 2004, **20**, 578-580
- [33] Nikitin, A., Egorov, S., Daraselia, N., Mazo, I., *Bioinformatics* 2003, **19**, 2155-2157.
- [34] Hoving, S., Voshol, H., van Oostrum, J., *Electrophoresis* 2000, **21**, 2617-2621.
- [35] Westbrook, J. A., Yan, J. X., Wait, R., Welson, S. Y., Dunn, M. J., *Electrophoresis* 2001, **22**, 2865-2871.
- [36] Pietrogrande, M. C., Marchetti, N., Dondi, F., Righetti, P. G., *Electrophoresis* 2003, **24**, 217-224.
- [37] Campostrini, N., Areces, L. B., Rappsilber, J., Pietrogrande, M. C., Dondi, F., Pastorino, F., Ponzoni, M., Righetti, P. G., *Proteomics* 2005, **5**, 2385-2395.
- [38] Sanchez, J. C., Wirth, P., Jaccoud, S., Appel, R. D., Sarto, C., Wilkins, M. R., Hochstrasser, D. F., *Electrophoresis* 1997, **18**, 638-641.
- [39] Righetti, P. G., Castagna, A., Herbert, B., Reymond, F., Rossier, J. S., *Proteomics* 2003, **42**, 269-321.
- [40] Fountoulakis, M., Langen, H., Evers, S., Gray, C., Takàcs, B., *Electrophoresis* 1997, **18**, 1193-1202.
- [41] Fountoulakis, M., Takàcs, B., *Protein Expr Purif.* 1998, **14**, 13-19.
- [42] Fountoulakis, M., Takàcs, M. F., Takàcs, B., *J. Chromatogr. A* 1999, **833**, 157-168.
- [43] Karlsson, K., Cairns, N., Lubec, G., Fountoulakis, M., *Electrophoresis* 1999, **20**, 2970-2976.

- [44] Righetti, P. G., Castagna, A., Antonioli, P., Boschetti, E., *Electrophoresis* 2005, **26**, 297-319.
- [45] Corthals, G. L., Wasinger, V. C., Hochstrasser, D. F., Sanchez, J. C., *Electrophoresis* 2000, **21**, 1104-1115.
- [46] Thulasiraman, V., Lin, S., Gheorghiu, L., Lathrop, J., Lomas, L., Hammond, D., Boschetti, E., *Electrophoresis* 2005, **26**, 3561-3571.
- [47] Righetti, P. G., Castagna, A., Antonucci, F., Piubelli, C., Cecconi, D., Campostrini, N., Rustichelli, C., Antonioli, P., Zanusso, G., Monaco, S., Lomas, L., Boschetti, E., *Clin. Chim. Acta* 2005, **357**, 123-139.
- [48] Washburn, M. P., Wolters, D., Yates, J. R., 3rd, *Nat. Biotechnol.* 2001, **19**, 242-247.
- [49] Gygi, S. P., Han, D. K., Gingras, A. C., Sonenberg, N., Aebersold, R., *Electrophoresis* 1999, **20**, 310-319.
- [50] Yip, T. T., Lomas, L., *Technol. Cancer Res. Treat.* 2002, **1**, 273-280.
- [51] Gretzer, M. B., Chan, D. W., van Rootselaar, C. L., Rosenzweig, J. M., Dalrymple, S., Mangold, L. A., Partin, A. W., Veltri, R. W., *Prostate* 2004, **60**, 325-331.
- [52] Petricoin, E. F., Ardekani, A. M., Hitt, B. A., Levine, P. J., Fusaro, V. A., Steinberg, S. M., Mills, G. B., Simone, C., Fishman, D. A., Kohn, E. C., Lotta, L. A., *Lancet* 2002, **359**, 572-577.
- [53] Li, J., Zhao, J., Yu, X., Lange, J., Kuerer, H., Krishnamurthy, S., Schilling, E., Khan, S. A., Sukumar, S., Chan, D. W., *Clin. Cancer Res.* 2005, **11**, 8312-8320.
- [54] Yang, S. Y., Xiao, X. Y., Zhang, W. G., Zhang, L. J., Zhang, W., Zhou, B., Chen, G., He, D. C., *BMC Cancer* 2005, **5**, 83.
- [55] Junker, K., Gneist, J., Melle, C., Driesch, D., Schubert, J., Claussen, U., Von Eggeling, F., *Int. J. Mol. Med.* 2005, **15**, 285-290.
- [56] Zhang, Y. F., Wu, D. L., Guan, M., Liu, W. W., Wu, Z., Chen, Y. M., Zhang, W. Z., Lu, Y., *Clin. Biochem.* 2004, **37**, 772-779.
- [57] Koopmann, J., Zhang, Z., White, N., Rosenzweig, J., Fedarko, N., Jagannath, S., Canto, M. I., Yeo, C. J., Chan, D. W., Goggins, M., *Clin. Cancer Res.* 2004, **10**, 860-868.
- [58] Paradis, V., Degos, F., Dargere, D., Pham, N., Belghiti, J., Degott, C., Janeau, J. L., Bezeaud, A., Del forge, D., Cubizolles, M., Laurendeau, I., Bedossa, P., *Hepatology* 2004, **41**, 40-47.
- [59] Wollscheidt, V., Wellmann, A., *Anticancer Res.* 2004, **24**, 1791-1796.
- [60] Yang, E. C., Guo, J., Diehl, G., DeSouza, L., Rodrigues, M. J., Romaschin, A. D., Colgan, T. J., Siu, K.W., *J. Proteome. Res.* 2004, **3**, 636-643.
- [61] Wong, Y. F., Cheung, T. H., Lo, K.W., Wang, V. W., Chan, C. S., Ng, T. B., Chung, T. K., Mok, S. C., *Cancer Lett.* 2004, **211**, 227-234.

- [62] Wilson, L. L., Tran, L., Morton, D. L., Hoon, D. S., *Ann. N. Y. Acad. Sci.* 2004, **1022**, 317-322.
- [63] Lin, Z., Jenson, S. D., Lim, M. S., Elenitoba-Johnson, K. S., *Mod. Pathol.* 2004, **17**, 670-678.
- [64] Figeys, D., McBroom, L. D., Moran, M. F., *Methods* 2001, **24**, 230-239.
- [65] Fields, S., Song, O., *Nature* 1989, **340**, 245-246.
- [66] Michnick, S. W., Remy, I., Campbell-Valois, F. X., Vallee-Belisle, A., Pelletier, J.N., *Methods Enzymol.* 2000, **328**, 208-230.
- [67] Wehrman, T., Kleaveland, B., Her, J. H., Balint, R. F., Blau, H. M., *Proc. Natl. Acad. Sci. U.S.A.* 2002, **99**, 3469-3474.
- [68] Suzuki, H., Fukunishi, Y., Kagawa, I., Saito, R., Oda, H., Endo, T., Kondo, S., Bono, H., Okazaki, Y., Hayashizaki, Y., *Genome Res.* 2001, **11**, 1758-1765.
- [69] Ekins, R. P., *J. Pharm. Biomed. Anal.* 1989, **7**, 155.
- [70] Ekins, R. P., Chu, F. W., *Trends in Biotechnology* 1999, **17**, 217-218.
- [71] Mocellin, S., Rossi, C. R., Traldi, P., Nitti, D., Lise, M., *TRENDS in Molecular Medicine* 2004, **10**, 24-32.
- [72] MacBeath, G., *Nat. Genet.* 2002, **32**, 526-532.
- [73] Phizicky, E., Bastiaens, P. I., Zhu, H., Snyder, M., Fields, S., *Nature* 2003, **422**, 208-215.
- [74] Kallioniemi, O. P., Wagner, U., Kononen, J., Sauter, G., *Hum. Mol. Genet.* 2001, **10**, 657-662.
- [75] Mann, M., Jensen, O. N., *Nature Biotechnology* 2003, **21**, 255-261.
- [76] Baumann, M., Meri, S., *Expert Rev. Proteomics* 2004, **1**, 207-17.
- [77] Reinders, J., Sickmann, A., *Proteomics* 2005, **5**, 4052-4061.
- [78] Areces, L. B., Matafora, V., Bachi, A., *Eur. J. Mass Spectrom.* 2004, **10**, 383-392.
- [79] Salih, E., *Mass Spectrom. Rev.* 2005, **24**, 828-846.
- [80] Ficarro, S. B., McClelland, M. L., Stukenberg, P. T., Burke, D. J., Ross, M. M., Shabanowitz, J., Hunt, D. F., White, F. M., *Nat. Biotechnol.* 2002, **20**, 301-305.
- [81] Stensballe, A., Jensen, O. N., *Rapid Commun. Mass Spectrom.* 2004, **18**, 1721-1730.
- [82] Jin, W. H., Dai, J., Zhou, H., Xia, Q. C., Zou, H. F., Zeng, R., *Rapid Commun. Mass Spectrom.* 2004, **18**, 2169-2176.
- [83] Goshe, M. B., Veenstra, T. D., Panisko, E. A., Conrads, T. P., Angell, N. H., Smith, R. D., *Anal. Chem.* 2002, **74**, 607-616.
- [84] Knight, Z. A., Schilling, B., Row, R. H., Kenski, D. M., Gibson, B. W., Shokat, K. M., *Nat. Biotechnol.* 2003, **21**, 1047-1054.
- [85] Steinberg, T. H., Agnew, B. J., Gee, K. R., Leung, W. Y., Goodman, T., Schulenberg, B., Hendrickson, J., Beechem, J. M., Haugland, R. P., Patton, W. F., *Proteomics* 2003, **3**, 1128-1144.

- [86] Schulenberg, B., Beechem, J. M., Patton, W. F., *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 2003, **793**, 127-139.
- [87] Calvo, K. R., Lotta, L. A., Petricoin, E. F., *Biosci. Rep.* 2005, **25**, 107-125.
- [88] Castagna, A., Cecconi, D., Sennels, L., Rappsilber, J., Guerrier, L., Fortis, F., Boschetti, E., Lomas, L., Righetti, P. G., *J. Proteome Res.* 2005, **4**, 1917-1930.
- [89] Edwards, J. J., Tollaksen, S. L., Anderson, N. G., *Clin. Chem.* 1982, **28**, 941-948.
- [90] Ball, G., Mian, S., Holding, F., Allibone, R. O., Lowe, J., Ali, S., Li, G., McCardle, S., Ellis, I. O., Creaser, C., Rees, R. C., *Bioinformatics* 2002, **18**, 395-404.
- [91] Rogers, M. A., Clarke, P., Noble, J., Munro, N. P., Paul, A., Selby, P. J., Banks, R. E., *Cancer Res.* 2003, **63**, 6971-6983.
- [92] Kolch, W., Mischak, H., Chalmers, M. J., Pitt, A., Marshall, A. G., *Rapid Commun. Mass Spectrom.* 2004, **18**, 2365-2366.
- [93] Fung, E. T., Weinberger, S. R., Gavin, E., Zhang, F., *Expert Rev. Proteomics* 2005, **2**, 847-862.
- [94] Kramer, R., Cohen, D., *Nat. Rev. Drug Discov.* 2004, **3**, 965-972.
- [95] Chapal, N., Molina, L., Molina, F., Laplanche, M., Pau, B., Petit, P., *Fundam. Clin. Pharmacol.* 2004, **18**, 413-422.
- [96] Righetti, P. G., Castagna, A., Antonioli, P., Cecconi, D., Campostrini, N., Righetti, S. C., *Expert Rev. Proteomics* 2005, **2**, 215-228.
- [97] Tureci, O., Usener, D., Schneider, S., Sahin, U., *Methods Mol. Med.* 2005, **109**, 137-154.
- [98] Imafuku, Y., Omenn, G. S., Hanash, S., *Dis. Markers* 2004, **20**, 149-153.
- [99] Hong, S. H., Misek, D. E., Wang, H., Puravs, E., Giordano, T. J., Greenson, J. K., Brenner, D. E., Simeone, D. M., Logsdon, C. D., Hanash, S. M., *Cancer Res.* 2004, **64**, 5504-5510.
- [100] Unwin, R. D., Harnden, P., Pappin, D., Rahman, D., Whelan, P., Crafen, R. A., Selby, P. J., Banks, R. E., *Proteomics* 2003, **3**, 45-55.
- [101] Brichory, F., Beer, D., Le Naour, F., Giordano, T., Hanash, S., *Cancer Res.* 2001, **61**, 7908-7912.
- [102] Le Naour, F., Misek, D. E., Krause, M. C., Deneux, L., Giordano, T. J., Scholl, S., Hanash, S. M., *Clin. Cancer Res.* 2001, **7**, 3328-3335.
- [103] Shalhoub, P., Kern, S., Girard, S., Beretta, L., *Dis. Markers* 2001, **17**, 217-223.
- [104] Gires, O., Munz, M., Schaffrik, M., Kieu, C., Rauch, J., Ahlemann, M., Eberle, D., Mack, B., Wollenberg, B., Lang, S., Hofmann, T., Hammerschmidt, W., Zeidler, R., *Cell. Mol. Life Sci.* 2004, **61**, 1198-1207.

CHAPTER 2

NEUROBLASTOMA

2.1 INTRODUCTION

Neuroblastoma (NB), together with lymphoma, osteosarcoma, Ewing's tumors, rhabdomyosarcoma and lymphoblastic leukemia, belongs to a group of undifferentiated paediatric malignancies known as the small round-cell tumors of childhood. NB, the most common extracranial solid tumor of infancy and childhood, is an embryonal tumor. It arises from primitive neuroepithelial cells of the neural crest and occurs in the adrenal medulla or paraspinal sympathetic ganglia of the abdomen, chest or neck [1]. The neural crest is a transitory structure that arises during the closure of the neural tube. The neural crest is a group of embryonic cells that disengages from the neural plate during neurulation and immigrate to give lineage to specialized adult cells (for example, the glial cells of the peripheral nervous system, medullary cells of the adrenal gland, calcitonin producing cells of thyroid, pigmented cells and mesectodermal derivatives) (Figure 1).

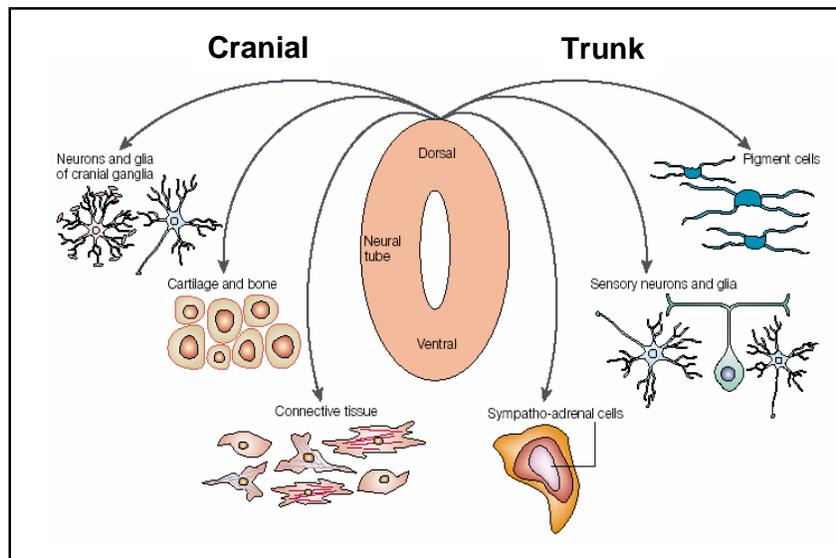


Figure 1. Specialized adult cells derived from the migration of the neural crest cells.

2.2 CLINICAL CHARACTERISTICS

The disease accounts for approximately 6% of childhood cancers, with an annual incidence of eight million children under the age of 15. The median age at diagnosis is 22 months, and more than 95% of cases present by 10 years of age [2-3]. Unfortunately, in the majority of patients Neuroblastoma is metastatic at the time of diagnosis, and it usually undergoes rapid progression with fatal outcome.

Tumor stage and patient age at diagnosis correlate strongly with survival [4-5]. NB can form relatively benign, localised and well-differentiated tumors that are successfully treated by surgical resection alone (stage I or II) or locally invasive (stage III) and metastatic (stage IV) tumors that are associated with a bad clinical outcome [6]. Distant metastases commonly occur in the regional lymph nodes, liver, bone marrow and bones [7] (Figure 2).

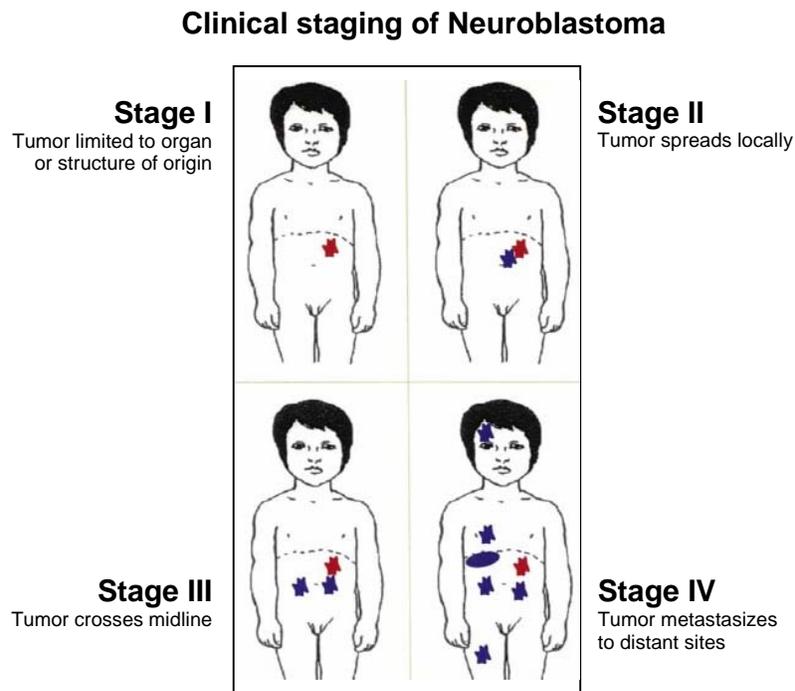


Figure 2: Clinical stages of human Neuroblastomas

However, in children less than one year old, metastases limited to the bone marrow, liver or skin, but absent from the bones (stage IV S) are associated with a favourable outcome. In the same patients, the tumor may regress spontaneously, *via* differentiation into benign ganglioneuroma [8]. For children older than 1 year with metastatic disease, the outcome is usually fatal. This wide range of clinical variability reflects NB's biological heterogeneity.

NB metabolizes catecholamines and, as a result, excretes various markers of tumor activity, including vanillylmandelic acid (VMA) or homovanillic acid (HVA), markers for which quantitation ratios can indicate tumor differentiation and may therefore even signify prognosis [9].

2.3 OVERVIEW OF NEUROBLASTOMA GENETICS

Neuroblastoma is typically a sporadic tumor, and only few isolated instances of familial clustering have been observed [10-11]. Neuroblastoma tumor cells are characterised by a wide diversity of genetic mutations. Some common genetic features include:

- **Amplification of the *N-myc* gene** is one of the most established genetic prognostic factors. The *N-myc* gene is located on the distal short arm of chromosome 2 (2p24). The process of amplification usually results in 50 to 400 copies of the gene per cell, with correspondingly high levels of protein expression (Figure 3). The amplification of *N-myc* proto-oncogene has been found in about 25%-30% of primary Neuroblastomas from untreated patients and it is predominantly associated with advanced stages of disease [12-13] and it is used as prognostic value, being associated with rapid tumor progression and poor prognosis [14]. The possible role of *N-myc* in tumorigenesis is also suggested by the observation that a decrease of *N-myc* expression is associated with the morphological differentiation induced by retinoic acid *in vitro*, and the finding that *N-myc* amplification is detected in very few IV-S stage tumors which often regress spontaneously [15]. The genomic region amplified with *N-myc* is large, therefore it has been postulated that additional genes located near *N-myc* may contribute to the ultimate tumor phenotype when coamplified. The RNA helicase gene *DDX1* maps within 300 kb 5' of *N-myc* and is coamplified in approximately 40% to 50% of Neuroblastomas with *N-myc* amplifications [16]. Nevertheless, *DDX1* amplification has not been identified in the absence of *N-myc* amplification. Thus, *N-myc* is primarily responsible for the aggressive nature of Neuroblastomas with amplification at the 2p24.1 locus, but *DDX1* may contribute to the highly malignant nature of some *N-myc* amplified tumors.

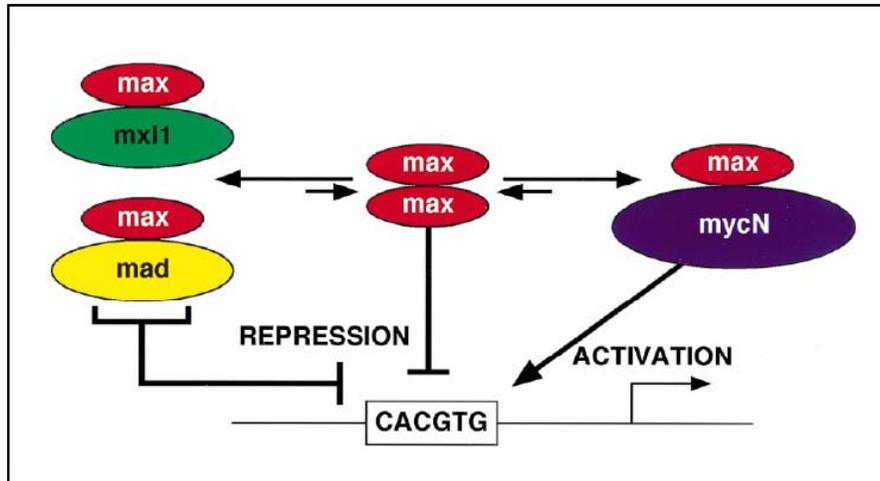


Figure 3: Model of MycN interactions with other proteins. Genomic amplification of MYCN leads to increased MycN nuclear protein. Heterodimerization of MycN and Max is favored in the presence of excess MycN (arrows). Transcriptional activation of target genes occurs after interaction of the MycN/Max complex with the promoter region through an E-box motif (CACGTG). In the absence of MycN, it is postulated that the transcriptional repressing protein complexes Max/Max, Max/Mad, and Max/Mxi1 predominate and function to repress transcription.

- Deletion of material from the chromosome 1p36 region** is also associated with adverse prognosis. Alterations of 1p characterize a wide range of human malignancies, including both solid tumors and hematological cancers [17]. Cytogenetic analysis of Neuroblastoma tumors has shown deletion of the distal short arm of chromosome 1 (1p36.1-1p36.3) in 70% of cases. This abnormality was first described by Brodeur *et al.* [18] as the most characteristic cytogenetic abnormality in human Neuroblastoma and was subsequently confirmed in 70%-80% of Neuroblastoma karyotyped. It has been recently suggested that two distinct tumor suppressor genes map in the 1p35-36 region and one of these would be associated with *N-myc* amplification [19].
- Gain of 17q material** is the most frequent genetic abnormality in Neuroblastoma. The frequent occurrence of 17q translocations and the significant association with poor overall survival strongly suggest that the region of chromosome 17 gain includes a gene, or genes, critical for tumor progression. If so, there are in principle two possible ways in which such a gene could be affected by these rearrangements: either the unbalanced translocation disrupts a gene close to the breakpoint or the gain of 17q material could alter the gene dosage of one or more genes localized distal to the breakpoints. Several candidate genes have been proposed to be responsible for the 17q gain effect on tumor growth characteristics. In principle, each

gene within the translocated 17q segment is a candidate, making the definitive proof for each candidate's contribution almost impossible. Survivin, an anti-apoptosis protein, mapped to 17q25, is one candidate as its expression correlates strongly with adverse clinical factors, such as age and stage [20]. Each translocation of 17q material is associated with a deletion event on a partner chromosome, most frequently chromosome 1p, followed by 11q. Deleted chromosome regions may contribute substantially to tumor phenotype. Unbalanced 17q gain is an adverse prognostic factor and is strongly associated with adverse clinical features, 1p deletion, and *N-myc* amplification.

- **Deletion of material from the chromosome 11q** is also associated with adverse prognosis. Cytogenetic analyses have reported 11q deletion in about 15% of Neuroblastomas [21]. Functional evidence for a tumor suppressing effect of chromosome 11 was demonstrated by transfer of an intact chromosome 11 into the Neuroblastoma cell line NGP inducing differentiation [22]. Deletions of 11q often have been seen in concert with a gain of 17q. After 1p, chromosome arm 11q is the second most common partner for 17q translocations [23]. Such translocations, resulting in concurrent loss of distal 11q and gain of 17q, account for approximately half of the 11q deletion.
- **Differential expression of the neurotrophin receptors** is strongly correlated with the biologic and clinical features of Neuroblastomas. *TRKA* expression is inversely related to disease stage and *N-myc* amplification status [24]. Thus, high *TRKA* expression is a marker of “favorable” Neuroblastomas and is correlated with an increased probability of disease survival. In contrast, full-length *TRKB* is expressed preferentially in advanced stage, *N-myc* amplified Neuroblastomas [25]. *TRKC* is expressed in favorable Neuroblastomas, essentially all of which also express *TRKA* [26].
- **Evidence for the prognostic value of ploidy in Neuroblastoma** comes from flow cytometric and cytogenetic analyses. Flow cytometric analysis can easily be used for determining cellular DNA content. Although this analysis cannot detect specific chromosome rearrangements, such as deletions, translocations, or even gene amplification, it has been shown to correlate with biological behavior of Neuroblastoma tumors, at least in a subset of patients. By using flow cytometric analysis, it was demonstrated that, in infants, DNA content was significantly linked to

tumor stage, with diploidy much more frequent in advanced tumor stages [27]. The DNA content was also shown to discriminate between good and poor responders to chemotherapy. Hyperdiploidy is mainly observed in low stage tumors of younger patients with a favorable clinical outcome, whereas diploid tumors are associated with advanced tumor growth and significantly reduced survival probability [28]. Diploidy was closely linked to advanced stages and the DNA content lost its prognostic significance for patients over 2 years of age [29].

2.4 DIAGNOSIS

Physicians often discover Neuroblastoma as a lump or tenderness in the child's abdomen or during a routine visit to the doctor. A simple urine test may help to detect Neuroblastoma. Other studies used to identify Neuroblastoma include:

- **Computed tomography (CT)** creates a computer-generated two-dimensional image of the tumor and surrounding tissue that includes greater detail than standard X-rays.
- **Magnetic resonance imaging (MRI)** uses a magnetic field and radio waves for generating cross-sectional pictures of the body.
- **Ultrasound** uses a wandlike device (transducer) to send and receive high-frequency sound waves that can be translated into images of internal organs.
- A **chest X-ray** displays the chest, lungs, heart, large arteries, ribs, and the diaphragms so that physicians can look for abnormalities or tumors.
- A **bone scan** shows changes in bones that may indicate a tumor or other abnormality.
- A **biopsy** provides tissue samples that can be analyzed under a microscope for cancer.
- **Blood tests** can reveal abnormalities such as anemia or an elevated white blood cell count.

2.5 NEUROBLASTOMA TREATMENT

2.5.1 STANDARD TREATMENT

- **Surgery** is typically used to treat Neuroblastoma. Depending on the localization of the tumor and whether it has spread, as much of the tumor as possible will be removed. If the tumor cannot be removed, a biopsy may be done instead.
- **Radiation therapy** is a cancer treatment that uses high-energy x-rays or other types of radiation for killing cancer cells. There are two types of radiation therapy. External radiation therapy uses a machine outside the body for sending radiation toward the cancer. Internal radiation therapy uses a radioactive substance sealed in needles, seeds, wires, or catheters that are placed directly into or near the cancer. The way the radiation therapy is given depends on the type and stage of the cancer being treated.
- **Chemotherapy** is a cancer treatment that uses drugs to stop the growth of cancer cells, either by killing the cells or by stopping the cells from dividing. When chemotherapy is taken by mouth or injected into a vein or muscle, the drugs enter the bloodstream and can reach cancer cells throughout the body (systemic chemotherapy). When chemotherapy is placed directly into the spinal column, an organ, or a body cavity such as the abdomen, the drugs mainly affect cancer cells in those areas (regional chemotherapy). The way the chemotherapy is given depends on the type and stage of the cancer being treated. The use of two or more anticancer drugs is called combination chemotherapy.
- **Watchful waiting** is closely monitoring a patient's condition without giving any treatment until symptoms appear or change.

2.5.2 OTHER TYPES OF TREATMENT

- **High-dose chemotherapy and radiation therapy with bone marrow transplant or stem cell transplant** is a method of giving high doses of chemotherapy and radiation therapy and replacing blood-forming cells destroyed by the cancer treatment. Stem cells (immature blood cells) are removed from the blood or bone marrow of the patient or a donor and are frozen and stored. After chemotherapy and radiation therapy are completed, the stored stem cells are thawed and given back to the patient through an infusion. These reinfused stem cells grow into (and restore) the body's blood cells.
- **Monoclonal antibody therapy** is a cancer treatment that uses antibodies made in the laboratory, from a single type of immune system cell. These antibodies can identify substances on cancer cells or normal substances that may help cancer cells grow. The antibodies attach to the substances and kill the cancer cells, block their growth, or keep them from spreading. Monoclonal antibodies are given by infusion. They may be used alone or for delivering drugs, toxins, or radioactive material directly to cancer cells. Among the new avenues for NB treatment that are being explored, immunotherapy has attracted much interest. Emphasis has been placed on monoclonal antibodies directed to tumor-associated antigens – in particular the disialoganglioside GD2 – that have been tested in the clinical setting with promising results. In addition, stimulation of cell-mediated antitumor effector mechanisms have been attempted – for example, by recombinant interleukin (IL)-2 administration [30].
- **Peptide nucleic acid (PNA)** are nucleic-acid-based gene-specific drugs. PNA are DNA analogs in which the sugar-phosphate backbone is replaced by a pseudopeptide chain constituted by *N*-(2-aminoethyl)glycine monomers covalently bonded to DNA bases [31]. PNAs form highly stable duplexes with complementary DNA and RNA strands; they are more stable than oligonucleotides due to their resistance to degradation by nucleases and proteases. PNAs have been shown to exert antisense inhibitory effects *in vitro* on important tumor proteins, such as PML-RAR-alpha and BCL-2 [32-33]. Recently, Sun *et al.* [34] and Pession *et al.* [35] reported remarkable results utilizing a PNA-based antisense strategy for inhibition of MYCN production in Neuroblastoma cells. Interestingly, PNAs designed to target the DNA coding strand (and thus not mRNA) have also shown anti-gene capacity both *in vitro* and *in vivo*, whether conjugated or not to nuclear carriers. Tonelli *et al.* [36] developed an anti-

gene peptide nucleic acid (PNA) for selective inhibition of MYCN transcription in Neuroblastoma cells, targeted against a unique sequence in the antisense DNA strand of exon 2 of MYCN and linked at its NH₂ terminus to a nuclear localization signal peptide. Selective activity of the PNA was shown by altering three point mutations, and by the observation that an anti-gene PNA targeted against the noncoding DNA strand did not exert any effect. These findings could encourage the development of an anti-gene PNA-based tumor-specific agent for Neuroblastoma (and other neoplasms) with MYCN expression.

- **Histone deacetylase inhibitors** constitute a promising new treatment for cancer due to their novel site of action and low toxicity. It is well known that the acetylation and deacetylation of histones play significant roles in transcriptional regulation of eukaryotic cells. The balance between acetylation and deacetylation is an important factor in regulating gene expression and is thus linked to the control of cell fate. The histone deacetylase inhibitors (HDIs) including the hydroxamic acids, such as suberoylanilide hydroxamic acid and pyroxamide, the benzamides MS-275 and CI-994 and the butyrate derivative 4-PBA are a new class of anti-neoplastic agents currently being evaluated in clinical trials. Moreover, new synthetic HDIs have been used recently in phase I and II clinical trials. Over the next few years experts believe that as first generation HDIs produce clinical benefits and second generation inhibitors are rationally designed with improved specificity, this class of drugs will emerge as a new way of cancer treatment. The first clinical studies have shown that histone hyperacetylation can be achieved safely in humans and that treatment of cancer with such agents seems to become possible. The use of HDIs, probably in association with classical chemotherapy drugs or in combination with DNA-demethylating agents, could be promising for cancer patients. Further evaluation is needed to establish the clinical activity of combination therapy using HDIs with cytotoxic drugs or differentiation induced agents [37]. It was recently reported [38] that histone deacetylase inhibitor treatment induces apoptosis of Neuroblastoma cells by increasing the acetylation of Ku70 in the cytoplasm, resulting in the release of Bax from Ku70. Subsequently, Bax releases cytochrome c from mitochondria causing apoptosis.
- **Angiogenesis inhibitors** are being investigated in a large variety of malignancies such as renal cell carcinoma, head and neck carcinoma, lung cancer, breast cancer, prostate cancer and a variety of haematological malignancies [39]. Over the last ten years several reports have been published concerning the relationship between

tumor progression and angiogenesis in Neuroblastoma in experimental models *in vitro* and *in vivo*. Moreover, a high vascular index in Neuroblastoma correlates with poor prognosis, suggesting dependence of aggressive tumor growth on active angiogenesis [40-41]. The existence of specific angiogenesis inhibitors was first postulated by Folkman in 1971 [42]. The term “antiangiogenesis” was introduced to describe treatment designed to prevent the induction of new blood vessels and perhaps reduce the number of already present ones. There are two classes of angiogenesis inhibitors, direct and indirect. Direct angiogenesis inhibitors target the microvascular endothelial cells that are recruited to the tumor and prevent them from responding to various mitogens (e.g., angiostatin, avastatin, endostatin). Indirect angiogenesis inhibitors generally prevent the expression or block the activity of a tumor protein that activates angiogenesis, such as fibroblast growth factor-2 (FGF-2) and vascular endothelial growth factor (VEGF), or block the expression of its receptors on endothelial cells. Approximately 75 antiangiogenic compounds have been developed and are under clinical testing. Individual tumor angiogenesis is not dependent upon a single angiogenic molecule, but rather NB tumors simultaneously express multiple angiogenic inducers and inhibitors. Eggert *et al.* [43] found that high expression levels of seven angiogenic factors strongly correlated with the advanced stage of NB and this suggests that several angiogenic peptides set in concert in the regulation of neovascularization in NB; this clearly indicates that therapy targeting a single angiogenic factor is unlikely to be effective. Experimental studies show that additive and/or synergistic antitumor activity in NB may be the result of the combination of various antiangiogenic agents or with chemotherapy and/or radiotherapy.

2.6 THE IMPORTANCE OF ANIMAL MODELS IN CANCER

Clinically relevant animal models of human cancer are important for studies of cancer biology, invasion and metastasis, and for investigating new forms of prognostic diagnosis and therapy. Mice have been used in cancer research since 1894. Initially, mice were used for same-species tumor transplantations and drug treatment studies. In 1921, inbred strains that were predisposed to getting tumors were started and disseminated among cancer researchers. In 1962, the discovery of a mutant mouse with low immunity led to human tumor transplantations, a valuable breakthrough for cancer research. A further breakthrough in the late 1980s led to transgenic mice, those whose genes have been altered for producing a desired characteristic. Oncogenes, or genes that cause cancer, could then be studied in greater detail. Mice adapt well to laboratory housing and can be housed socially or individually. They possess a surprising genetic similarity to humans. These features, combined with a rapid rate of reproduction, make mice the mammal of choice for fine-tuned genetic manipulation. Mice with many different special features have been bred or created, including some described here:

- **Inbred Mice:** the inbreeding of mice predisposed to developing cancer has led to a variety of specialized strains. In 1921, Leonell Strong established many inbred strains that frequently and spontaneously developed cancer. Serving as a virtually unlimited source of many types of tumors, these inbred mice have made it possible to study the growth and general characteristics of tumors.
- **Nude Mice:** the nude mouse is a major breakthrough for cancer research because it allows human tumors to be studied in another animal. The nude mouse, a hairless mutant reported in 1962, is immunodeficient, and thus does not reject tumor transplantations from other species. It lacks a thymus, which is essential for the production of T-cells, lymphocytes that are essential to the immune system. By transplanting an actual human tumor into a nude mouse, the tumor can be studied in a whole animal system. The absence of functioning T cells prevents nude mice from rejecting not only allografts, but they cannot even reject xenografts; that is, grafts of tissue from another species. Before discovery of the nude mouse, human tumors were grafted and grown in immune-privileged sites, such as the anterior chamber of the eye, the brain and the cheek pouch. These locations are inconvenient, and the tumors are eventually rejected. The recessive *nu* gene, which is responsible for the lack of a thymus in nude mice, has since been introduced into many types of inbred strains of mice with other immunodeficiencies. Nude mice have served in the

laboratory to gain insights into the immune system, leukemia, solid tumors, AIDS and other forms of immune deficiency.

- **SCID Mice:** in 1983, mice with severe combined immune deficiency (SCID) were reported. SCID mice are even more immunodeficient than nude mice. Tumors from other species are easily transplanted into SCID mice and will grow without being rejected. For certain specific tumors, SCID mice show improved transplantability over nude mice. In addition, SCID mice are ideal for the growth of hybridomas *in vivo* to produce a continuous supply of antibodies (Abs). Sometimes referred to as a reagent, Ab is necessary for a wide range of diagnostic, clinical and experimental procedures.
- **Transgenic Mice:** in the late 1980s the methodology for engineering transgenic mice made it possible to create mice to address specific questions and problems. Transgenic mice result from genetically altered embryos: a gene or combination of genes is microinjected into developing oocytes. The genetic alteration affects the germ plasm, and subsequently can be transmitted to progeny. Through selective breeding, it is then possible to maintain a strain of mice consisting of individuals with particular traits of interest. A specific trait, such as a predisposition to develop a particular type of tumor, can be introduced into a mouse strain by injecting into the embryo an oncogene, a gene that causes cancer. Transgenic mice permit the study of cancer in specific tissues, including initial tumor development.

2.6.1 NEUROBLASTOMA MODELS

The complexity of the interaction between NB tumors and their regional microenvironment emphasizes the importance of *in vivo* models for such investigations. During the past decade, several types of NB animal models have been described. These models include syngenic murine tumors that display neural differentiation consistent with NB, transgenic mice that spontaneously develop NB-like tumors, and human NB cell lines xenografted into immunocompromised mice [44-45]. Murine tumor xenograft models have the disadvantage of tumor growth in immunocompromised host where human tumor cells interact with a murine microenvironment. Their advantage over other models is that the tumor cell is human and that the genetic diversity of NB can be represented by using numerous well-characterized human NB cell lines previously described [46]. Most NB xenograft models characterized to date consist of primary tumor growth at heterotopic (*i.e.* subcutaneous) sites. Orthotopic (*i.e.* intra-adrenal) models for most cancers have the advantage of more relevant tumor biology, including response to therapy and metastatic behaviour, than heterotopic tumor models [47]. In order to provide additional models that can account for the biological diversity of NB, Khanna and coworkers [48] initiated studies for defining five adrenal NB xenograft models. Initial studies comparing heterotopic and orthotopic injections of tumor cells confirmed the findings of others that orthotopic cancer models were superior from the standpoint of both primary tumor biology and spontaneous distant metastasis. Adrenal tumors were well vascularized with each xenograft model demonstrating distinctive patterns of tumor vascularization (angiogenic phenotype). Murine xenograft models should be used in conjunction with syngenic murine models, and genetically-engineered mice in the study of NB. Each model system contributes important information and allows different investigative questions to be addressed. The relevant tumor biology that is reproduced by orthotopic injection of human NB cells within the model system enhances the value of xenograft in the study of NB biology and therapy.

2.7 BIBLIOGRAPHY

- [1] Brodeur, G. M., Maris, J. M., *Principles and Practice of Pediatric Oncology*, J. B. Lippincott Company, Philadelphia, 2002, pp. 895-938.
- [2] Castleberry, R. P., *Eur. J. Cancer* 1997, **33**, 1430-1437.
- [3] Maris, J. M., Matthay, K. K., *J. Clin. Oncol.* 1999, **17**, 2264-2279.
- [4] Evans, A. E., D'Angio, G. J., Propert, K., Anderson, J., Hann, H. W., *Cancer* 1987, **59**, 1583-1589.
- [5] Shimada, H., Ambros, I. M., Dehner, L. P., Hata, J., Joshi, V. V., Roald, B., *Cancer* 1999, **86**, 349-363.
- [6] Matthay, K. K., Perez, C., Seeger, R. C., Brodeur, G. M., Shimada, H., Atkinson, J. B., Black, C. T., Gerbing, R., Haase, G. M., Stram, D. O., Swift, P., Lukens, J. N., *J. Clin. Oncol.* 1998, **16**, 1256-1264.
- [7] Reynolds, C. P., Seeger, R. C., *Am. J. Pediatr. Hematol. Oncol.* 2001, **23**, 150-152.
- [8] Brodeur, G. M., Maris, J. M., Yamashiro, D. J., Hogarty, M. D., White, P. S., *J. Pediatr. Hematol. Oncol.* 1997, **19**, 93-101.
- [9] Fitzgibbon, M. C., Tormey, W. P., *Ann. Clin. Biochem.* 1994, **31**, 1-11.
- [10] Kushner, B. H., Gilbert, F., Helson, L., *Cancer* 1986, **57**, 1887-1893.
- [11] Maris, J. M., Chatten, J., Meadows, A. T., Biegel, J. A., Brodeur, G. M., *Med. Pediatr. Oncol.* 1997, **28**, 1-5.
- [12] Kohl, N. E., Kanda, N., Schreck, R. R., Bruns, G., Latt, S. A., Gilbert, F., Alt, F. W., *Cell* 1983, **35**, 359-367.
- [13] Schwab, M., Alitalo, K., Klempnauer, K. H., Varmus, H. E., Bishop, J. M., Gilbert, F., Brodeur, G., Goldstein, M., Trent, J. M., *Nature* 1983, **305**, 245-248.
- [14] Shimada, H., Stram, D. O., Chatten, J., Joshi, V. V., Hachitanda, Y., Brodeur, G. M., Lukens, J. N., Matthay, K. K., Seeger, R. C., *J. Natl. Cancer Inst.* 1995, **87**, 1470-1476.
- [15] Brodeur, G. M., Fong, C., *Cancer Genet. Cytogenet.* 1989, **41**, 153-174.
- [16] George, R. E., Kenyon, R. M., McGuckin, A. G., *Oncogene* 1996, **12**, 1583-1587.
- [17] Schwab, M., Praml, C., Amler, L. C., *Genes Chromosomes Cancer* 1996, **16**, 211-229.
- [18] Brodeur, G. M., Green, A. A., Hayes, F. A., Williams, K. J., Williams, D. L., Tsiatis, A. A., *Cancer Res.* 1981, **41**, 4678-4686.
- [19] Cheng, N. C., Roy, N. V., Chan, A., Beitsma, M., Westerveld, A., Speleman, F., Versteeg, R., *Oncogene* 1995, **10**, 291-297.
- [20] Adida, C., Berrebi, D., Peuchmaur, M., Reyes-Mugica, M., Altieri, D. C., *Lancet* 1998, **351**, 882-883.

- [21] Mertens, F., Johansson, B., Hoglund, M., Mitelman, F., *Cancer Res.* 1997, **57**, 2765-2780.
- [22] Bader, S. A., Fasching, C., Brodeur, G. M., Stanbridge, E. J., *Cell Growth Differ.* 1991, **2**, 245-255.
- [23] Van Roy, N., Laureys, G., Cheng, N. C., Willem, P., Opdenakker, G., Versteeg, R., Speleman, F., *Genes Chromosomes Cancer* 1994, **10**, 103-114.
- [24] Nakagawara, A., Arima, M., Azar, C. G., Scavarla, N. J., Brodeur, G. M., *Cancer Res.* 1992, **52**, 1364-1368.
- [25] Nakagawara, A., Azar, C. G., Scavarda, N. J., Brodeur, G. M., *Mol. Cell. Biol.* 1994, **14**, 759-767.
- [26] Ryden, M., Sehgal, R., Dominici, C., Schilling, F. H., Ibanez, C. F., Kogner P., *Br. J. Cancer* 1996, **74**, 773-779.
- [27] Look, A. T., Hayes, F. A., Nitschke, R., McWilliams, N. B., Green, A. A., *N. Engl. J. Med.* 1984, **311**, 231-235.
- [28] Cohn, S. L., Rademaker, A. W., Salwen, H. R., Franklin, W. A., Gonzales-Crussi, F., Rosen, S. T., Bauer, K. D., *Am. J. Pathol.* 1990, **136**, 1043-1052.
- [29] Look, A. T., Hayes, F. A., Shuster, J. J., Douglass, E. C., Castleberry, R. P., Bowman, L. C., Smith, E. I., Brodeur, G. M., *J. Clin. Oncol.* 1991, **9**, 581-591.
- [30] Prigione, I., Corrias, M. V., Airoldi, I., Raffaghello, L., Moranti, F., Bocca, P., Cocco, C., Ferrone, S., Pistoia, V., *Ann. N. Y. Acad. Sci.* 2004, **1028**, 69-80.
- [31] Nielsen, P. E., Egholm, M., Berg, R. H., Buchardt, O., *Science* 1991, **254**, 1497-1500.
- [32] Mologni, L., Marchesi, E., Nielsen, P. E., Gambacorti-Passerini, C., *Cancer Res.* 2001, **15**, 5468-5473.
- [33] Cutrona, G., Carpaneto, E. M., Ulivi, M., Rondella, S., Landt, O., Ferrarini, M., Boffa, L. C., *Nat. Biotechnol.* 2000, **18**, 300-303.
- [34] Sun, L., Fuselier, J. A., Murphy, W. A., Coy, D. H., *Peptides* 2002, **23**, 1557-1565.
- [35] Pession, A., Tonelli, R., Fronza, R., Sciamanna, E., Corradini, R., Sforza, S., Tedeschi, T., Marchelli, R., Montanaro, L., Camerin, C., Franzoni, M., Paolucci, G., *Int. J. Oncol.* 2004, **24**, 265-272.
- [36] Tonelli, R., Purgato, S., Camerin, C., Fronza, R., Bologna, F., Alberesi, S., Franzoni, M., Corradini, R., Sforza, S., Faccini, A., Shohet, J. M., Marchelli, R., Pession A., *Mol. Cancer Ther.* 2005, **4**, 779-786.
- [37] Kouraklis, G., Theocharis, S., *Oncol. Rep.* 2006, **15**, 489-494.
- [38] Subramanian, C., Ovipari, A. W. Jr., Castle, V. P., Kwok, R. P., *Cell Cycle* 2005, **4**, 1741-1743.
- [39] Rhee, J., Hoff, P. M., *Expert Opin. Pharmacother.* 2005, **6**, 1701-1711.

-
- [40] Canete, A., Navano, S., Bermudez, J., Pellin, A., Castel, V., Llombardt-Bosh, A., *J. Clin. Oncol.* 2000, **18**, 27-34.
- [41] Ribatti, D., Vacca, A., Nico, B., De Falco, G., Montaldo, P. G., Ponzoni, M., *Eur. J. Cancer* 2002, **38**, 750-757.
- [42] Folkman, J., *New Engl. J. Med.* 1971, **285**, 1182-1186.
- [43] Eggert, A., Ikegaki, N., Kwiatkowski, J., Ihoo, H., Brodeur, G. M., Himelstein, B. P., *Clin. Cancer Res.* 2000, **6**, 1900-1908.
- [44] Ziegler, M. M., Ishizu, H., Nagabuchi, E., Takada, N., Arya, G., *Cancer* 1997, **79**, 1757-1766.
- [45] Beltinger, C., Debatin, K. M., *Int. J. Cancer* 2001, **92**, 313-318.
- [46] Thiele, C. J., *Neuroblastoma in Human Cell culture*, Masters J. Lancaster, UK, 1998, pp. 21-53.
- [47] Killion, J. J., Radinsky, R., Fidler, I. J., *Cancer Metastasis Rev.* 1998, **17**, 279-284.
- [48] Khanna, C., Jaboin, J. J., Drakos, E., Tsokos, M., Thiele, C. J., *In Vivo* 2002, **16**, 77-85.

CHAPTER 3

PROTEOMIC ANALYSIS OF AN ORTHOTOPIC NEUROBLASTOMA ANIMAL MODEL

3.1 INTRODUCTION

Despite various therapeutical treatments, including radio-, immune- and chemio-therapy, at elevated doses, with or without bone marrow transplant, childhood cancers caused by neural crest cells, are still characterised by a high percentage of relapse and a high rate of mortality. The effective treatment of NB, either at advanced stages or at minimal residual disease, remains one of the major challenges in paediatric oncology. Prognosis for patients with this disease has improved with advances in medical care but the overall 5 year survival is still less than 60% [1]. Indeed, the incidence of fatal relapses is still high and long-term survival remains very low [2]. Starting from these considerations, there is a great need for designing novel therapeutic strategies. Innovative therapies are thus required and should be focused on the genes and biological pathways that might contribute to malignant transformation or progression of this neoplasia. To be highly effective, cancer therapy has to be individualised. This requires a detailed understanding of the properties of each tumor with regard to metastasising propensity and drug sensitivity, information that is difficult or even impossible to gain from microscopic analysis of hematoxylin-stained tissue sections. Detailed information regarding mutations in growth-controlling genes, expression of genes controlling growth, metastasis and drug resistance must be obtained. Proteome analysis, today, is perhaps one of the most valuable tools for defining markers useful for tumor diagnosis [3]. In addition, proteomics has the potential of unravelling basic tumor biological questions regarding mechanisms involved in the pathogenesis of cancer. In this study, we present a first proteomic approach to NB analysis, based on an orthotopic Neuroblastoma animal model.

3.1.1 BIOLOGICAL RELEVANCE OF THE ORTHOTOPIC NEUROBLASTOMA XENOGRAFT MODEL

Previous *in vivo* experiments, by using directly an *i.v.* administration of NB cells in nude mice, have proved the importance of choosing a specific xenograft animal model that mimics the metastatic spread observed in advanced-stage human NB patients [4-5]. However, a further and more realistic view of a comparative study could be obtained if a tumor model was available that better reflected the growth of advanced NB in children (*i.e.* large adrenal gland tumors and multiple small metastatic lesions). All current data support this concept and recommend that orthotopic implantation of tumor cells in recipient animals is mandatory for studies of tumor progression, angiogenesis, invasion, and metastasis [6]. In order to provide a well characterised, relevant, highly reproducible, angiogenic, and metastatic orthotopic model of NB, Pastorino and coworkers [4] initiated studies to define an adrenal NB xenograft model. They decided to use the intra-adrenal injection of NXS2 murine NB cells in mice because, 2 weeks after injection, adrenal gland tumors were always found in all animals. This model best reflected the typical growth pattern of human NB, since orthotopic injection of NXS2 cells resulted in solid adrenal tumors that were highly vascular, locally invasive into surrounding tissues, and metastatic to distant sites. Indeed, macroscopic metastases always occurred after 3-4 weeks of injection in the ovary and spleen, while micrometastases were apparent frequently in the contralateral adrenal gland, kidneys, liver, bone marrow and lung.

3.2 EXPERIMENTAL PROCEDURES

3.2.1 ORTHOTOPIC NEUROBLASTOMA ANIMAL MODEL

Five-week-old female nude (nu/nu) mice were purchased from Harlan Laboratories (Harlan Italy-S. Pietro al Natisone, UD). Mice were anaesthetised and injected with 5×10^4 cultured murine NXS2 Neuroblastoma cells [7] (kindly provided by Dr. Reisfield, Scripps Clinic, La Jolla, CA, USA) in 20 μ L of HEPES buffer, after laparotomy, in the capsule of the left adrenal gland. The lethality of the method was 0%. Mice were monitored at least two times weekly for evidence of tumor development, quantification of tumor size, and evidence of tumor-associated morbidity. Mice were sacrificed and organs, after washing in PBS, were frozen and stored in liquid nitrogen. All experiments involving animals have been reviewed and approved by the licensing and ethical committee of the National Cancer Research Institute and by the Italian Ministry of Health. Four samples, divided in two classes, constituted the dataset: four samples from healthy adrenal nude mouse glands (whose pool was used to perform four 2-D maps); four samples from adrenal nude mouse glands carrying primary Neuroblastoma tumors (the pool of which was used for performing four 2-D maps).

3.2.2 2D-PAGE

Healthy and tumor adrenal glands from nude mice were stored at -80 °C. All samples were homogenised (5% homogenate) with a lysis solution containing 7 M urea, 2 M thiourea, 3% CHAPS, 40 mM Tris, 5 mM TBP, 0.5% carrier ampholytes pH 3-10, 5 mM EDTA, 1 mM PMSF and 50 U/mL DNase. After 1 h lysis, samples were centrifuged at 4 °C at 6,000 rpm for eliminating all residual particles. The collected supernatant was alkylated for 1 h with 10 mM acrylamide (1 M stock solution) and the reaction was blocked with 10 mM DTT (from 1 M stock solution) [8]. Interfering substances (lipids, salts) were removed by a precipitation step. A cold mix of acetone and methanol (8:1) was added and, after 2 h at -20 °C, the solution was centrifuged at 13,000 rpm for 30 min. Pellets were finally resuspended in sample solution containing urea, thiourea, CHAPS and Tris. The first dimension run was performed on strips (17 cm length, 0.5 mm thickness) with a linear pH gradient from pH 3 to 10 [9]. The IPG strips were rehydrated with 400 μ g of pooled samples (four strips with pooled healthy samples and four strips with pooled tumor samples) and containing traces of bromophenol blue for monitoring the electrophoretic run. Isoelectric focusing (IEF) was carried out with a Protean IEF Cell (Bio-Rad), with a low initial voltage and then by applying a voltage gradient up to 10,000 V with a limiting current of 50 μ A/strip. The total product time \times voltage applied was 75,000 Vh for each strip and the temperature was set at 20 °C. Each strip was equilibrated with a SDS denaturing solution containing 6 M urea, 2% SDS, 20% glycerol, 0.375 M Tris-HCl (pH 8.8). This step was conducted for 27 min in tubes containing 20 mL each of the equilibration solution. The IPG strips were then laid on a 7-20%T gradient SDS-PAGE with 0.5% agarose solubilised in the cathodic buffer (192 mM glycine, 0.1% SDS, Tris to pH 8.3). The anodic buffer was a solution of 0.375 M Tris-HCl, pH 8.8. The electrophoretic run was performed by setting a current of 2 mA for each gel for 1 h, then 5 mA/gel for 2

h and 10 mA/gel until the end of the run. During the whole run the temperature was set at 11 °C. All gels were fixed with a solution containing 40% ethanol and 10% acetic acid for 30 min. After this step the gels were stained with Sypro Ruby (70 mL for each gel) overnight and then destained with 10% methanol, 7% acetic acid for 1 h. The gels were finally washed with milliQ water. All the samples were then digitised with the Versa Doc Scanner.

3.2.3 PROTEIN PATTERN AND STATISTICAL ANALYSIS

The digitised images were acquired with the software PDQuest (version 6.2), which was used for cropping and orienting the images, for detecting and identifying spots, for comparing and matching spots, for normalising and analysing the data and for preparing a report. A match set was created from the protein patterns of the two independent cellular extracts (healthy mouse adrenal gland, mouse adrenal gland carrying NB primary tumors). A standard gel was generated out of the image with the highest spot number. Spot quantities of all gels were normalised by removing non expression-related variations in spot intensity; for that, the raw quantity of each spot in a gel was divided by the total quantity of all the spots in that gel and included in the standard. The final synthetic image was a Gaussian scan image that contained all the Gaussian spots with a defined volume and quality. All subsequent spot matching and analysis steps in the PDQuest software were performed on Gaussian spots. The results were evaluated in terms of spot optical density (OD). Statistical analysis (Student's *t*-test) via PDQuest allowed the study of proteins that were significantly increased or decreased in pathological samples (with a significance level α of 0.05).

3.2.4 PROTEIN IDENTIFICATION BY MASS SPECTROMETRY

The spots of interest were carefully excised from the gel with a razor blade and placed in Eppendorf tubes. The gel pieces were washed twice with a solution of acetonitrile/Tris 5 mM pH 8.5 (50/50) followed by a single wash with only Tris 5 mM pH 8.5. These pieces were dehydrated in a Speedvac device at room temperature and covered with 15 μ L of Trypsin (0.02 mg/mL) in NH_4HCO_3 buffer (40 mM, pH 8.5) and left at 37 °C overnight. The peptides were extracted two times in 50 μ L of acetonitrile/ H_2O 1% (v/v) formic acid (50/50). The extraction was conducted in an ultrasonic bath for 15 min each time. After sonication, the excess of acetonitrile was evaporated and the peptides were resuspended in 10 μ L of 0.1% TFA, concentrated and cleaned using ZipTip microcolumns (C18). The extracted peptides were loaded onto the MALDI target plate by mixing 1 μ L of each sample with the same volume of a matrix solution, prepared fresh every day by dissolving 10 mg/mL cyano-4-hydroxycinnamic acid in acetonitrile/ethanol (1:1, v/v), and allowed to dry. Measurements were performed using a ToFSpec 2E MALDI-TOF instrument (Micromass, Manchester, UK), operated in the reflectron mode, with an accelerating voltage of 20 kV [10]. The laser wavelength was 337 nm and the laser repetition rate was 4 Hz. The final mass spectra were produced by averaging 50-200 laser shots. Peptide masses were searched against SWISS-PROT, TrEMBLE and NCBI nr databases by utilizing the ProteinLynx program from Micromass, Profound from Prowl and Mascot from Matrix Science.

3.2.5 IMMUNOBLOT ANALYSIS

Proteins derived from healthy and tumor adrenal glands, previously separated by 2D-PAGE, were transferred onto PVDF membranes (Immuno-Blot™ PVDF membrane, Bio-Rad) for 2 h at 60 V, using a Mini Trans-Blot system (Bio-Rad). Membranes were blocked with 3% BSA in TBST (10 mM Tris, 150 mM NaCl, 0.1% Tween 20, pH 7.5) overnight at 4 °C and then incubated for 1 h at room temperature with the anti-OP18/stathmin antibody (1:5,000, Sigma), recognising the C-terminus of this protein. Blots were developed with an enhanced chemiluminescence system (ECL plus, Amersham) and stathmin visualised on a radiography film (Kodak, Rochester). Films were scanned using a densitometer (GS-710, Bio-Rad).

3.3 RESULTS

Differential 2-D maps were generated for healthy and tumoral samples. Figure 1 shows an example of the 2D-PAGE maps thus obtained.

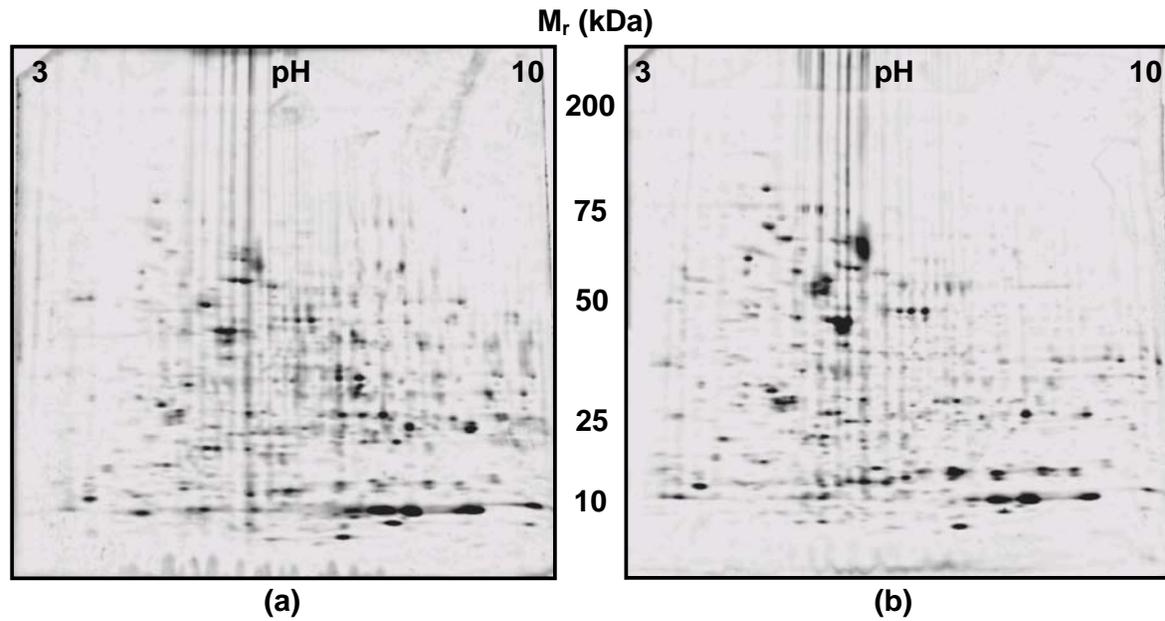


Figure 1. 2-D maps representing healthy (a) and tumoral (b) adrenal glands.

Figure 2 shows the master map of the Neuroblastoma developed in nude mice: about 750 spots could be counted with Sypro Ruby staining. They seem to be quite evenly distributed in the pH 3-10 interval, with a M_r distribution from ca. 9,000 to 140,000 Da. The circles mark 84 spots found to be differentially expressed in pathological samples, of which 39 were up-regulated and 45 down-regulated. Moreover, a number of protein spots that were newly expressed and newly silenced in the tumor samples as compared to healthy ones could also be detected. Additionally, the numbers refer to the 14 spots which could be identified by MS analysis (an additional two, numbers 15 and 16, although producing good quality spectra, could not be matched to any entries available in the databases).

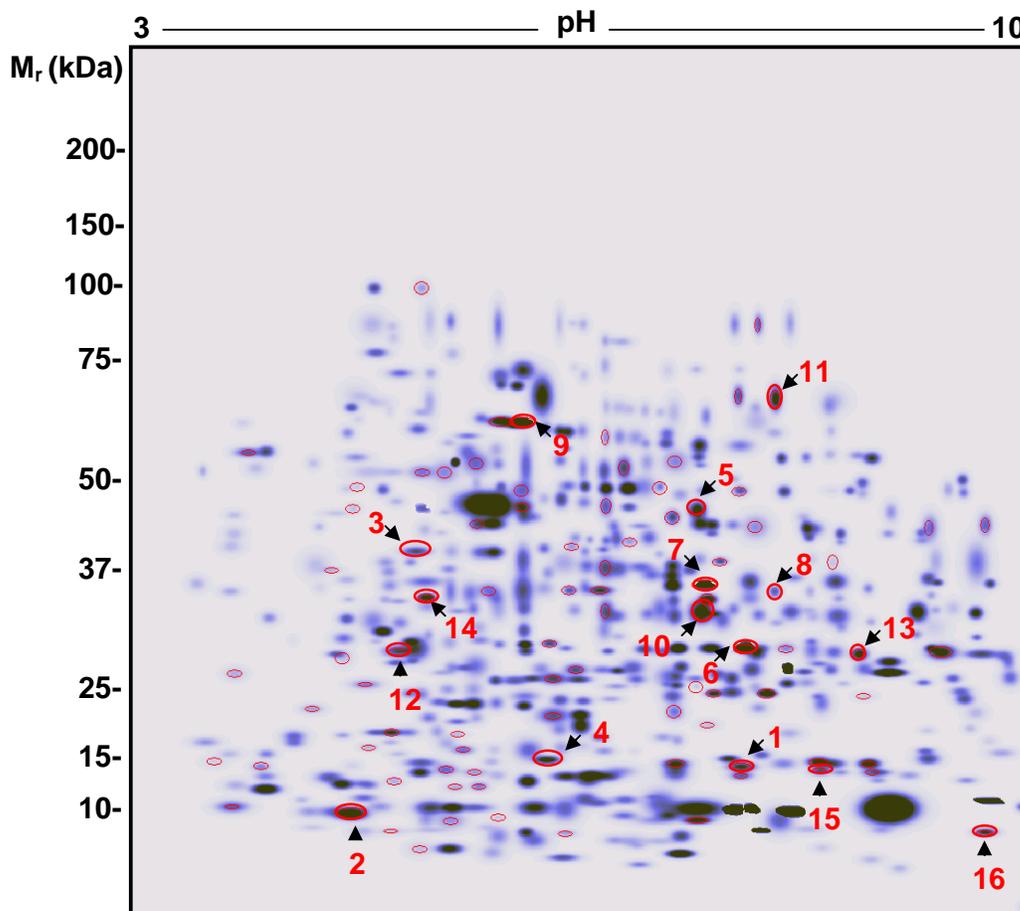


Figure 2. Master map of a Neuroblastoma tissue excised from the adrenal glands of nude mice. The 84 spots differentially expressed are circled and the identified spots are numbered. The circled numbered spots refer to the 14 spots which could be identified by mass spectrometry upon excision (two of them, spots 15 and 16, gave good spectra but could not be found in any of the databases available).

Figure 3 shows a small gel area with a few spots up- and down-regulated and newly expressed in pathological versus control specimens. The spot with a question mark corresponds to a two-fold up-regulated protein which could not be identified in any of the databases available, although it gave good MS spectra.

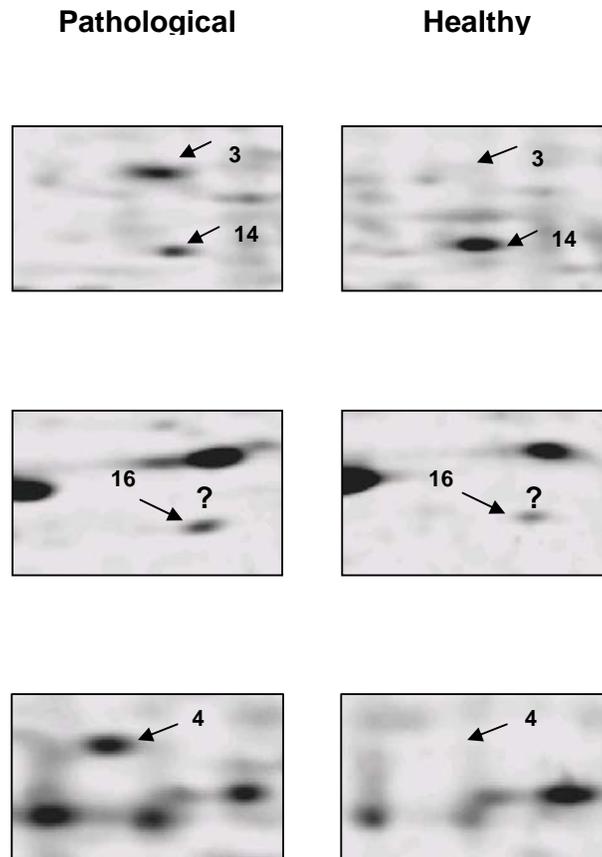


Figure 3. Comparison of two-dimensional gel patterns of some proteins in Neuroblastoma (left panel) and healthy (right panel) tissues. The corresponding spot number are shown in Table 1.

The identified spots, together with experimental and theoretically predicted pI and M_r values, are given in Table 1. The classification of these proteins is summarized in Figure 4.

Table 1. Identified proteins from the Neuroblastoma tissue.

SSP no.	Exp Mr (Da)	Exp pI	Theor Mr	Theor pI	Z-Score	MOWSE-Score	Protein name	SwissProt Accession number	% cov.	No. Peptide	Fold of Variation
1	~15000	7,08	17840	8,03	2,39	1,93 E8	Peptidyl-prolyl <i>cis-trans</i> isomeraseA	P17742	59,5	13	↑ 5
2	~13000	5	20123	5,05	1,84	1,6 E4	Adrenodoxin	P46656	25	8	↓ 21
3	~40000	4,05	32560	4,06	2,38	2,51 E5	Nucleophosmin (NPM)	Q61937	31	8	> 10
4	~17000	6,05	17143	6	2,33	3,78 E7	Stathmin (OP18)	P54227	64,9	17	> 10
5	~50000	7,05	46661	6,09	2,31	4,46 E8	Isocitrate dehydrogenase	O88844	33,3	15	< 10
6	~30000	7,08	29366	7,04	1,67	1,78 E5	Carbonil anhydrase III	P14141	28	8	↓ 8
7	~37000	7,05	35857	7,02	2,29	1,77 E11	Aldose reductase related protein	P21300	59,3	18	↓ 8
8	~35000	8	33335	8,02	2,1	2,68 E7	Thiosulfate sulfurtransferase	P52196	39	13	< 10
9	~65000	6,05	60956	6,01	2,37	2,61 E21	60 kDa heat shock protein	P19226	57	37	↓ 4
10	~34000	7,05	34973	8,06	2,38		Electron transfer flavoprotein	NCBI gi21704230	41	12	↓ 4
11	~72000	8,01	67631	7,07	2,41	2,09 E9	Transketolase	P40142	28	18	↓ 2
12	~30000	5,01	27771	4,08	2,06	2,57 E6	14-3-3 protein zeta/delta (protein kinase C inhibitor protein-1)	P35215	42	9	↑ 2
13	~27000	8,09	31517	8,06	2,28	6,78 E6	Enoyl-CoA hydrase	P14604	34	10	↓ 4
14	~36000	5,03	35753	4,08	2,38	2,7 E13	Annexin V	P48036	64	22	↓ 2
15	~15000	8,05	?	?	?	?	?	?	?	?	↑ 3
16	~8500	9,08	?	?	?	?	?	?	?	?	↑ 2

Classification of the proteins was obtained by the program FatiGo (Figure 4).

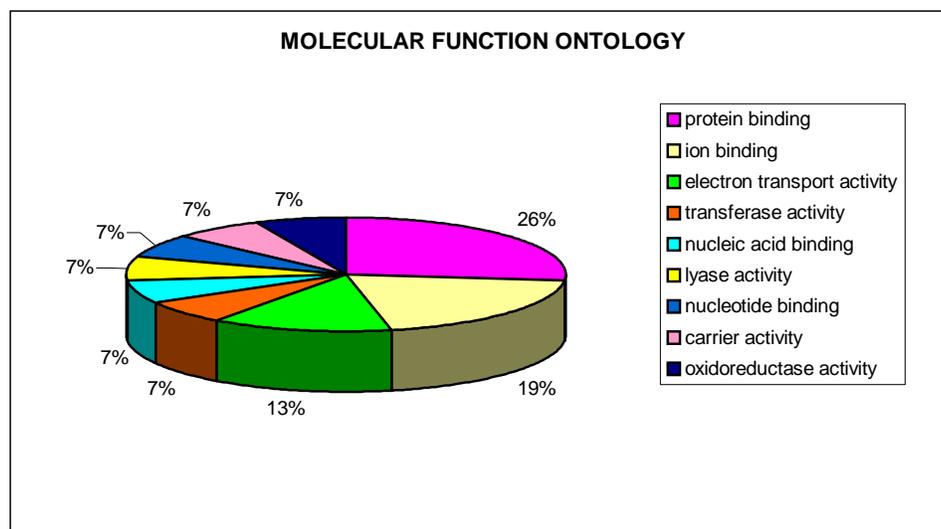


Figure 4. Classification of the differentially expressed proteins in Neuroblastoma tissues according to the biological function in which they are involved, obtained by the program FatiGo (<http://fatigo.bioinfo.cnio.es>)

In general, there was a good agreement between the theoretically-predicted and experimentally found pI and M_r values, except for a large deviation in the case of adrenodoxin, suggesting that this protein might be extensively modified at the post-translational level in this kind of tissue, or be a pre-protein, as discussed below. Adrenodoxin (ADX) is a small iron-sulfur protein present in the mitochondrial matrix, where it transfers electrons from adrenodoxin reductase to mitochondrial forms of cytochrome *P*-450. Adrenodoxin is encoded as larger precursors of 188 amino acids, and the signal peptide (64 amino acids) is cleaved during transfer into mitochondria. The M_r of the mature adrenodoxin chain (124 amino acids) is 13,617 Da: this is in agreement with our experimental data. The mouse adrenodoxin pre-protein is 93, 74, 70 and 69% identical to the rat, human, bovine and porcine adrenodoxin pre-proteins, respectively. The mature adrenodoxin is 98% identical to rat adrenodoxin and 90% identical to the human, bovine and porcine adrenodoxin [11]. In order to further confirm the data obtained with the master map, we performed an immunoblot analysis on some identified proteins. For example on stathmin, a protein which, together with nucleophosmin (NPM), appears to be expressed only on tumor samples (Figure 3). As shown in Figure 5, with this method low levels of stathmin can also be detected in normal murine tissue (right panel), whereas in the tumor sample (left panel) the same spot appears to be over-expressed by a factor of at least >15 fold. Additionally, when the same experiment

was repeated in a narrower pH gradient (pH 4-7) this very intense spot (which appeared oblong in the wide pH range) could be resolved into a well-separated string of minor spots, more acidic than the parental protein.

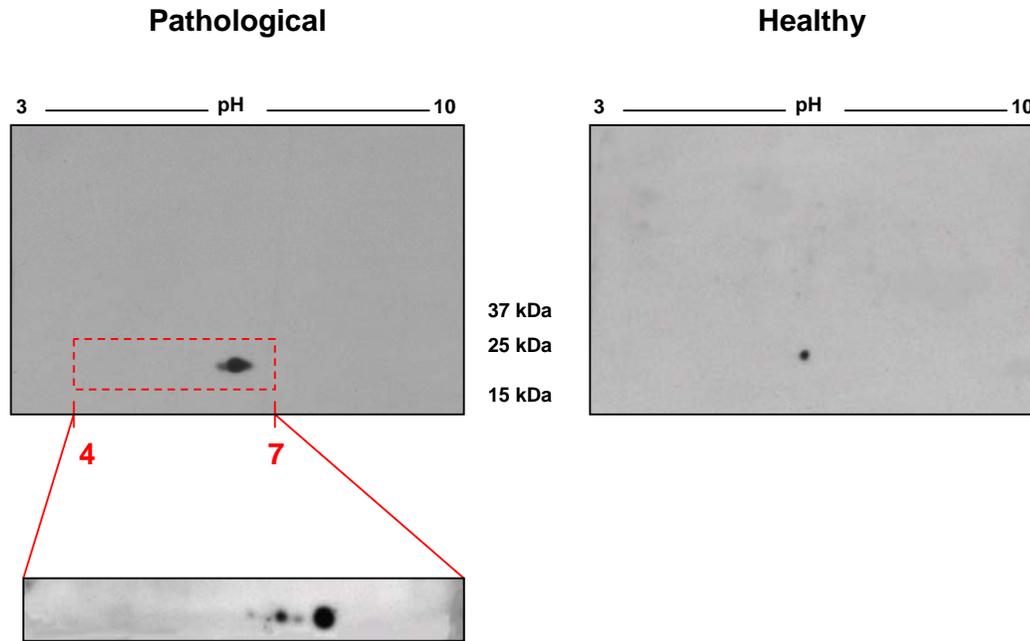


Figure 5. Immunoblots of 2D-PAGE slabs from normal murine samples (right panel) and tumor tissues (left panel), developed by using an anti-OP18/stathmin antibody, recognising the C-terminus of this protein. Lower strip on the left: same boxed area of the upper gel, but run in a shallower (pH 4-7) IPG gradient. Note the string of spots to the left of the parental spot, suggesting a phosphorylated train.

This experiment suggests two important points, which might have escaped detection up to the present. First of all, it might not be true that a gene is fully silenced or newly turned-on in pathological tissues (or during ontogenesis or any other biologically-relevant phenomenon, such as differentiation, cell-cycle arrest, apoptosis, cell modification induced by drug treatments) as compared to the control, healthy one. Indeed, extremely low levels might be present in the control samples, and *vice versa*. What our experiments suggest is that, in the control samples, the levels of stathmin were so minute as to be below the sensitivity limit of Sypro Ruby staining. With immuno-detection (the most sensitive detection method today available) detectable levels of stathmin could indeed be found also in control tissue (Sypro Ruby has a sensitivity at least three orders of magnitude lower than our chemiluminescence detection method on the immunoblot) [12]. The other very important point, is that only with such a sensitive method we could observe that, in tumor tissues, stathmin seems to be

modified, possibly at the post-synthetic level, the p/ variations in the string of minor spots suggesting that there could be phosphorylation events.

3.4 DISCUSSION

It is believed that cancer develops through a multistep process involving the accumulation of genetic alterations that lead to altered expression patterns, modifications in protein structures and functions. To many, the key to a cure for patients with cancer is early diagnosis. In the post-genomics era, proteomics promises the discovery of biomarkers and tumor markers for early detection and diagnosis, novel protein-based drug targets for anticancer therapy, and new endpoints for the assessment of therapeutic efficacy and toxicity [13].

Proteomic approaches complement global gene expression approaches, which are powerful tools for identifying differentially expressed genes but are hampered by imperfect correlation of the levels of mRNA and proteins. To date, one of the most consistently successful proteomic methodology is the combination of two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) for protein separation and visualization, followed by mass spectrometric (MS) protein identification using peptide mass fingerprints and tandem MS peptide sequencing. Differential protein expression profiles detected by 2D-PAGE and MS have been reported for various types of human cancers, such as pancreatic ductal adenocarcinoma, renal-cell carcinoma, colorectal cancer, breast cancer, hepatocellular carcinomas [14-15].

Among the 14 proteins which were identified by MS analysis, of particular interest are the two up-regulated proteins Cyclophilin A (CypA) and 14-3-3 protein zeta-delta, as well as the two highly expressed proteins Nucleophosmin (NPM) and Stathmin (OP18). Their role will be briefly discussed below.

1. 14-3-3 protein zeta-delta (up-modulated 2 fold in primary tumor)

14-3-3 proteins form a family of highly conserved proteins capable of binding to more than 200 different mostly phosphorylated proteins. They are present in all eukaryotic organisms investigated. 14-3-3 binding partners are involved in almost every cellular process and 14-3-3 proteins play a key role in these processes. Among the proteins that interact with the 14-3-3 proteins are kinases (Raf-1, ASK-1), phosphatases (Cdc25), modulators of protein G (RGS7) and regulators of apoptotic induction (Bcl-2, Bad). The various isoforms of 14-3-3 carry out important functions in apoptotic control, by interacting with proteins such as Bad and Raf-1. The interaction of 14-3-3 with Bad seems to block activation of the mitochondrial proapoptotic pathway, by preventing interaction of Bad with Bcl-2 and Bcl-x_L. Masters and collaborators [16] have recently demonstrated that the over-expression of any of the seven isoforms of 14-3-3 results in a strong quenching of the apoptotic pathway.

2. Cyclophilin A (up-modulated 5 fold in primary tumor)

CypA, a 5-fold up-regulated protein in Neuroblastoma primary tumor tissues, is found in normal cells and comprises ~0.1% of cytoplasmic protein [17]. It is a member of the peptidyl-prolyl isomerases, a group of proteins that catalyzes cis-trans isomerization about peptidyl-prolyl bonds during protein folding or conformational changes. CypA was first identified as the primary intracellular target of the immunosuppressant cyclosporin A (CsA) [18]. Through inhibition of the phosphatase calcineurin, CypA:CsA down-regulates the nuclear translocation of nuclear factors of activated T cells and downstream gene transcription (*i.e.*, interleukin 2 in T cells) [19].

Recent evidence, however, points to possible roles for CypA overexpression in tumorigenesis. In two separate reports, CypA was shown to interact with the retinoblastoma susceptibility gene product p105Rb (Rb) in Jurkat T-cell extracts. Nuclear translocation of Rb:CypA correlated with retinoic acid-induced neural differentiation [20-21]. Although CypA seems to have a role in apoptosis, it is not clear if it is involved in activation or inhibition, or both. CypA was shown to be involved in caspase activation in neurons [22], and to cooperate with apoptosis-inducing factor during apoptosis associated chromatinolysis [23]. However, CypA was also reported to bind the calcium-sequestering protein calreticulin by yeast two hybrid screen [24], which suggests that it may inhibit calcium fluxes required for apoptosis [25]. Recently, CypA was identified as one of the most dominantly expressed proteins in non-small-cell lung cancer [26], and a novel Cyp, which is similar to CypA, also has been associated with metastasis and is overexpressed in bladder cancer, hepatocellular carcinoma, sarcoma, and breast carcinoma [27] and represents a potentially novel therapeutic target.

3. Nucleophosmin (up-modulated more than 10 fold in primary tumor)

Nucleophosmin (NPM) also called B23, is a ubiquitously expressed phosphoprotein involved in ribosome assembly/transport, cytoplasmic/nuclear trafficking, regulation of DNA polymerase alpha activity, centrosome duplication, and regulation of p53. NPM continuously shuttles between the nucleus and cytoplasm. It has been shown to bind nucleic acid, prevent protein aggregation *via* its chaperon activities, protect enzymes during thermal denaturation, and facilitate renaturation of chemically denatured proteins. In its cellular structure role, there is evidence suggesting that NPM is associated with the centrosome. It is the substrate of CDK2/cyclin E during duplication of centrosomes (cellular division). Due to the NPM gene interaction with several tumor-associated chromosome translocations, NPM is thought to be a portion of several fusion proteins: NPM-ALK, NPM-RAR, and NPM-MLF1. While it is not

thought to be part of the transforming potential of these fusion proteins, it is believed to act as the interface for oligomerization and oncogenic conversion of these tumor promoting fusion proteins.

NPM is overexpressed in actively proliferating cells and cancer cells; its overexpression inhibits apoptosis, while knockdown of NPM induced cell death [28].

4. Stathmin (up-modulated more than 10 fold in primary tumor)

Stathmin/OP18 is a 19-kDa cytoplasmic protein, whose function is related to microtubule turnover. Stathmin/OP18 promotes tubulin catastrophes by binding to a microtubule plus end, which stimulates the exposed β -subunit of the terminal tubulin dimer to hydrolyze its GTP, causing tubulin to undergo a conformational change that favors transition to depolymerization. Thus, stathmin/OP18 increases microtubule turnover and contributes to rapid reorganization of the microtubule cytoskeleton. Microtubules are essential for a wide variety of cellular functions including mitosis, cell shape, and cell motility. OP18 is also highly expressed in a number of human malignancies, such as acute leukemias, lymphomas, neuroblastomas, prostatic adenocarcinomas, and breast carcinomas [29-31].

3.5 CONCLUSIONS

The present data show that a proteomic approach, combined with image analysis and mass spectrometry, could be a valuable tool for gathering important information about proteins associated with tumor development. To our knowledge, this is the first attempt, at a proteome level, of analysis of the paediatric solid tumor Neuroblastoma. At present, we were forced to adopt a clinically relevant animal model, by injecting murine NB tumor cells in the capsule of the left adrenal gland of nude mice, due to the fact that human studies are hampered by the difficulty to procure proper, healthy controls. Although preliminary in nature, this study will be quite valuable in view of extending it to human cases.

3.6 BIBLIOGRAPHY

- [1] Niethammer, D., Handgretinger, R., *Eur. J. Cancer* 1995, **31A**, 568-571.
- [2] Ladenstein, R., Philip, T., Lasset, C., Hartmann, O., Garaventa, A., Pinkerton, R., Michon, J., Pritchard, J., Klingebiel, T., Kremens, B., Pearson, A., Coze, C., Paolucci, P., Frappaz, D., Gadner, H., Chauvin, F., *J. Clin. Oncol.* 1998, **16**, 953-965.
- [3] Righetti, P. G., Stoyanov, A., Zhukov, M., *The Proteome Revisited: Theory and Practice of All Relevant Electrophoretic Steps*, Elsevier, Amsterdam, 2001.
- [4] Pastorino, F., Brignole, C., Marimpietri, D., Sapra, P., Moase, E. H., Allen, T. M., Ponzoni, M., *Cancer Res.* 2003, **63**, 86-92.
- [5] Raffaghello, L., Pagnan, G., Pastorino, F., Cosimo, E., Brignole, C., Marimpietri, D., Montaldo, P. G., Gambini, C., Allen, T. M., Bogenmann, E., Ponzoni, M., *Int. J. Cancer* 2003, **104**, 559-567.
- [6] Chambers, A. F., Groom, A. C., MacDonald, I. C., *Nat. Rev. Cancer* 2002, **2**, 563-572.
- [7] Greene, L. A., Shain, W., Chalazonitis, A., Breakfield, X., Minna, J., Coon, H. G., Nirenberg, M., *Proc. Natl. Acad. Sci. U.S.A.* 1975, **72**, 4923-4927.
- [8] Herbert, B., Galvani, M., Hamdan, M., Olivieri, E., McCarthy, J., Pedersen, S., Righetti, P. G., *Electrophoresis* 2001, **22**, 2046-2057.
- [9] Righetti, P. G., *Immobilized pH Gradients: Theory and Methodology*, Elsevier, Amsterdam, 1990.
- [10] Hamdan, H., Righetti, P. G., *Mass Spectrom. Rev.* 2002, **21**, 287-302.
- [11] Stromstedt, M., Waterman, M. R., *Biochim. Biophys. Acta* 1995, **1261**, 126-128.
- [12] Patton, W. F., *BioTechniques* 2000, **28**, 944-948.
- [13] Wu, W., Hu, W., Kavanagh, J. J., *Int. J. Gynecol. Cancer* 2002, **12**, 409-423.
- [14] Shekouh, A. R., Thompson, C. C., Prime, W., Campbell, F., Hamlett, J., Herrington, C. S., Lemoine, N. R., Crnogorac-Jurcevic, T., Buechler, M. W., Friess, H., Neoptolemos, J. P., Pennington, S. R., Costello, E., *Proteomics* 2003, **3**, 1988-2001.
- [15] Roberts, K., Bhatia, K., Stanton, P., Lord, R., *Proteomics* 2004, **4**, 784-792.
- [16] Masters, S. C., Subramanian, R. R., Truong, A., Yang, H., Fujii, K., Zhang, H., Fu, H., *Biochem. Soc. Trans.* 2002, **30**, 360-365.
- [17] Ryffel, B., Woerly, G., Greiner, B., Haendler, B., Mihatsch, M., Foxwell, J., *Immunology* 1991, **72**, 399-404.
- [18] Handschumacher, R. E., Harding, M. W., Rice, J., Drugge, R. J., Speicher, D. W., *Science* 1984, **226**, 544-547.
- [19] Ivery, M., *Med. Res. Rev.* 2000, **20**, 452-484.
- [20] Cui, Y., Mirkia, K., Fu, F., Zhu, L., Yokoyama, K. K., Chiu, R., *J. Cell. Biochem.* 2002, **86**, 630-641.

-
- [21] Song, J., Lu, Y., Yokoyama, K., Rossi, J., Chiu, R., *J. Biol. Chem.* 2004, **279**, 24414-24419.
- [22] Capano, M., Virji, S., Crompton, M., *Biochem. J.* 2002, **363**, 29-36.
- [23] Cande, C., Vahsen, N., Kouranti, I., Schmitt, E., Daugas, E., Spahr, C., Luban, J., Kroemer, R. T., Giordanetto, F., Garrido, C., Penninger, J. M., Kroemer, G., *Oncogene* 2004, **23**, 1514 -1521.
- [24] Reddy, P. A., Atreya, C. D., *Int. J. Biol. Macromol.* 1999, **25**, 345-351.
- [25] Demaux, N., Distelhorst, C., *Science* 2003, **300**, 65-67.
- [26] Campa, M. J., Wang, M. Z., Howard, B., Fitzgerald, M. C., Patz, E. F. Jr., *Cancer Res.* 2003, **63**, 1652-1656.
- [27] Meza-Zepeda, L. A., Forus, A., Lygren, B., Dahlberg, A. B., Godager, L. H., South, A. P., Marenholz, I., Lioumi, M., Florenes, V. A., Maelandsmo, G. M., Serra, M., Mischke, D., Nizetic, D., Ragoussis, J., Tarkkanen, M., Nesland, J. M., Knuutila, S., Myklebost, O., *Oncogene* 2002, **21**, 2261-2269.
- [28] Ye, K., *Cancer Biol. Ther.* 2005, **4**, 918-923.
- [29] Roos, G., Brattsand, G., Landberg, G., Marklund, U., Gullberg, M., *Leukemia* 1993, **7**, 1538-1546.
- [30] Melhem, R. F., Strahler, J. R., Hailat, N., Zhu, X. X., Hanash, S. M., *Biochem. Biophys. Res. Commun.* 1991, **179**, 1649-1655.
- [31] Antonucci, F., Chilosi, M., Parolini, C., Hamdan, M., Astner, H., Righetti, P. G., *Electrophoresis* 2003, **24**, 2376-2385.

CHAPTER 4

STUDY OF HUMAN NEUROBLASTOMA PROGRESSION USING PROTEOME ANALYSIS

4.1 INTRODUCTION

Tumor metastasis is the dominant cause of death in cancer patients. However, the molecular and cellular mechanisms underlying tumor metastasis are still elusive. The identification of protein molecules with their expressions correlating to the metastatic process would help to understand the metastatic mechanisms and thus facilitate the development of strategies for the therapeutic interventions and clinical management of cancer. Cancer cells acquire properties required to form metastatic colonies, a process known as tumor progression (Figure 1). This process can be rapid or may require years or decades to provide detectable metastasis. Properties attained by tumor cells during metastatic progression may include (a) disaggregation of cells in the primary tumor; (b) invasion: increased motility of primary tumor cells and enhanced degradation of extracellular matrix; (c) induction of new blood vessels (angiogenesis); (d) intravasation: entry of tumor cells into the circulation; (e) extravasation: exit of tumor cells into tissue in distant organs; (f) formation of micrometastases; and (g) conversion of micrometastases to macrometastases. The mechanisms underlying progression include both genetic (mutational) and epigenetic changes that result in the alteration of protein profiles in cancer cells over time [1].

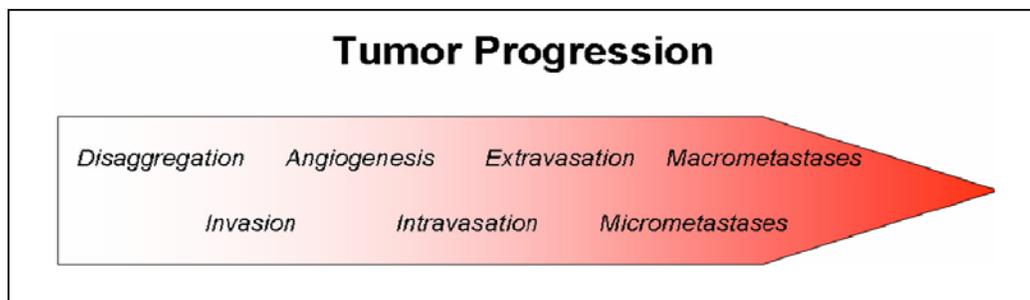


Figure 1. Elements of tumor progression.

Proteomics is a systematic research approach aiming to provide the global characterization of protein expression and function under given conditions. Proteomic technology has been widely used in biomarker discovery and pathogenetic studies including tumor metastasis [2]. There are two main expected outcomes from proteomic analysis of tumor metastasis. The first is to discover new molecular markers from the profiling of metastatic tumors. The second is to decipher the intracellular signaling pathways that lead cancer cells to be metastatic. Expression and functional proteomics are suited for the purposes of the metastasis studies. The resulting data would provide knowledge bases for the early detection and prediction of metastasis and for the identification of novel targets for drug development and therapeutic intervention.

Metastasis models in proteomics

Before research data can be correlated with clinical practice, a fundamental problem is the establishment of an ideal metastatic model [3]. So far, most of experiments employed cell lines, biopsy or tissue samples as analytic materials. Each of them has strength and weakness. Cell lines have the advantages of reproducibility, availability in large numbers, and homogeneity in cell lineage. Lots of experiments attempted to identify biomarkers or obtain biological insights of tumor metastasis by comparing the protein expressions between cell lines that differ in metastatic potential. However, no cell lines in culture are fully representative of tumor *in vivo*. Researches have proved that without a supporting tumor microenvironment, only cancer cells alone are not enough to confer metastatic status [4] and that tumor-microenvironmental interaction has a decisive role in controlling local cancer growth, invasion and distant metastasis [5]. Biopsies and tumor blocks are ideal experiment materials in cancer research. The data derived from such samples exhibit the real condition of cancer development. But a major difficulty is the cellular heterogeneity of tumors. Besides cancer cells, tumor tissues also contain other cell types, including myoepithelial cells, fibroblasts and endothelial cells. It is possible that a proteomics analysis may be confounded by opposed protein changes in different cell types and then the sensitivity of the analysis may be greatly reduced. Laser capture microdissection (LCM) technology may provide help, enabling researchers to collect specific cell types from a tissue sample. Other commonly used metastatic models were formed by inoculating tumor cells or tissue blocks to nude mice (or in general immunocompromised mice) to develop visible metastases at secondary sites. Those metastasis-associated molecules then were identified based on these models. The complexity of the interaction between NB tumors and their regional microenvironment emphasizes the importance of an *in vivo* model for such investigation. We started our studies with a human NB cell line (SH-SY5Y) orthotopically xenografted into immunocompromised

mice (SCID mice). This model best reflects the typical growth pattern of human NB, since injection of SH-SY5Y cells in the adrenal glands resulted in solid tumors that were locally invasive into surrounding tissues, and metastatic to distant sites. In order to evaluate NB progression and the mechanism underlying the metastatic process and to elucidate possible therapeutical targets, as well as putative molecular targets regulating metastasis, we performed proteomic analysis on primary adrenal tumors vs liver metastases (liver is one of the most important organ affected by NB metastasis).

4.2 EXPERIMENTAL PROCEDURES

4.2.1 XENOGRAFT NEUROBLASTOMA ANIMAL MODEL

Neuroblastoma cells were injected in the adrenal glands of mice, as previously reported [6]. Briefly, five-week-old female SCID mice purchased from Harlan Laboratories (Harlan Italy-S.Pietro al Natisone, UD), were anaesthetised and injected with 2×10^6 SH-SY5Y Neuroblastoma cells in 20 μ L of HEPES buffer, after laparotomy, in the capsule of the left adrenal gland. Mice were monitored at least two times weekly for evidence of tumor development, quantification of tumor size, and evidence of tumor-associated morbidity. Mice were sacrificed and organs, after washing in PBS, were frozen and stored in liquid nitrogen. In all cases with tumor cells (either primary cancer or metastasis), histology confirmed that more than 95% of the tissue samples used for proteomics contained cancer cells.

All experiments involving animals have been reviewed and approved by the licensing and ethical committee of the National Cancer Research Institute and by Italian Ministry of Health.

4.2.2 2D-PAGE

Adrenal glands carrying primary NB tumors and NB liver metastases, from SCID mice, were stored at -80 °C. All samples were homogenised (5% homogenate) with a lysis solution containing 7 M urea, 2 M thiourea, 3% CHAPS, 40 mM Tris, 5 mM TBP, 0.5% carrier ampholytes pH 3-10, 5 mM EDTA, 1 mM PMSF. After 1 hour lysis, samples were centrifuged at 4 °C at 6,000 rpm for eliminating all residual particles. The collected supernatant was reduced and alkylated as described by Herbert and coworkers [7]. Interfering substances (lipids, salts) were removed by a precipitation step with a cold mix of acetone and methanol (8:1). Pellets were finally resuspended in sample solution containing urea, thiourea, CHAPS and Tris. The first dimension run was performed on strips (17 cm length, 0.5 mm thickness) with a linear pH gradient from pH 3 to pH 10 [8]. The IPG strips were rehydrated with 500 μ g of pooled samples (5 strips with pooled adrenal glands carrying primary NB tumor and 5 strips with pooled liver metastasis samples) and containing traces of bromophenol blue for monitoring the electrophoretic run. Isoelectric focusing (IEF) was carried out with a Protean IEF Cell (Bio-Rad), with a low initial voltage and then by applying a voltage gradient up to 10,000 V with a limiting current of 50

$\mu\text{A}/\text{strip}$. The total product time \times voltage applied was 75,000 Vh for each strip and the temperature was set at 20 °C. Each strip was equilibrated with a SDS denaturing solution containing 6 M urea, 2% SDS, 20% glycerol, 0.375 M Tris-HCl (pH 8.8). This step was conducted for 27 min in tubes containing 20 mL each of the equilibration solution. The IPG strips were then laid on a 7-20%T gradient SDS-PAGE with 0.5% agarose solubilised in the cathodic buffer (192 mM glycine, 0.1% SDS, Tris to pH 8.3). The anodic buffer was a solution of 0.375 M Tris-HCl, pH 8.8. The electrophoretic run was performed by setting a current of 2 mA for each gel for 1 h, then 5 mA/gel for 2 h and 10 mA/gel until the end of the run. During the whole run the temperature was set at 11 °C. All gels were fixed with a solution containing 40% ethanol and 10% acetic acid for 30 min. After this step the gels were stained with Sypro Ruby (70 mL for each gel) overnight and then destained with 10% methanol, 7% acetic acid for 1 h. The gels were finally washed with milliQ water.

The images of gels stained with Sypro Ruby were scanned with a Versa Doc (Bio-Rad), and analyzed with the software PDQuest Version 6.2 (Bio-Rad, Labs, Hercules, CA) as described in section 3.2.3 of chapter 3.

4.2.3 PROTEIN IDENTIFICATION BY MASS SPECTROMETRY

Classical “in gel digestion” was performed basically according to the standard protocol [9] but in a 96-well plate for parallel processing. In brief, the excised gel pieces were transferred manually into an individual well of a V-bottom polypropylene 96-well plate (Pink Microtiter plates-pierced, Genomic Solutions) and washed sequentially once with 100 μL of water and 100 μL of 25 mM NH_4HCO_3 ; dehydrated by adding 100 μL of acetonitrile. The gel pieces were then swelled in 100 μL of 10 mM DTT for 30 min at RT and subsequently shrunken with acetonitrile. The proteins were alkylated with 100 μL of 55 mM iodoacetamide for 20 min at RT in the dark. All the liquid material was transferred by centrifugation at 1000 rpm for 2 min. After drying completely, the gel pieces were reswollen with 20-25 μL of digestion buffer (12.5 ng/ μL trypsin in 25 mM NH_4HCO_3) for 30 min at 4 °C. Afterwards, the digestion buffer was replaced by the same buffer without trypsin to ensure that the gel plugs were kept wet during digestion process. The plate was covered and incubated at 37 °C for 12-18 h.

Following digestion, the peptide mixture was desalted and concentrated using home-made C_{18} -Stage tips [10]. The peptides were eluted directly onto a collection plate using 10 μL of 80% acetonitrile/1% TFA by vacuum (4 to 6 Hg). Following elution, the volume of the extracted peptides was reduced to 2 μL using a centrifugal vacuum concentrator (Eppendorf, concentrator 5301). The total tryptic digest samples were analyzed by reverse-phase nanoLC-MS/MS using an Agilent 1100 nanoflow LC system (Agilent Technologies Inc.). The LC-system was coupled to a QSTAR XL hybrid QqTOF mass spectrometer (Applied Biosystem, MDS Sciex). Binding and chromatographic separation of peptides was achieved in a 15 cm long fused silica emitter (100 μm ID from Picotip New Objective) packed in-house with reverse phase C18 resin (Maisch, Germany). The tryptic peptide mixtures were loaded using an auto-sampler and at a flow rate of 700 nL/min onto the packed column. The peptides were separated in a linear gradient of 5 to 24% acetonitrile in 0.5% acetic acid over 15 min and then in further 5 min to 81% acetonitrile at 300 nL/min. The QSTAR was operated in data-dependent acquisition mode to automatically switch between MS and MS/MS. Protein identification was done

using Mascot (Matrix Science, London, UK) searching the mammalian NCBI non-redundant protein database. Search parameters were: MS and MS/MS tolerance 0.2 Da, tryptic specificity allowing for up to 1 missed cleavages, fixed modification: carbamidomethylation of cysteine, variable modification: oxidation of methionine.

4.2.4 IMMUNOBLOT ANALYSIS

Proteins derived from primary NB tumor and NB liver metastases, previously separated by 12% SDS-PAGE were transferred onto PVDF membranes (Immuno-Blot™ PVDF membrane, Bio-Rad) for 2 h at 60 V, using a Mini Trans-Blot system (Bio-Rad). Membranes were blocked with 10% non fat dry milk in TBST (10 mM Tris, 150 mM NaCl, 0.1% Tween 20, pH 7.5) overnight at 4 °C followed by incubation with primary antibody opportunistically diluted in 0.01% Tween 20% and 3% non fat dry milk in TBS, for 1 h at room temperature. After that, blots were rinsed in TTBS solution, by applying 4 x 15 minutes washes in gentle agitation. Secondary anti-mouse, anti-rabbit, anti-goat peroxidase-conjugated antibodies were diluted in the same solution used for the primary ones, and were applied to blots for 1 h at room temperature, followed by another series of 4 washes in TTBS. Proteins were visualized by enhanced chemiluminescence (ECL kit) on Kodak photographic slab provided by Sigma-Aldrich Chemie GmbH (Steinheim, Germany). The following antibodies were used: anti-RKIP (in goat antibody, dilution ratio 1:100), anti-OP18 (in rabbit antibody, dilution ratio 1:5000), anti-SEPT2 (in goat antibody, dilution ratio 1:500), anti-PSMD7 (in rabbit antibody, dilution ratio 1:500), anti-eIF4E (in rabbit antibody, dilution ratio 1:500), anti-β actin (in mouse antibody, dilution ratio 1:5000 dilution). Films were scanned using a densitometer (GS-710, Bio-Rad).

4.2.5 BIOCHEMICAL FUNCTION AND PATHWAY ANALYSIS

Proteins functions were obtained from the SWISS-PROT Protein knowledge base (when available). Additionally, the software package PathwayAssist (Stratagene, La Jolla, CA) was used to identify functional relationships among the proteins. This software explores gene interaction networks represented in the ResNet database. ResNet is a comprehensive database of molecular networks compiled by proprietary natural language processing techniques applied to the whole PubMed database. The database contains more than 100,000 events of regulation, interaction and modification between 15,000 proteins, cell processes and small molecules.

4.3 RESULTS and DISCUSSION

Differential 2-D maps were generated for primary tumor and liver metastasis samples. Figure 2 shows an example of the 2D-PAGE maps thus obtained. Five replicates maps were obtained for each sample.

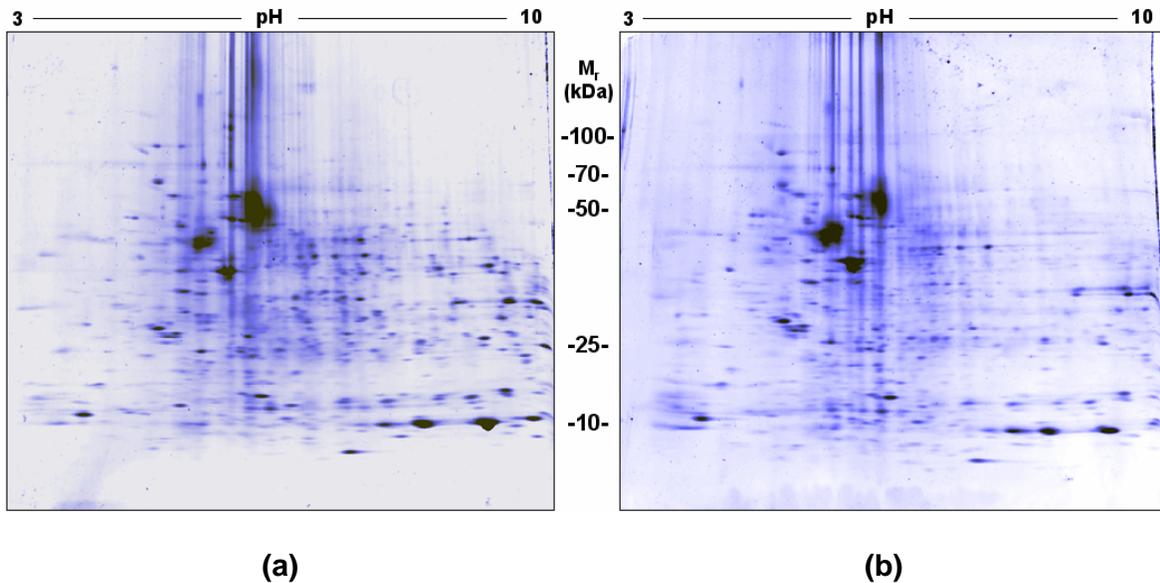


Figure 2. 2-D maps representing primary tumor (a) and liver metastasis (b) samples.

Figure 3 shows the master map of the Neuroblastoma cells: about 270 spots could be counted with Sypro Ruby staining and 85 spots were found to be differently expressed of which 30 were up-regulated, 15 down-regulated, 28 were highly expressed (more than 10 times) and 12 highly silenced (less than 10 times). All these spots were eluted and treated as described in section 4.2.3. The 44 identified spots, by means of LC-MS/MS analysis, are circled in the standard map and are indicated with their SSP numbers.

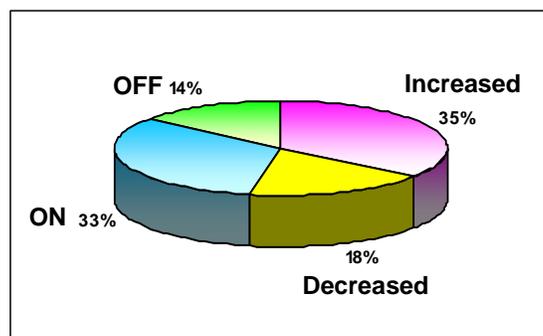
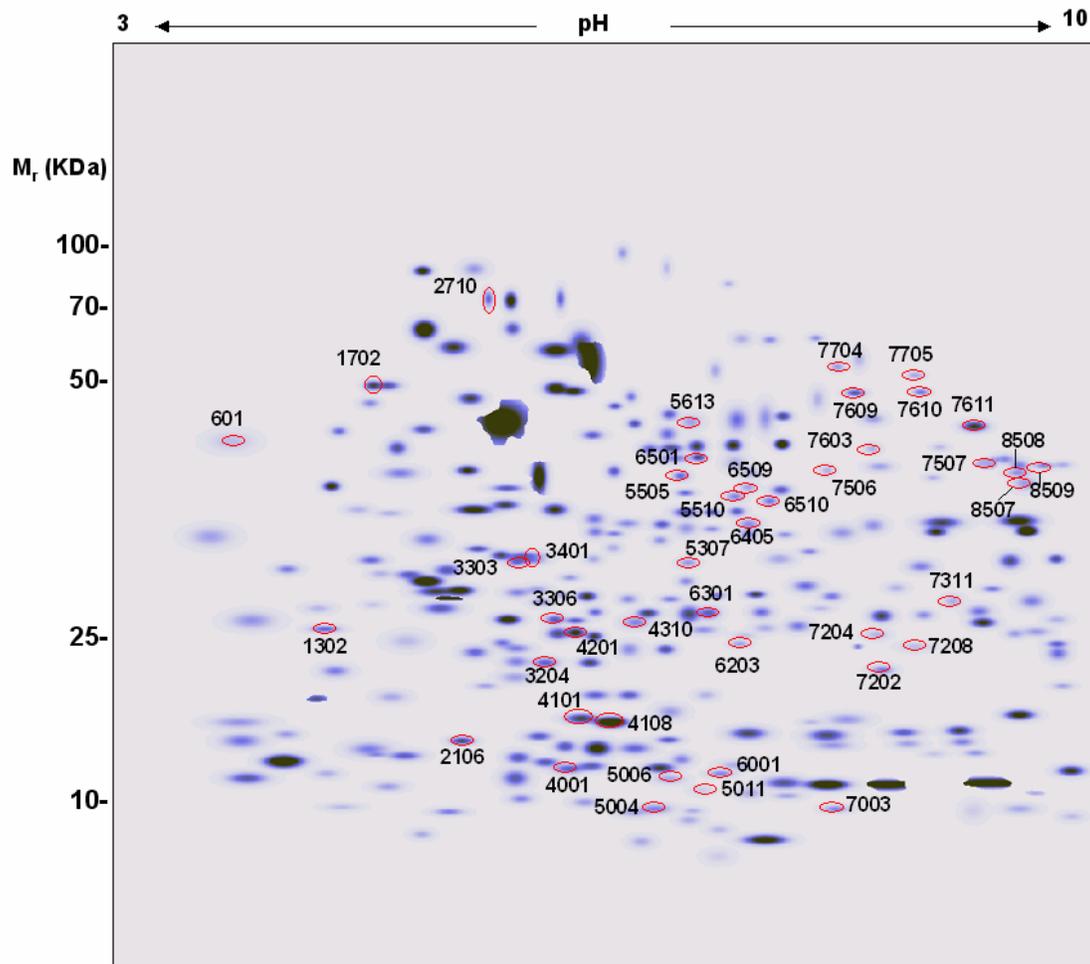


Figure 3. Master map of Neuroblastoma cells. The 44 identified spots are circled and numbered (see Table 1 for their classification). Pie chart: summary of all proteins up- or down-regulated and/or switched on/off.

Table 1. Identified proteins

SSP ID	Fold of variation	Protein name	NCBI accession no.	HUGO gene name	Reported function
601	↑ 2,4	Y box-binding protein 1	gi 112410	YB1	signaling DNA damage and cell proliferation
3303	↑ 3,5	cytoskeleton associated protein	gi 736704	CKAP1	protein folding, microtubule-based process
4310	↑ 2,7	eukaryotic translation initiation factor 4E	gi 3219774	eIF-4E	translational initiation, protein biosynthesis, regulation of translation
5006	↑ 3,9	Histidine triad nucleotide-binding protein 1	gi 1633186	HINT1	signal transduction
5011	↑ 3,9	Histidine triad nucleotide-binding protein 1	gi 1633186	HINT1	signal transduction
5307	↑ 2,5	DNAJC14 protein	gi 16877384	DNAJC14	protein folding
		3-hydroxyanthranilate 3,4-dioxygenase	gi 20810280	HAAO	synthesis of the excitotoxin quinolinic acid (QUIN) from 3-hydroxyanthranilic acid
5613	↑ 2,5	Chain A, Human Mitochondrial Aldehyde Dehydrogenase Complexed With Nad+ And Mn2+	gi 6137677	ALDH2	carbohydrate metabolism, metabolism, alcohol metabolism
6203	↑ 2,4	thioredoxin peroxidase	gi 5453549	PRDX2	oxygen and reactive oxygen species metabolism, response to oxidative stress, anti-apoptosis
6405	↑ 3,9	arginase	gi 178995	ARG	arginine catabolism, response to wounding, urea cycle, arginine metabolism
7204	↑ 3,4	calreticulin precursor	gi 4757900	CALR	regulation of transcription, DNA-dependent, protein folding
7208	↑ 2,2	peroxiredoxin 1	gi 4505591	PRX1	skeletal development, cell proliferation, response to oxidative stress
7311	↑ 2,8	mitochondrial short-chain enoyl-coenzyme A hydratase 1 precursor	gi 12707570	ECHS1	fatty acid metabolism, energy pathways, metabolism, fatty acid beta-oxidation
7503	↑ 3,1	acetyl-CoA C-acetyltransferase 1 precursor, mitochondrial	gi 86728	ACAT1	mitochondrially localized enzyme that catalyzes the reversible formation of acetoacetyl-CoA from two molecules of acetyl-CoA.
7506	↑ 3,6	proteasome endopeptidase complex p42 chain	gi 7435741	PSMC6	ATP-dependent degradation of ubiquitinated proteins
7507	↑ 5,2	Phosphoglycerate kinase 1	gi 6679291	PGK1	glycolysis
7603	↑ 2,6	3-hydroxy-3-methylglutaryl-Coenzyme A synthase 2 (mitochondrial)	gi 5031751	HMGCS2	ketone body biosynthesis, acetyl-CoA metabolism, cholesterol biosynthesis
8508	↑ 2,7	acetyl-CoA C-acetyltransferase 1 precursor, mitochondrial	gi 86728	ACAT1	mitochondrially localized enzyme that catalyzes the reversible formation of acetoacetyl-CoA from two molecules of acetyl-CoA.
2106	> 10	DNAJC14 protein	gi 16877384	DNAJC14	protein folding
		3-hydroxyanthranilate 3,4-dioxygenase	gi 20810280	HAAO	synthesis of quinolinic acid
5510	> 10	26S proteasome regulatory chain, p40	gi 1085272	PSMD7	protein biosynthesis, ribosome biogenesis
6509	> 10	fumarylacetoacetate hydrolase, FAH	gi 253320	FAH	tyrosine catabolism, regulation of transcription, DNA-dependent, metabolism, aromatic amino acid family metabolism, L-phenylalanine catabolism
6510	> 10	poly(rC) binding protein 1	gi 5453854		single-stranded nucleic acid binding protein
7704	> 10	Delta-1-pyrroline-5-carboxylate dehydrogenase, mitochondrial precursor	gi 2506350	ALDH4A1	aldehyde metabolism, retinoic acid metabolism, metabolism

7705	> 10	Delta-1-pyrroline-5-carboxylate dehydrogenase, mitochondrial precursor	gi 2506350	ALDH4A1	aldehyde metabolism, retinoic acid metabolism, metabolism
7609	> 10	catalase	gi 179950	CAT	response to oxidative stress, electron transport
7610	> 10	catalase	gi 179950	CAT	response to oxidative stress, electron transport
7611	> 10	ATP synthase, H+ transporting, mitochondrial F1 complex	gi 4757810	ATP5A1	ATP synthesis coupled proton transport, proton transport, ATP biosynthesis
8509	> 10	Phosphoglycerate kinase 1	gi 129902	PGK1	glycolysis
1302	↓ 2,0	GTP binding protein	gi 4092054	RAN	intracellular protein transport, protein-nucleus import, DNA metabolism, RNA-nucleus export, cell growth and/or maintenance, spermatid development, small GTPase mediated signal transduction, signal transduction, regulation of cell cycle
1702	↓ 2,2	calreticulin precursor	gi 4757900	CALR	regulation of transcription, DNA-dependent, protein folding
2710	↓ 3,0	heat shock 90kDa protein 1, beta	gi 20149594	HSPCB	response to heat, response to unfolded protein, positive regulation of nitric oxide biosynthesis, protein folding
3306	↓ 2,2	ubiquitin thiolesterase	gi 88254		protein metabolism
4101	↓ 2,0	stathmin 1	gi 5031851	OP18	cell growth and/or maintenance, mitotic spindle assembly, intracellular signaling cascade, microtubule depolymerization, microtubule-based process, axonogenesis
4108	↓ 2,0	stathmin 1	gi 5031851	OP18	see spot no. 4101
4201	↓ 2,1	glutathione transferase	gi 87564	GST	Conjugation of reduced glutathione to a wide number of exogenous and endogenous hydrophobic electrophiles.
		ubiquitin thiolesterase	gi 88254	OTUB	hydrolase that can remove conjugated ubiquitin from proteins in vitro and may therefore play an important regulatory role at the level of protein turnover by preventing degradation.
5004	↓ 2,7	S100 calcium binding protein A11 (calgizzarin)	gi 5032057	S100A11	negative regulation of cell proliferation, negative regulation of DNA replication
		BolA-like protein 2 (My016 protein)	gi 12001972	BOLA2	transcription regulator activity
6501	↓ 2,2	actin-related protein	gi 381964	ACTR1A	vesicle-mediated transport
		pancreatic tumor-related protein	gi 189597		anchoring to other cellular components
7202	↓ 2,6	raf kinase inhibitor protein	gi 913159	RKIP	interacts with Raf-1 and seems to inhibit it
3204	< 10	40s ribosomal protein S7	gi 337518	RPS7	protein biosynthesis, ribosome biogenesis
		ATP synthase, H+ transporting, mitochondrial F0 complex, subunit d isoform a	gi 5453559	ATP5J	energy pathways, ATP synthesis coupled proton transport, proton transport
		thiol-specific antioxidant protein	gi 438069	PRX2	oxygen and reactive oxygen species metabolism, response to oxidative stress, anti-apoptosis
3401	< 10	Fructose-1,6-bisphosphatase	gi 119740	FBP1	gluconeogenesis, carbohydrate metabolism, fructose metabolism
4001	< 10	Retinoic acid-binding protein I, cellular	gi 266904	CRABP1	development, signal transduction, transport
5505	< 10	Human Diff6,H5,CDC10 homologue	gi 1040689	NEDD5	cell cycle, cytokinesis
6001	< 10	Histidine triad nucleotide-binding protein 1	gi 1633186	HINT1	see spots no. 5006, 5011
6301	< 10	Chain A, Triosephosphate Isomerase	gi 999892	TPI1	metabolism, fatty acid biosynthesis, glycolysis, pentose-phosphate shunt, lipid biosynthesis, gluconeogenesis
		Chain A, Horf6 A Novel Human Peroxidase Enzyme	gi 3318841		
8507	< 10	Phosphoglycerate kinase 1	gi 129902	PGK1	glycolysis

Figure 4 shows a small gel area with a few spots up- and down-regulated in liver metastasis (LM) vs primary tumor (PT).

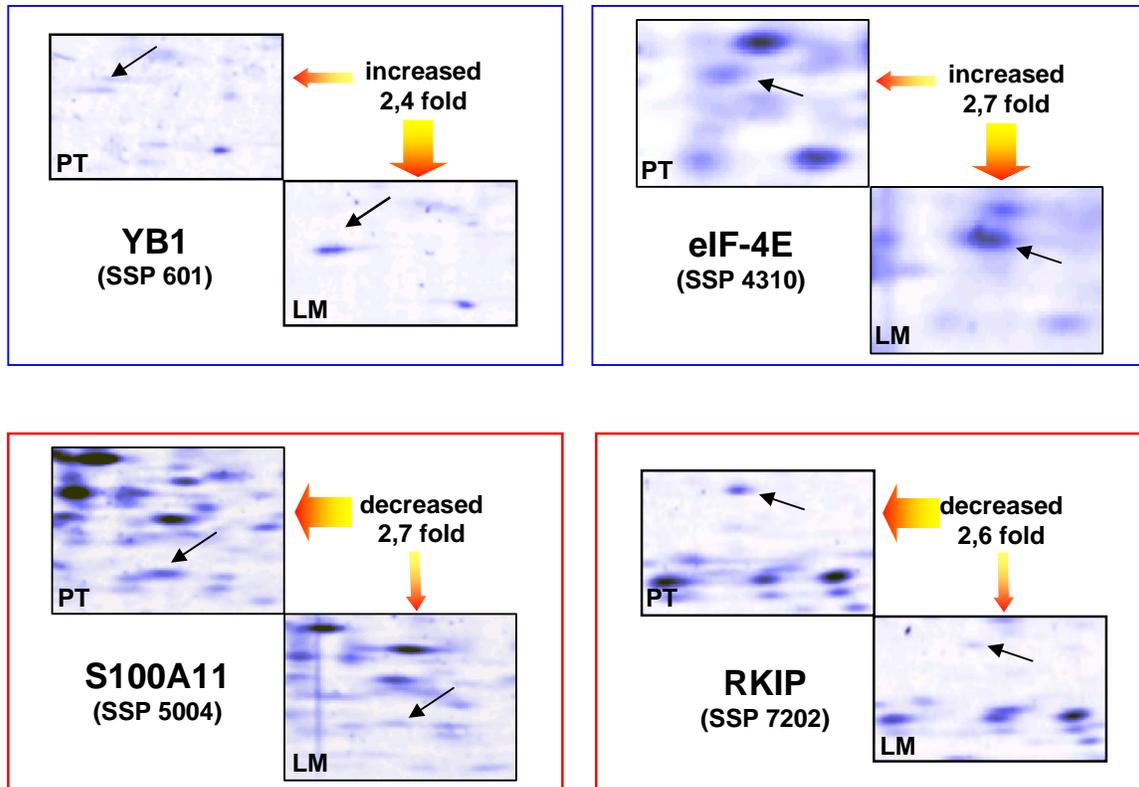


Figure 4. Examples of identified proteins: YB1, eIF-4E, S100A11, RKIP. For each protein is shown the position in the gel and the expression trend in the comparison.

To search for functional associations and common targets of the identified proteins, we used the PathwayAssist software and the data obtained are shown in Figure 5. Among the identified proteins, the most notable groups were proteins thought to be involved in metastasis and processes related to metastasis such as transformation, invasion, motility, angiogenesis, apoptosis, differentiation and proliferation.

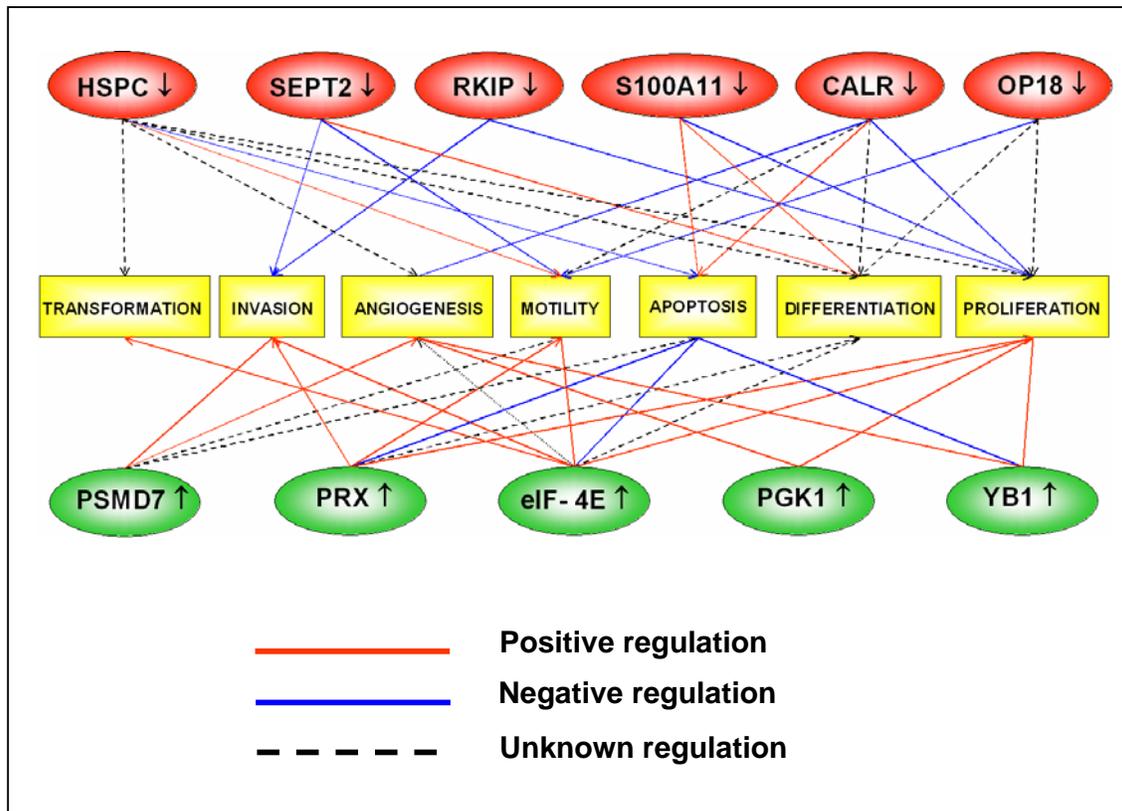


Figure 5. Common targets (cell processes) for identified proteins in liver metastasis vs primary tumor. To bring all the relevant information about proteins together, we have used the PathwayAssist software (Stratagene) for generating a network of molecules that have common targets (transformation, invasion, angiogenesis, motility, apoptosis, differentiation, proliferation). The activating, inhibiting or unclear connections are indicated with the colors (red), (blue), (black), respectively. The meaning of the different shape is the following: ellipse → identified proteins, rectangle → cell processes.

4.3.1 DIFFERENTIAL PROTEOMIC ANALYSIS

Among the identified proteins, the most notable groups were proteins thought to be involved in metastasis and processes related to metastasis such as transformation, invasion, motility, angiogenesis, apoptosis, differentiation and proliferation.

The role of the most important proteins detected is briefly discussed below.

4.3.1.1 DIFFERENTIALLY DOWN-MODULATED PROTEINS

1. Septin-2, Human Diff6, CDC10 homologue (SEPT2 , down-modulated more than 10 fold in liver metastasis)

Septins are an evolutionarily conserved family of genes that encode a P loop-based GTP-binding domain flanked by a polybasic domain and (usually) a coiled-coil region. They have roles in cytokinesis, vesicle trafficking, polarity determination, membrane diffusion barriers formation, as well as in microtubule and actin dynamics. Septins can form hetero-oligomeric complexes and possibly function as dynamic protein scaffolds. Recently, it has been shown that there are at least 13 human septin genes that exhibit extensive alternate splicing. There are complex patterns of human septin gene expression and recently it has been found that alterations in septin expression are seen in human diseases including neoplasia [11]. It was found that a high level of Septin-2 mRNA expression in NB was associated with younger age at diagnosis (≤ 1 year of age), favorable clinical stage (I, II, and IVs), whereas a low level of Septin-2 mRNA expression was suggested to be associated with the progression of clinical stage and N-myc gene amplification [12].

2. Raf kinase inhibitor protein (RKIP, down-modulated 2,6 fold in liver metastasis)

Raf kinase inhibitor protein (RKIP), a member of the phosphatidyl ethanolamine binding protein (PEBP) family, is a small, cytosolic protein originally purified from bovine brain [13-15]. The RKIP family of proteins is highly conserved and does not share significant homology with any other protein family [16]. Recent investigations have identified that RKIP inhibits Raf-mediated activation of MEK, which accounts for its current name. Aberrant RKIP activity is associated with an increasing number of diseases *via* its association with signal transduction pathways.

RKIP blocks the activation of several signaling pathways including MEK, G-proteins and NF- κ B [17]. Immunohistochemical analysis of prostate cancer primary tumors and metastases revealed that RKIP protein expression was decreased in metastases. These results demonstrate that RKIP is a metastasis suppressor gene (MSG) which suggest that it or

proteins it interacts with are putative molecular targets to control metastasis. Loss of RKIP enhances both angiogenesis and vascular invasion, and protects against apoptosis. Moreover RKIP expression is decreased in metastases of prostate cancer patients, compared to normal prostate or the primary prostate tumor [18].

3. S100 calcium binding protein A11-Calgizzarin (S100A11, down-modulated 2,7 fold in liver metastasis)

The protein encoded by this gene is a member of the S100 family of proteins containing 2 EF-hand calcium-binding motifs. S100 proteins regulate numerous intracellular functions, which include protein phosphorylation, enzyme activation, interaction with cytoskeletal components, and calcium homeostasis. In addition, it is currently thought that S100 proteins are involved in the regulation of many cellular processes such as cell cycle progression and differentiation. For example, S100A11 has been implicated in growth inhibition of human fibroblasts and in Ca^{2+} -induced growth inhibition of human keratinocytes in culture. Moreover down-regulation of S100A11 was found associated with bladder cancer progression and poor survival [19].

4. Calreticulin precursor (CARL, down-modulated 2,2 fold in liver metastasis)

CARL is a widely-expressed calcium-binding protein found mainly in the endoplasmic reticulum, but also in other cellular compartments. It is also found in the nucleus, suggesting that it may have a role in transcription regulation. It may inhibit target gene transcription by interacting with steroid hormone receptors, thereby masking their DNA-binding sites and triggering the onset of the apoptotic process. Reduced levels of CARL were observed in highly metastatic MHCC97-H cells [20]. This result is in accordance with the report that calreticulin and calreticulin fragments inhibit angiogenesis and suppress tumor growth [21]. The down-regulation of CARL was also demonstrated in hepatoma cells [22] and in colorectal cancer [23].

4.3.1.2 DIFFERENTIALLY UP-MODULATED PROTEINS

1. 26S proteasome regulatory chain, p40 (PSMD7, up-modulated more than 10 fold in liver metastasis)

The 26S proteasome is a highly conserved multicatalytic protease responsible for ATP- and ubiquitin-dependent degradation of all short-lived and of 70–90% of all long-lived proteins including cyclin A, B and E, p21 and p27, p53, cJun, cFos, and I κ B. As such, the 26S proteasome controls cell cycle, signal transduction pathways, apoptosis, angiogenesis and major functions of the immune system [24]. The proteasome is a potential target for pharmacological agents. Bortezomib, the first proteasome inhibitor to reach clinical trials, has shown *in vitro* and *in vivo* activity against a variety of malignancies, including myeloma, chronic lymphocytic leukemia, prostate cancer, pancreatic cancer, and colon cancer. The proteasome inhibitor bortezomib exhibits antiproliferative, proapoptotic, antiangiogenic, and antitumor activities in several cancer models. Proteasome inhibitors block apoptosis in thymocytes and in primary neuronal cells, conversely, they induce apoptosis in a variety of different tumor cells. Thus, the increased sensitivity of transformed cells to the transient proteasome inhibition provides a reasonable therapeutic window for the use of proteasome inhibitors in cancer therapy [25].

2. Phosphoglycerate kinase 1 (PGK1, up-modulated 5,2 fold in liver metastasis)

PGK1 is known as the sixth enzyme of the glycolytic pathway, in which it equilibrates transfer between position 1 of 1,3-bisphosphoglycerate and the γ -phosphate of MgATP²⁻. PGK1 also influences DNA replication and repair in mammalian cell nuclei. It is also secreted by tumor cells and participates in tumor angiogenesis as a disulfide reductase [26]. The hypoxic nature of solid tumors triggers expression of vascular endothelial cell growth factor, which in turn stimulates both angiogenesis and glycolytic enzyme activity, including that of PGK1, with the ability to facilitate anaerobic production of ATP [27].

3. Eukaryotic translation initiation factor 4E (eIF-4E, up-modulated 2,7 fold in liver metastasis)

Eukaryotic translation initiation factor 4E (eIF-4E) is a 25 kDa messenger RNA cap-binding phosphoprotein and is involved in the initiation of protein synthesis. Although eIF-4E regulates the recruitment of mRNA to ribosomes, and thereby globally regulates cap-dependent protein synthesis, eIF-4E contributes to malignancy by selectively enabling the translation of a limited pool of mRNAs, those that generally encode key proteins involved in cellular growth, angiogenesis, survival and malignancy (*e.g.*, cyclin D1, c-myc, vascular

endothelial growth factor, matrix metalloprotease 9). The contribution of the mRNA cap-binding protein, eIF-4E, to malignant transformation and progression has been illuminated over the past decade. eIF-4E overexpression has been demonstrated in human tumors of the breast, head and neck, colon, prostate, bladder, cervix and lung, and has been related to disease progression. Overexpression of eIF-4E in experimental models dramatically alters cellular morphology, enhances proliferation and induces cellular transformation, tumorigenesis and metastasis. Conversely, blocking eIF-4E function by expression of antisense RNA, or overexpression of the inhibitory eIF-4E binding proteins (4E-BPs), suppresses cellular transformation, tumor growth, tumor invasiveness and metastasis [28]. A deeper understanding of the role of eIF-4E in regulating the translation of the diverse gene products involved in all aspects of malignancy will improve the capacity to exploit eIF-4E as a therapeutic target and as a marker for human cancer progression [29].

4. Peroxiredoxin 1 (PRX1 up-modulated 2,2 fold in liver metastasis) and Peroxiredoxin 2 (PRX2 up-modulated 2,4 fold in liver metastasis)

Peroxiredoxins (PRXs) are a novel group of peroxidases endowed with high antioxidant efficiency and some of them having also effects on cell differentiation and apoptosis [30]. The mammalian PRX family has six distinct members located in various subcellular locations, including peroxisomes and mitochondria, places where oxidative stress is most evident. There is fastly growing evidence that oxidative stress is important not only for normal cell physiology but also for many pathological processes like atherosclerosis, neurodegenerative diseases, allergies, and cancer. ROS participate in carcinogenesis in all its stages, *e.g.*, initiation, promotion, and progression.

PRX1 participates in the signaling cascades of growth factors and tumor necrosis factor- α by regulating the intracellular concentration of H_2O_2 [31]. These facts suggest that PRX1 activity may be associated with, not only proliferation, but also with formation of reactive oxygen species (ROS). ROS are produced by inflammatory cell infiltration and indicate a cellular response. PRX1 has been demonstrated to be overexpressed in follicular thyroid neoplasms and thyroiditis [32], malignant mesothelioma [33], lung cancer [34], in breast cancer [35], and also in a xenografted human breast cancer resistant to tamoxifen [36]. PRX2 is the immediate enzyme that links reduction of H_2O_2 to thioredoxin. Might participate in the signaling cascades of growth factors and tumor necrosis factor- α by regulating the intracellular concentrations of H_2O_2 .

5. Y Box binding protein 1 (YB1, up-modulated 2,4 fold in liver metastasis)

YB1 is a member of the highly conserved Y-box family of proteins that regulate gene transcription by binding to either double- or single-stranded TAACC elements (the Y box) contained within many eukaryotic promoters [37-38]. Y-boxes are located on the promoter of numerous genes, such as DNA topoisomerase II alpha (Topo IIalpha), proliferating cell nuclear antigen (PCNA) and multidrug resistance 1 (MDR1). YB1 is activated in response to genotoxic stress and is associated with drug resistance, and both YB1 protein level and nuclear localization appear to be prognostic for some human cancers. In addition, YB1 regulates expression of several tumor associated genes. These include epidermal growth factor receptor (EGFR or c-ErbB1), matrix metalloproteinase-2 (MMP-2), fas, mdr1, DNA topoisomerase II and MHC Class II. YB1, acts as a negative regulator of p53. The p53 tumor suppressor plays a major role in preventing tumor development by transactivating genes to remove or repair potentially tumorigenic cells. Using reporter assays, it was shown that YB1 represses transcription of the p53 promoter in a sequence-specific manner. YB1 reduces endogenous levels of p53, which in turn reduces p53 activity. Conversely, inhibiting YB1 in a variety of tumor cell lines induces p53 activity, resulting in significant apoptosis via a p53-dependent pathway. These data suggest that YB1 may be a good target for the development of new therapeutics [39]. In agreement with these findings, high levels of YB1 expression in breast cancer is associated with tumor aggressiveness and poor prognosis [40].

4.3.2 WESTERN BLOT ANALYSIS

We used Western blots for validating protein identification and expression patterns from two-dimensional gels by using antibodies specific to the candidates proteins and anti- β actin antibody for the normalization of the optical density values (Figure 6). Each Western blot was replicated three times. These proteins, namely RKIP and SEPT2 were selected because they could be putative molecular targets able to control the metastatic process, PSMD7 and eIF-4E because they could be potential therapeutic targets and OP18 for its implication in migration and metastasis of tumor cells [41].

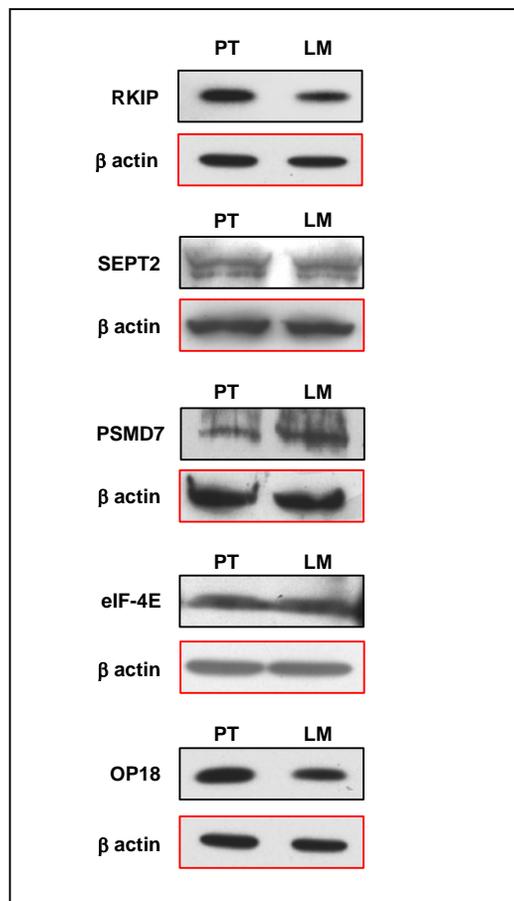


Figure 6. SDS-PAGE followed by Western blot and immuno-detection of the proteins Raf kinase inhibitor protein (RKIP), Septin 2 (SEPT2), 26S proteasome regulatory chain, p40 (PSMD7), Eukariotic translation initiation factor 4E (eIF-4E), Stathmin (OP18) in primary tumor (PT) and liver metastasis (LM).

Although for the majority of analyzed proteins the trends of variation (up or down regulation) were confirmed by Western blot analysis, data did not show a statistically significant variation. These results suggested that Western blot (with SDS-PAGE) is not an optimal method to validate proteomics results. Actually the different isoforms of one protein (separated and identified by 2D-PAGE) can not be observed by a Western blot analysis of a monodimensional SDS-PAGE where all the isoforms of one protein are collected in the same band. Furthermore the Western blot analysis (even if performed with replicate samples and replicate experiments) is only semi-quantitative, since the transfer of a protein onto PVDF membrane and also the development of photographic films to detect ECL signal, give variable results due to technical variability.

4.4 CONCLUDING REMARKS

Current research data demonstrate the value of proteomic analysis for understanding the molecular mechanism involved in metastasis activity. The proteomic methodology of 2-dimensional polyacrylamide gel electrophoresis/mass spectrometry also provides an improved opportunity to understand the natural history of NB and develop novel chemotherapeutic agents for this prevalent childhood malignancy with a dismal outcome.

In this study we were able to identify a lot of proteins involved in the process of metastasis. Among them there were some of particular interest: at least three potential therapeutic targets have been identified, namely the 26S proteasome regulatory chain, the Eukaryotic translation initiation factor 4E, and the Y Box binding protein-1. Moreover, at least two putative molecular targets able to control the metastatic process have been also identified in the Septin-2 and the Raf kinase inhibitor proteins and finally Stathmin for its implication in migration and metastasis of tumor cells was also identified. Further evaluation and characterization of the genomic and proteomic variations may lead to the identification of biomarkers that can be specifically applied to metastatic assay and diagnosis in clinic. In addition, integration of proteomics with other experimental disciplines, particularly biochemistry, cell biology, molecular genetics and chemistry, will continue to extend the application of proteomics in the study of tumor metastasis.

4.5 BIBLIOGRAPHY

- [1] Everley, P. A., Setter, B. R., *Ann. N. Y. Acad. Sci.* 2005, **1059**, 1-10.
- [2] Cai, Z., Chiu, J. F., He, Q. Y., *Genomics Proteomics Bioinformatic* 2004, **2**, 152-166.
- [3] Chambers, A. F., Groom, A. C., MacDonald, I. C., *Nat. Rev. Cancer* 2002, **2**, 563-572.
- [4] Kang, Y., Siegel, P. M., Shu, W., Drobnjak, M., Kakonen, S. M., Cordon-Cardo, C., Guise, T. A., Massague, J., *Cancer Cell* 2003, **3**, 537-549.
- [5] Chung, L. W., Baseman, A., Assikis, V., Zhau, H. E., *J. Urol.* 2005, **173**, 10-20.
- [6] Pastorino, F., Brignole, C., Marimpietri, D., Cilli, M., Gambini, C., Ribatti, D., Longhi, R., Allen, T. M., Corti, A., Ponzoni, M., *Cancer Res.* 2003, **63**, 7400-7409.
- [7] Herbert, B., Galvani, M., Hamdan, M., Olivieri, E., McCarthy, J., Pedersen, S., Righetti, P. G., *Electrophoresis* 2001, **22**, 2046.
- [8] Righetti, P. G., *Immobilized pH Gradients: Theory and Methodology*. Elsevier, Amsterdam, 1990, pp. 1-400.
- [9] Shevchenko, A., Wilm, M., Vorm, O., Mann, M., *Anal. Chem.* 1996, **68**, 850-858.
- [10] Rappsilber, J., Ishihama, Y., Mann, M., *Ana. Chem.* 2003, **75**, 663-670.
- [11] Russell, S. E., Hall, P. A., *Br. J. Cancer.* 2005, **93**, 499-503.
- [12] Nagata, T., Takahashi, Y., Asai, S., Ishii, Y., Mugishima, H., Suzuki, T., Chin, M., Harada, K., Koshinaga, S., Ishikawa, K., *J. Surg. Res.* 2000, **92**, 267-275.
- [13] Simister, P. C., Banfield, M. J., Brady, R. L., *Acta Crystallogr. D Biol. Crystallogr.* 2002, **58**, 1077-1080.
- [14] Vallee, B. S., Tauc, P., Brochon, J. C., Maget-Dana, R., Lelievre, D., Metz-Boutigue, M. H., Bureaud, N., Schoentgen, F., *Eur. J. Biochem.* 2001, **268**, 5831-5841.
- [15] Yeung, K., Seitz, T., Li, S., Janosch, P., McFerran, B., Kaiser, C., Fee, F., Katsanakis, K. D., Rose, D. W., Mischak, H., Sedivy, J. M., Kolch, W., *Nature* 1999, **401**, 173-177.
- [16] Banfield, M. J., Barker, J. J., Perry, A. C., Brady, R. L., *Structure* 1998, **6**, 1245-1254.
- [17] Keller, E. T., *Anti-Cancer Drugs* 2004, **15**, 663-669.
- [18] Keller, E. T., Fu, Z., Yeung, K., Brennan, M., *Cancer Lett.* 2004, **207**, 131-137.
- [19] Memon, A. A., Sorensen, B. S., Meldgaard, P., Fokdal, L., Thykjaer, T., Nexø, E., *Clin. Cancer Res.* 2005, **11**, 606-611.
- [20] Ding, S. J., Li, Y., Shao, X. X., Zhou, H., Zeng, R., Tang, Z. Y., Xia, Q. C., *Proteomics* 2004, **4**, 982-994.
- [21] Pike, S. E., Yao, L., Jones, K. D., Cherney, B., Appella, E., Sakaguchi, K., Nakhasi, H., Teruya-Feldstein, J., Wirth, P., Gupta, G., Tosato, G., *J. Exp. Med.* 1998, **188**, 2349-2356.
- [22] Yu, L. R., Zeng, R., Shao, X. X., Wang, N., Xu, Y. H., Xia, Q. C., *Electrophoresis* 2000, **21**, 3058-3068.

- [23] Alfonso, P., Nunez, A., Madoz-Gurpide, J., Lombardia, L., Sanchez, L., Casal, J. I., *Proteomics* 2005, **5**, 2602-2611.
- [24] Fekete, M. R., McBride, W. H., Pajonk, F., *BMC Cancer* 2005, **5**, 114.
- [25] Rajkumar, S. V., Richardson, P. G., Hideshima, T., Anderson, K. C., *J. Clin. Oncol.* 2005, **23**, 630-639.
- [26] Lay, A. J., Jiang, X. M., Kisker, O., Flynn, E., Underwood, A., Condron, R., Hogg, P. J., *Nature* 2000, **408**, 869-873.
- [27] Shichijo, S., Azuma, K., Komatsu, N., Ito, M., Maeda, Y., Ishihara, Y., Itoh, K., *Clin. Cancer Res.* 2004, **10**, 5828-5836.
- [28] De Benedetti, A., Graff, J. R., *Oncogene* 2004, **23**, 3189-3199.
- [29] Mamane, Y., Petroulakis, E., Rong, L., Yoshida, K., Ler, L. W., Sonenberg, N., *Oncogene* 2004, **23**, 3172-3179.
- [30] Karihtala, P., Mantyniemi, A., Kang, S. W., Kinnula, V. L., Soini, Y., *Clin. Cancer Res.* 2003, **9**, 3418-3424.
- [31] Kang, S., Chae, H., Seo, M., Kim, K., Baines, I., Rhee, S., *J. Biol. Chem.* 1998, **273**, 6297-6302.
- [32] Haklar, G., Sayin-Ozveri, E., Yuksel, M., Aktan, A. O., Yalcin, A. S., *Cancer Lett.* 2001, **165**, 219-224.
- [33] Kinnula, V. L., Lehtonen, S., Sormunen, R., Kaarteenaho-Wiik, R., Kang, S. W., Rhee, S. G., Soini, Y., *J. Pathol.* 2002, **196**, 136-323.
- [34] Chang, J. W., Jeon, H. B., Jeung, H. L., Yoo, J. S., Chun, J. S., Kim, J. H., Yoo, Y., *J. Biochem. Biophys. Res. Commun.* 2001, **289**, 507-512.
- [35] Noh, D. Y., Ahn, S. J., Lee, R. A., Kim, S. W., Park, I. A., Chae, H. Z., *Anticancer Res.* 2001, **21**, 2085-2090.
- [36] Besada, V., Diaz, M., Becker, M., Ramos, Y., Castellanos-Serra, L., Fichtner, I., *Proteomics* 2006, **6**, 1038-1048.
- [37] Wolffe, A. P., *Bioessays* 1994, **16**, 245-251.
- [38] Mantovani, R., *Nucleic Acids Res.* 1998, **26**, 1135-1143.
- [39] Lasham, A., Moloney, S., Hale, T., Homer, C., Zhang, Y. F., Murison, J. G., Braithwaite, A. W., Watson, J. J., *Biol. Chem.* 2003, **278**, 35516-35523.
- [40] Huang, J., Tan, P. H., Li, K. B., Matsumoto, K., Tsujimoto, M., Bay, B. H., *Int. J. Oncol.* 2005, **26**, 607-613.
- [41] Iancu-Rubin, C., Atweh, G. F., *Trends Cell Biol.* 2005, **15**, 346-348.

CHAPTER 5

PROTEOMIC ANALYSIS OF ANTI-ANGIOGENIC EFFECTS BY A COMBINED TREATMENT WITH VINBLASTINE AND RAPAMYCIN IN AN ENDOTHELIAL CELL LINE

5.1 ANGIOGENESIS

Angiogenesis, a biological process by which new capillaries are formed from pre-existing vessels, is an important natural process used by the body for reproduction and for healing injured tissues [1]. It plays a fundamental role in the neoplastic process and is essential for the local progression and metastatic spread of solid and hematologic tumors. In solid tumor growth, a specific clinical turning point is the transition from the avascular to the vascular phase [2]. Having developed an intrinsic vascular network, the neoplastic mass is able to grow indefinitely (unlike all the other forms, tumor angiogenesis is not limited in time) both *in situ* and at distant sites (metastasis) since its cells enter the vascular bed and colonize other organs [3]. New vessels promote tumor cell growth by carrying oxygen and nutrients and removing catabolites, whereas endothelial cells (ECs) secrete growth factors for tumor cells and a variety of matrix-degrading proteinases that facilitate invasion (Figure 1).

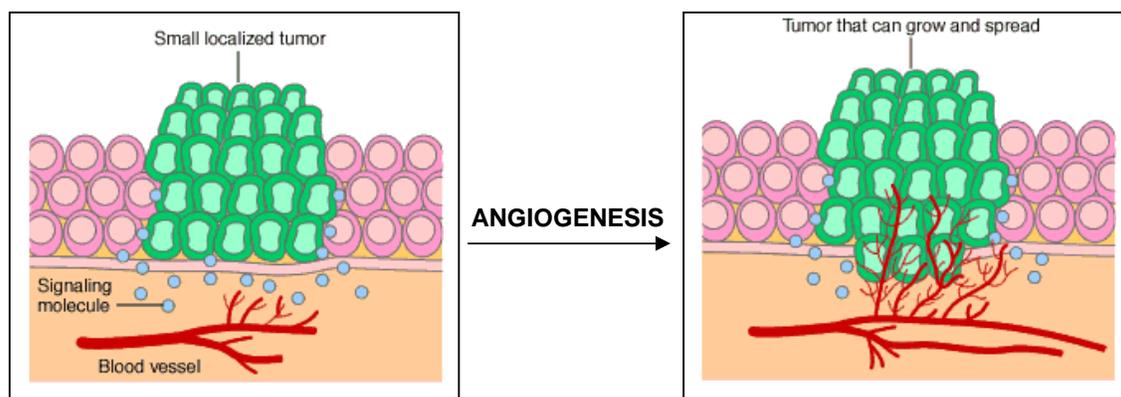


Figure 1. The development of new capillaries from pre-existing vessels.

An expanding endothelial surface also gives tumor cells more opportunities to enter the circulation and metastasize, whereas the release of antiangiogenic factors by tumor cells may account, at least in part, for control exerted by primary tumors over metastasis. Tumor angiogenesis depends mainly on the release by neoplastic cells of growth factors specific for endothelial cells and able to stimulate growth of the host's blood vessels [4]. Activators of endothelial cell proliferation and migration are mainly receptor tyrosine kinase ligands, such as vascular endothelial growth factor (VEGF), fibroblast growth factors (FGFs), platelet derived growth factor (PDGF), and epidermal growth factor (EGF).

5.1.1 ANTI ANGIOGENESIS THERAPY

Anti-angiogenesis management should be effective against many tumors, and the development of drugs that inhibit angiogenesis has become an attractive approach to cancer therapy [5]. Long-term, regular low-dose administration of such agents ("metronomic", *i.e.* very frequent or continuous low-dose chemotherapy) could inhibit EC proliferation, angiogenesis, and tumor growth [6] and, by attacking the vasculature, could circumvent chemotherapy-induced drug resistance, because ECs are considered to be genetically more stable than tumor cells [7]. Much evidence, mostly *in vitro*, indicates that the 'activated' endothelial cells of newly forming blood-vessel capillaries are highly and selectively sensitive to very low doses of various chemotherapeutic drugs [8-9]. For example, several studies have been undertaken to test the antiproliferative, migration-inhibitory and sometimes cytotoxic effects of picomolar concentrations of chemotherapeutic drugs on various human cell types, including fibroblasts, lymphocytes, tumour cells, epithelial cells from various tissues, and microvascular or macrovascular endothelial cells. Some of the most interesting studies involve various microtubule inhibitors, such as vinblastine, paclitaxel and docetaxel. In that experiments, ultra-low concentrations of these drugs were reported to inhibit proliferation or migration of endothelial cells, but not of other cell types examined. For example, Wang and coworkers reported that 10–100,000-fold higher concentrations of paclitaxel were required to inhibit proliferation and migration of human astrocytes, fibroblasts, mammary epithelial cells, keratinocytes, prostate epithelial cells or smooth-muscle cells, compared with epithelial cells [10]. Taxanes, however, must be formulated in certain vehicles for injection, to prevent their binding to serum proteins. Clinically relevant concentrations of these vehicles or binding proteins can significantly inhibit the anti-angiogenic activity of taxanes, meaning that higher doses of such injectable taxanes would have to be used *in vivo* to induce anti-angiogenic effects [11]. Moreover antiangiogenic therapy is preferable to conventional chemotherapy because the last one kills all dividing cells in the body, aiming to do the most damage to dividing cancer cell. However, normally dividing cells (hair cells,

intestinal cells, mucous membranes, bone marrow cells) are also destroyed. This leads to well-known severe side effects of chemotherapy: hair loss, diarrhea, mouth ulcer, infection, and low blood counts. It has been suggested that combination of two or more angiogenesis inhibitors may be more effective than single-drug treatment. Combination of inhibitors with different mechanism of action attacking different biochemical pathways should shut down the multifactor stimulated cascade of biochemical angiogenic processes.

In the present study two drugs (Vinblastine and Rapamycin) were evaluated for their anti-angiogenic synergistic activity on an endothelial cell line (Figure 2).

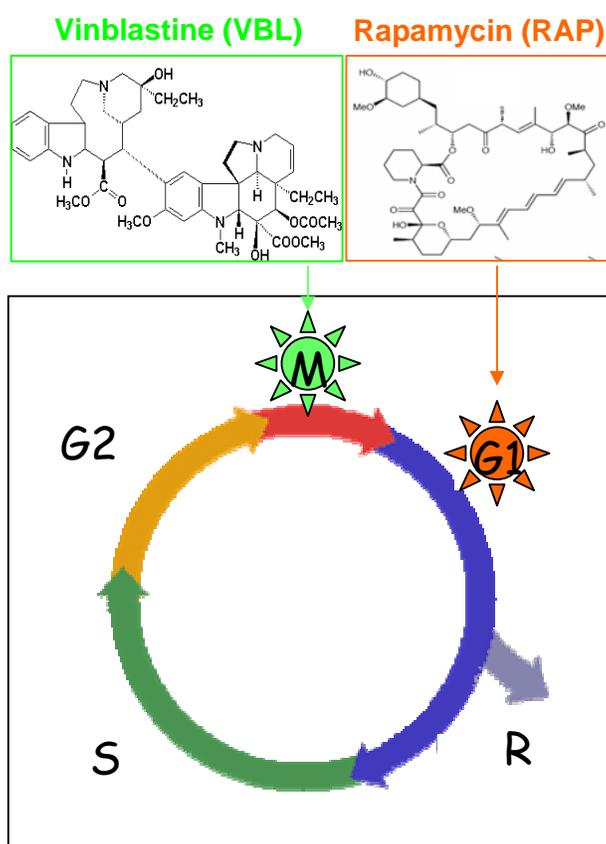


Figure 2. Anti-angiogenic drugs Vinblastine (VBL) and Rapamycin (RAP) and their point of action in the cell cycle.

Vinblastine (VBL) is a chemotherapeutic agent which is used to treat many kinds of cancer (lymphoma, Hodgkin's disease, testicular cancer, breast cancer, choriocarcinoma, mycosis fungoides, Kaposi's sarcoma). This alkaloid binds to tubulin, thus preventing the cell from making the spindles it needs to be able to move its chromosomes around as it divides and it also seem to interfere with cells' ability to synthesize DNA and RNA. Rapamycin (RAP) is a macrolide antibiotic with immunosuppressant properties that prevents rejection of organ transplants in many animals. Rapamycin inhibits the activity of the protein, mTOR, (mammalian target of rapamycin) which functions in a signaling pathway to promote tumor growth. These two drugs exert their antiangiogenic activity with different mechanisms [9,12].

Previously Marimpietri and coworkers have observed that VBL and RAP display synergistic effects in Neuroblastoma as antiangiogenic agents both *in vitro* and *in vivo* [13]. Each compound alone was able to induce a significant dose- and time-dependent response by inhibiting *in vitro* endothelial cells (EA.hy926 and HUVEC) growth, with a synergistic effect, when both drugs were combined. Morphological changes were induced by each molecule in ECs and their combination resulted in a clear and stronger effect. Moreover, apoptosis was found on endothelial cells treated with VBL, showing an increase in the percentage of cells with a G2-M and sub-G1 DNA content, whereas in those treated with RAP a block in the G1 cell fraction and inhibition of progression to the S phase were observed. The drug combination resulted in a synergistic cell cycle arrest and induction of apoptosis [13].

The present study deals with the differential protein expression profiles obtained from an EA.hy926 endothelial cell line untreated or treated with VBL and/or RAP by using two-dimensional gel electrophoresis. Mass spectrometry analysis identified 54 of the 113 differentially expressed polypeptide chains. Some of the identified proteins are involved in the processes of angiogenesis, proliferation, migration and apoptosis and are regulated according to the antiangiogenic effect induced by VBL and RAP alone or in combination.

5.2 EXPERIMENTAL PROCEDURES

5.2.1 CHEMOTHERAPEUTICS

VBL (Lilly France SA, Saint Cloud, Paris, France) was solubilized in phosphate-buffered saline (PBS) and diluted with the culture medium. RAP (ICN Biomedicals Inc., South Chillicothe Road Aurora, OH, USA) was solubilized in DMSO and diluted with the culture medium.

5.2.2 CELL CULTURE

The human endothelial-like immortalized cell line EA.hy926 [14] was maintained in logarithmic growth in 75-cm² plastic culture flasks (Corning, Bibby Sterilin, UK) in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% Fetal bovine serum (FBS) (Sigma Chemical Co., St Louis, MO, USA), 50 IU/mL sodium penicillin G, 50 µg/mL streptomycin sulphate, and 2 mM L-glutamine, as previously described by Ponzoni and collaborators [15].

5.2.3 PROLIFERATION ASSAY

ECs were plated (2×10^3 /well) in 96-well plates in their complete medium. After 24h, medium was removed and replaced in quadruplicate by the same medium (control) or supplemented with 1 nM VBL and 100 nM RAP. The cell number was estimated after 24 and 48 hours by the colorimetric assay of Kueng and coworkers [16] modified by Marimpietri [13]. Briefly, cells were fixed for 20 min at room temperature with 4% paraformaldehyde (PFA), stained with 0.1% crystal violet in 20% methanol for 20 min, washed in PBS, solubilized with 10% acetate and read at 595 nm in a microplate reader (SpectraFluor Plus, Tecan, Austria). The cell number was derived from a calibration curve set up with known numbers of cells plated in sextuplicate.

5.2.4 PHOSPHATIDYLSERINE EXPOSURE

Phosphatidylserine (PS) exposure was assessed with a human Annexin V-FITC Kit (Bender MedSystems, Vienna Austria). Briefly, cultured and treated cells with 1nM VBL and/or 100 nM RAP for 24 hours were collected, washed and incubated for 10 min with 5 µL Annexin V-FITC, washed once with PBS, resuspended in 190 µL prediluted binding buffer containing 10 µL of a 20 µg/mL propidium iodide (PI) stock solution and examined by two color flow cytometry using a FACScan (Beckton Dickinson, San Jose, CA) equipped with a xenon lamp and a filter set for excitation at 488 nm; FITC and PI fluorescence intensities were recorded through a 520-530 (FL1-H) and 575 (FL2-H) nm filters respectively. At least 10,000 events were collected in each dot plot and analyzed using Cell Quest software (Beckton Dickinson, San Jose, CA). Cells that stained negative for both Annexin V-FITC and PI staining were viable and did not undergo measurable apoptosis; those that stained positive for Annexin V-FITC and negative for PI were in the early stages of apoptosis; cells that stained positive for both Annexin V-FITC and PI were either in the later stages of apoptosis or dead.

5.2.5 2D-PAGE

Protein extraction from cells untreated, treated with 1 nM VBL for 24 h, and/or 100 nM RAP for 24 h was performed with lysis buffer (3% w/v CHAPS, 7 M urea, 2 M thiourea, 40 mM Tris, 5 mM tributylphosphine, 0.5% carrier ampholytes pH 3–10 and a protease inhibitor cocktail). Cells were left in lysis buffer for 30 min in ice and the cell lysates were prepared by sonication for 5 periods of 30 seconds. After centrifugation at 13,000 $\times g$ at 4 °C (for removal of particulate material) the protein solutions were collected and the supernatants were reduced and alkylated as described by Herbert and collaborators [17]. Nonlinear immobilised pH gradient strips [18] (18 cm long, pH 3–10; Bio-Rad, Hercules, CA, USA) were rehydrated for 8 h with 450 μ L of the sample solution (final total protein concentration of 1.8 mg/mL) containing trace of bromophenol blue for monitoring the electrophoretic run. IEF was conducted at 20°C for 75,000 Vh using a Protean IEF Cell (Bio-Rad, Hercules, CA, USA) with a low initial voltage (1,000 V) and then by applying a voltage gradient up to 10,000 V with a limiting current of 50 μ A/strip. For the second dimension, the IPG strips were equilibrated for 27 min by rocking in a solution containing: 375 mM Tris-HCl (pH 8.8), 6 M urea, 20% glycerol (Sigma), 2% SDS (Fluka). 10–20%T gradient SDS-PAGE gel slabs were prepared in-house, by using the Protean Multi-Gel Casting Chamber (Bio-Rad) and the IPG strips were then laid on top of the gels with 0.8% agarose dissolved in the cathode buffer (192 mM glycine, 15 mM Tris, 0.1% SDS, pH 8.3). Gels were run at 11°C and 2 mA/gel for 2 h, 5 mA/gel for 1 h and 10 mA/gel overnight in a Protean Plus Dodecacell apparatus (Bio-Rad), hosting twelve gels simultaneously. The protein zones were revealed with Sypro Ruby stain [19].

5.2.6 PROTEIN PATTERN DIFFERENTIAL ANALYSIS

The 2-DE gels were scanned with a Versa Doc (Bio-Rad), and analyzed with the software PDQuest Version 7.3 (Bio-Rad, Labs, Hercules, CA). A match set was created from the protein patterns of six replicate gels for each independent cellular extract (control, VBL, RAP, VBL+RAP). A standard gel was generated out of the image with the highest spot number. Spot quantities of all gels were normalized so as to remove non expression-related variations in spot intensity. Then, the raw quantity of each spot in a gel was divided by the total quantity of all the spots in that gel that have been included in the standard. The results were evaluated in terms of spot OD (optical density). Statistical analysis of PDQuest by *t*-test allowed the study of proteins that were significantly increased or decreased after the treatments of the cell line. A *p*-value <0.05 was considered as significant. A total of 113 spots were found to be differentially expressed.

5.2.7 PROTEIN IDENTIFICATION BY MASS SPECTROMETRY

The spots of interest were excised from the gel with a razor blade, placed in Eppendorf tubes, and destained by washing for 5 min with 50 μ L of 0.1 M NH_4HCO_3 . After that, 50 μ L of 100% acetonitrile were added to each tube and let to incubate for 5 min. The liquid was discarded and the washing step was repeated one more time. After washing, the gel plugs were shrunk by addition of acetonitrile. The dried gel pieces were re-swollen with 7.0 ng/ μ L trypsin (Promega, Madison, WI) in 50 mM NH_4HCO_3

and digested overnight at 37°C. Peptides were concentrated with ZipTip®µC₁₈ pipette tips. Co-elution was performed directly onto a MALDI target with 1 µL of α-cyano-4-hydroxycinnamic acid matrix (5 mg/mL in 50% acetonitrile, 0.1% TFA).

MALDI-MS and MALDI-MS/MS were performed on an Applied Biosystems 4700 Proteomics Analyzer with TOF/TOF ion optics. Data were acquired in positive MS reflector mode with five spots of standard (ABI4700 Calibration Mixture) for calibration. Mass spectra were obtained from each sample spot by 30 sub-spectra accumulation (each consisting of 50 laser shots) in a 750 to 4,000 mass range. Five signal-to-noise best peaks of each spectrum were selected for MS/MS analysis. For MS/MS spectra, the collision energy was 1 keV and the collision gas was air.

The interpretation of both the MS and MS/MS data was carried out by using the GPS Explorer software (Version 1.1, Applied Biosystems) which acts as an interface between the Oracle database containing raw spectra and a local copy of the MASCOT search engine (Version 1.8). Peptide mass fingerprints obtained from MS analysis were used for protein identification in the NCBI non-redundant database. All peptide mass values are considered monoisotopic and mass tolerance was set between 50 and 100 ppm. Trypsin was given as the digestion enzyme, 1 missed cleavage site was allowed, methionine was assumed to be partially oxidized and serine, threonine and tyrosine partially phosphorylated. Mascot (Matrix Science) scores greater than 61 were considered to be significant ($p < 0.005$). For MS/MS analysis, all peaks with a signal-to-noise ratio greater than 5 were searched against the NCBI database using the same modifications as the MS database. Fragment tolerance less than 0.3 Da was considered [20].

5.2.8 BIOCHEMICAL FUNCTION AND PATHWAY ANALYSIS

Proteins functions were obtained from the SWISS-PROT Protein knowledge base (when available). Additionally, the software package PathwayAssist (Stratagene, La Jolla, CA) was used to identify functional relationships among the proteins. This software explores gene interaction networks represented in the ResNet database. ResNet is a comprehensive database of molecular networks compiled by proprietary natural language processing techniques applied to the whole PubMed database. The database contains more than 100,000 events of regulation, interaction and modification between 15,000 proteins, cell processes and small molecules.

5.2.9 STATISTICS

All *in vitro* data are from three or more independent experiments. Results are expressed as means ± 95% confidence intervals. The significance of differences between experimental groups and controls was determined by the ANOVA with Tukey's multiple comparison test in GraphPad Prism. These findings were considered significant if *P* values were less than 0.05.

5.3 RESULTS

5.3.1 INHIBITION OF CELL PROLIFERATION

We first investigated the antiproliferative effects on EA.hy926 cells following exposure to VBL and RAP alone or in combination. A time-dependent antiproliferative effect was obtained with each single compound; however they did not induce complete inhibition of EA.hy926 cells after 48 hours. The combination of 1 nM VBL and 100 nM RAP resulted in a significant lower percentage of proliferating cells after 24 hours and an almost complete cell growth inhibition after 48 hours (Figure 3). Synergy is defined as a combination of two drugs that has a greater therapeutic effect than that expected by the simple additive effects of each drug. To determine whether our drug scheduling exhibited increased synergistic antiproliferative activity, combined data of VBL and RAP were plotted in an isobologram generated from the dose-response curves where the IC_{80} was calculated. The values derived from the combination treatment fell below the theoretical line indicating a synergistic activity. To rule out the possibility of specimen heterogeneity, the ANOVA test was applied to calculate the significance of the different effects, indicating a statistically significant increment when the two drugs were added in combination (Figure 3).

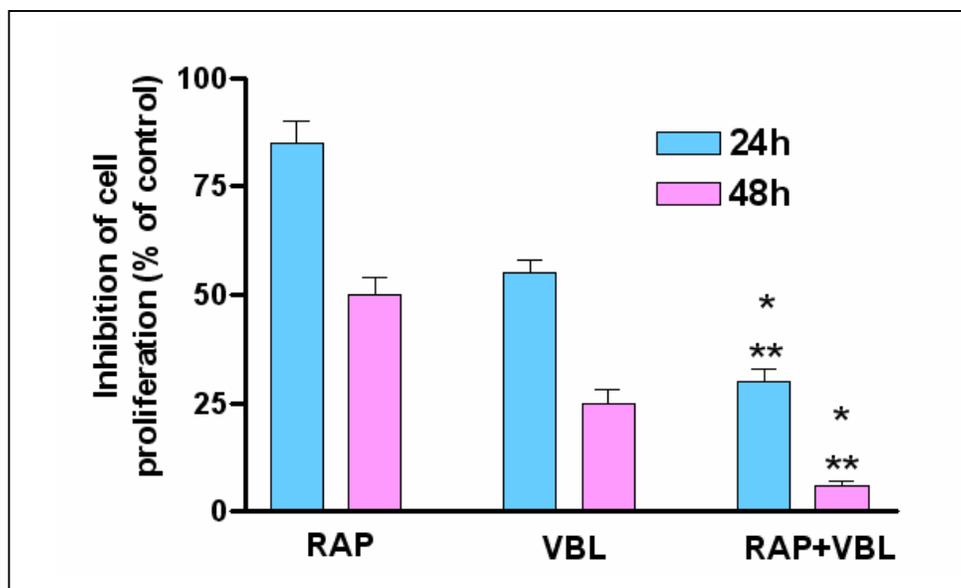


Figure 3. Synergistic effect of VBL and RAP on the growth of the endothelial cell line EA.hy926. EA.hy926 cells were exposed for 24 and 48 hours to complete medium alone (control) or supplemented with 1 nM VBL and/or 100 nM RAP. Cell proliferation was measured by staining with crystal violet in quadruplicate wells. Results are expressed as a mean percentage of optical density in the control wells. *, significantly different from VBL treatment at $P < 0.05$; ** $P < 0.01$ vs RAP treatment. Each column represents mean \pm 95% confidence intervals. The experiments were repeated three times with similar results.

5.3.2 INDUCTION OF APOPTOSIS

To determine whether the observed response to VBL and RAP treatments occurred via induction of apoptosis, EA.hy926 cells were exposed for 24 hours with 1 nM VBL and 100 nM RAP, either alone or in combination. A clear quantification of apoptosis percentage was achieved by exposing the EA.hy926 cell line to the drugs, followed by flow-cytometry after double annexin V-FITC conjugate and PI staining. As shown in Figure 4, FACS analysis revealed negligible apoptosis in control untreated cells and only a marginal apoptotic effect following treatment with RAP, whereas induction of apoptosis was observed in cells treated with VBL. The combination of the two molecules resulted in a significantly higher percentage of apoptotic cells, demonstrating a synergistic pro-apoptotic effect on EA.hy926 cells after 24 hours of treatment with Vinblastine and Rapamycin.

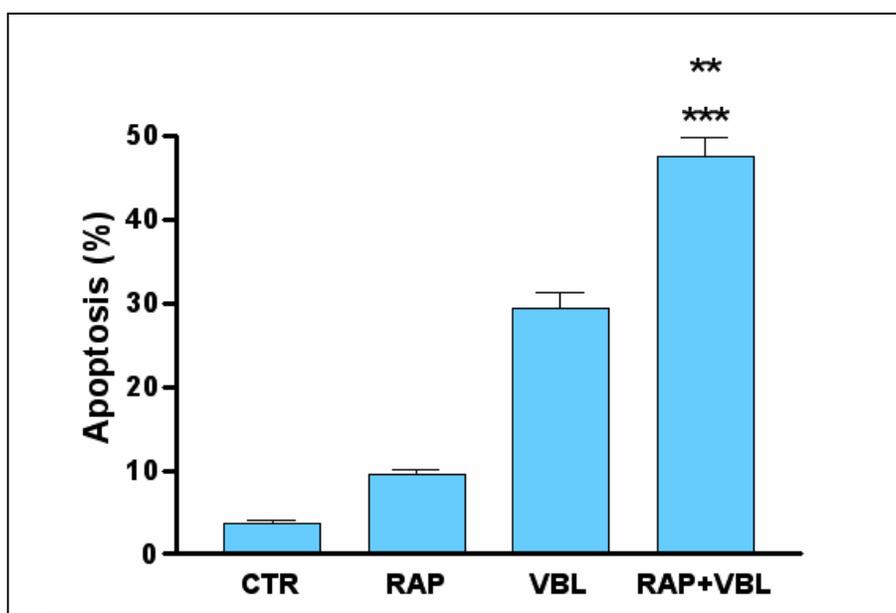


Figure 4. Vinblastine and Rapamycin trigger apoptosis in EA.hy926 cells. Cells were incubated for 24 hours with complete medium (control, CTR) or supplemented with 1 nM VBL and 100 nM RAP alone or in combination. Induction of apoptosis was evaluated by phosphatidylserine exposure, measured by flow cytometric analysis of annexin V (FITC) and Propidium Iodide double staining. A synergistic apoptotic effect was observed in cells treated with VBL and RAP in combination. Results are expressed as the percentage of apoptosis, and represent the average \pm 95% confidence intervals of three independent experiments (** $P < 0.01$ vs VBL; *** $P < 0.001$ vs CTR and RAP).

5.3.3 DIFFERENTIAL PROTEOMIC ANALYSIS AND PATHWAY ANALYSIS

Proteomics has been used for the direct determination of the effect of drugs at the molecular level. Differential 2D maps were generated for controls and cells treated with VBL and RAP alone and in combination. The differentially expressed proteins were identified *via* mass spectrometry, and their functional associations and common targets were analyzed *via* dedicated software.

We measured differential protein expression between control cells and cells treated with VBL or RAP alone or in combination. Representative two-dimensional maps of the four conditions are reported in Figure 5.

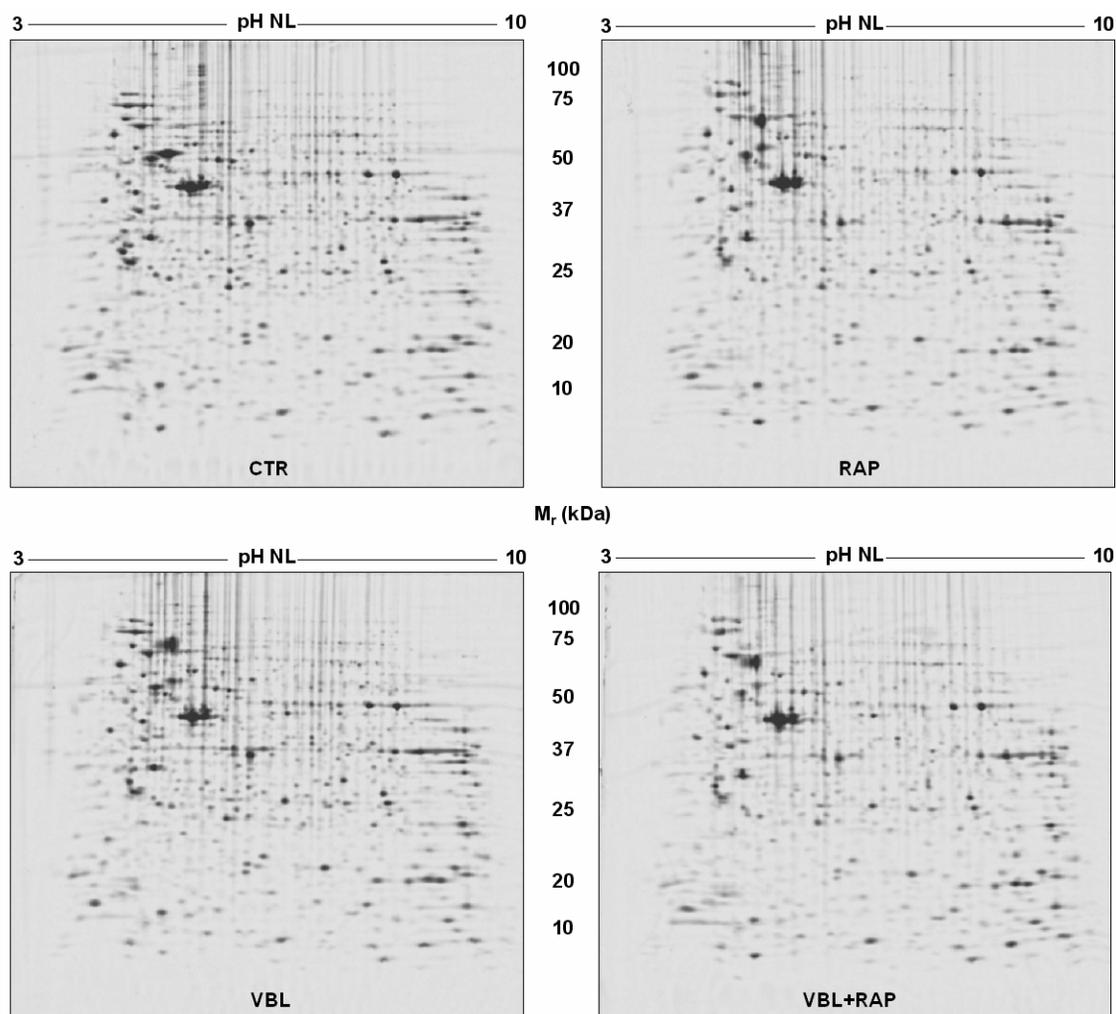


Figure 5. Two-dimensional maps of EA.hy926 cell line either untreated or treated with 1 nM VBL and/or 100 nM RAP for 24 h (CTR = control).

A total of 113 spots were identified as differentially expressed after VBL and/or RAP treatments (41 up-regulated and 72 down-regulated) (Figure 6), by applying a threshold of 2.0 fold. Out of the 113 differentially expressed spots, 54 were identified by MS analysis and were found to collapse into 42 unique proteins, while the remaining differentially expressed spots were not identified because of their relatively low concentrations.

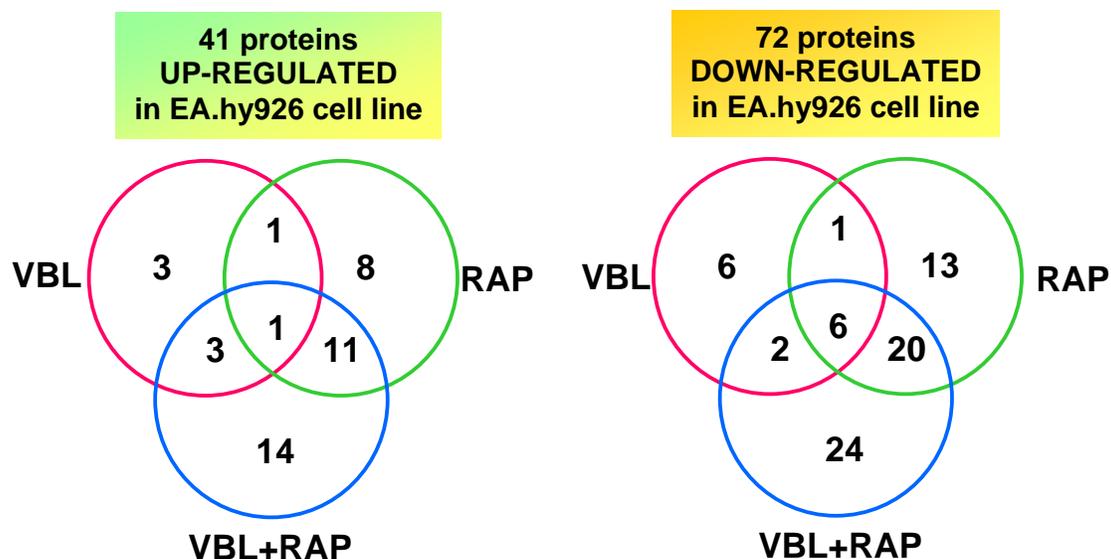


Figure 6. Summary of the 113 differentially expressed proteins after the indicated treatments. Numbers of the overlapping circles in the Venn diagrams indicate the numbers of spots which are regulated by more than one treatment.

In Figure 7, the identified spots are marked with the numbers corresponding to the standard spot number (SSP) in the Table 1.

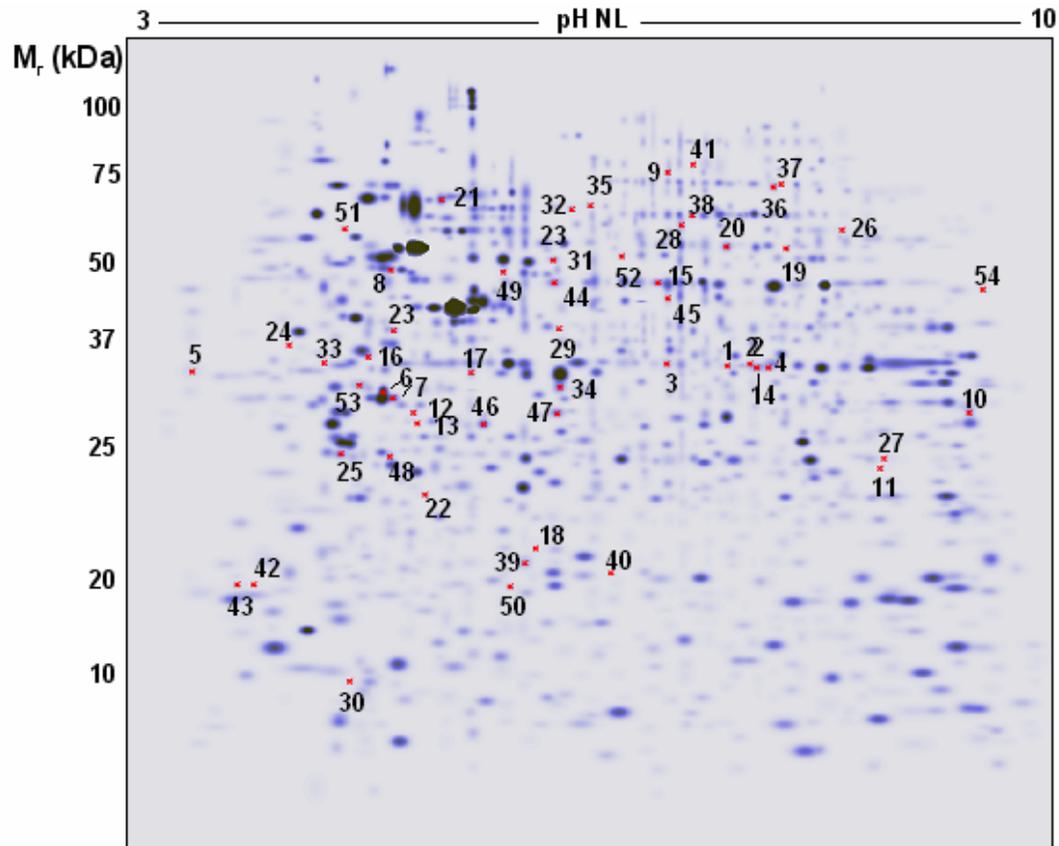


Figure 7. Standard map of the EA.hy926 endothelial cell line treated with VBL and RAP alone and in combination. The identified proteins are marked with the numbers corresponding to the standard spot number (SSP) in the Table 1.

Table 1. Identified proteins from the EA.hy926 endothelial cell line 2D gels.

Protein name	SSP	HUGO gene name	NCBI accession number	Prot. score	Pept. count	Fold of variations in VBL	Fold of variations in RAP	Fold of variations in VBL+RAP
Annexin A1	1	ANXA1	gi 442631	118	18		> 10	
Annexin A1	2	ANXA1	gi 442631	107	16		> 10	> 10
Annexin A1	3	ANXA1	gi 442631	86	15		↓ 2.9	↓ 2.0
Annexin A2 isoform 1	4	ANXA2	gi 50845388	79	17		> 10	> 10
Annexin A2, isoform 2	5	ANXA2	gi 50845388	74	18		↓ 2.5	↓ 7.1
Annexin A5	6	ANXA5	gi 17391477	231	20		> 10	> 10
Annexin A5	7	ANXA5	gi 17391477	82	21			↑ 2.0
ATP synthase beta subunit	8	ATP5B	gi 1374715	195	20	< 10	< 10	< 10
Caldesmon 1 isoform 2	9	CALD1	gi 4826657	72	21			↓ 2.5
Carbonyl reductase 1	10	CBR1	gi 54696548	65	12		> 10	> 10
CTCL tumor antigen HD-CL-01	11	C10orf118	gi 21262188	87	26		< 10	< 10
EF-hand domain-containing protein 2	12	EFHD2	gi 20149675	78	12		< 10	< 10
EF-hand domain-containing protein 2	13	EFHD2	gi 20149675	169	16		↑ 3.8	↑ 3.2
ERC protein 2	14	ERC2	gi 51701377	63	31		↓ 2.7	↓ 4.3
Eukaryotic translation Elongation factor 1-gamma	15	EEF1G	gi 4503481	151	16	↓ 2.2	↓ 3.0	↓ 3.8
Eukaryotic translation Elongation factor 1-gamma	16	EEF1G	gi 4503481	234	10		< 10	< 10
F-actin capping protein alpha-1 subunit	17	CAPZA1	gi 5453597	137	7		< 10	< 10
Ferritin, light polypeptide	18	FTL	gi 18203882	78	3	> 10		
Glucose-6-phosphate-dehydrogenase	19	G6PD	gi 26224874	147	25		↓ 3.7	↓ 5.2
Glucose-6-phosphate-dehydrogenase	20	G6PD	gi 26224874	73	16		↓ 2.5	↓ 2.1
Heat shock 70 kDa protein 8	21	HSPA8	gi 24234686	198	17	↓ 2.0	↓ 2.1	↓ 6.3
Hematological and neurological expressed 1 isoform 2	22	C16orf34	gi 50345294	179	7	↓ 2.5		↓ 2.7
Heterogeneous nuclear ribonucleoprotein C isoform a	23	HNRPC	gi 14110428	75	9			↓ 2.5
Heterogeneous nuclear ribonucleoprotein C isoform b	24	HNRPC	gi 4758544	87	7	↑ 2.8	< 10	
HIST1H4F	25	HIST1H4F	gi 4504305	115	6	> 10		> 10
Histone H1.0	26	H1F0	gi 472449	64	10		↑ 3.5	↑ 4.9
Hydroxyacyl-CoA dehydrogenase type II	27	HADH2	gi 4758504	143	11		↑ 2.1	↑ 2.5

Hypothetical protein	28	CAH18390	gi 51476843	86	18	< 10	< 10
Hypothetical protein FAM33A	29	FAM33A	gi 17389709	67	11	< 10	< 10
Keratin, type I cytoskeletal 10	30	KRT10	gi 40354192	69	21	↓ 2.1	< 10
Keratin, type II cytoskeletal 8	31	KRT8	gi 39645331	186	28		↓ 3.4
Lamin A/C isoform 1 precursor	32	LMNA	gi 27436946	165	31	> 10	> 10
LIM and SH3 protein 1	33	LASP1	gi 545310	78	9		↓ 2.2
L-lactate dehydrogenase B chain	34	LDHB	gi 5496396	66	13	↓ 2.9	
LMNA protein	35	LMNA	gi 30584609	146	27	↓ 2.0	↓ 2.0
LMNA protein	36	LMNA	gi 30584609	164	33		↑ 2.2
LMNA protein	37	LMNA	gi 30584609	245	36	↑ 4.8	
LMNA protein	38	LMNA	gi 30584609	105	23		↓ 2.0
Membrane-type 1 matrix metalloproteinase cytoplasmic tail binding protein-1	39	MTCBP-1	gi 8922762	78	13	↓ 2.5	
MGC5987 protein	40	MGC5987	gi 28422494	67	6	↑ 5.0	↑ 4.8
Minichromosome maintenance protein 7 isoform 1	41	MCM7	gi 33469968	99	28	↓ 2.6	
Myosin regulatory light chain MRCL3	42	MRCL3	gi 5453740	170	8	↓ 2.0	
Myosin regulatory light chain MRCL3	43	MRCL3	gi 5453740	170	11	↑ 4.5	↑ 3.1
Proteasome 26S non-ATPase subunit 2	44	PSMD2	gi 4506209	114	20	↓ 3.1	↓ 2.5
Proteasome 26S non-ATPase regulatory subunit 9	45	PSMD9	gi 2150046	63	18	↓ 4.8	↓ 2.6
Proteasome activator subunit 2	46	PSME2	gi 30410792	144	14	↓ 2.7	↓ 2.0
Proteasome activator subunit 3	47	PSME3	gi 30410796	416	16	↓ 2.6	↓ 2.4
Rho GDP-dissociation inhibitor (GDI) alpha	48	ARHGDI	gi 30582607	136	10	> 10	> 10
RuvB-like 2	49	RUVBL2	gi 5730023	97	16	↓ 2.3	< 10
Stathmin 1	50	OP18	gi 15680064	94	9		↓ 2.1
TPM4-ALK fusion oncoprotein type 2	51	TPM4-ALK	gi 10441386	197	12	> 10	> 10
Tryptophanyl-tRNA synthetase	52	WARS	gi 47419916	79	15		↓ 2.3
Ubiquitin aldehyde binding 1	53	OTUB1	gi 8923114	64	10	↓ 4.2	↓ 4.2
Vasodilator-stimulated phosphoprotein	54	VASP	gi 11414808	82	13		↓ 3.4

To search for functional associations and common targets of the identified proteins, we used the PathwayAssist software and the data obtained are shown in Figure 8.

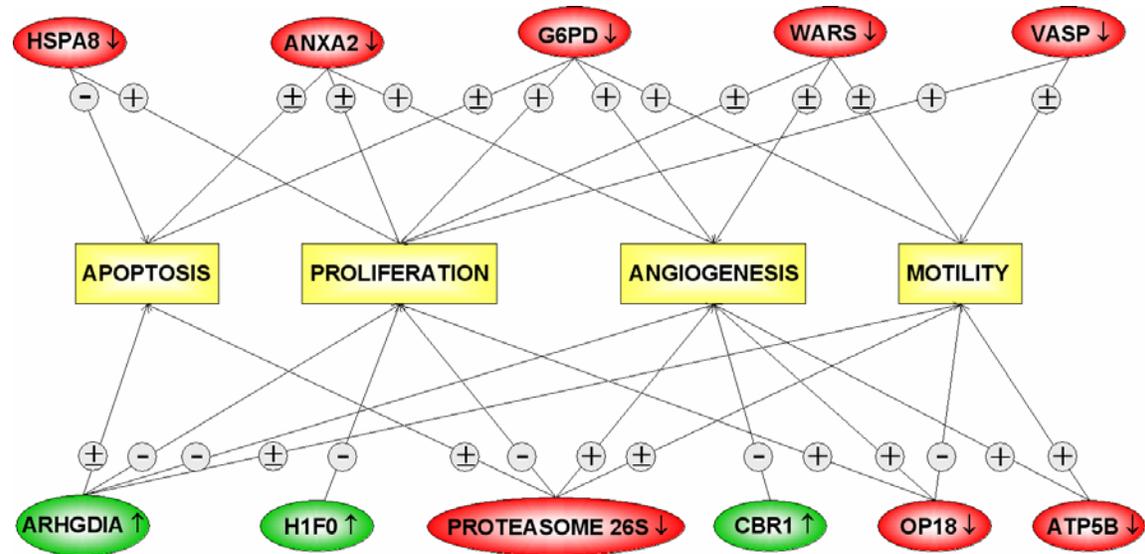


Figure 8. Common targets (cell processes) for identified proteins modulated by VBL+RAP. To bring all the relevant information about proteins together, we have used the PathwayAssist software (Stratagene) for generating a network of molecules that have common targets (angiogenesis, proliferation, apoptosis, motility). The activating, inhibiting or unclear connections are indicated with the symbols (+), (-), (±), respectively. The meaning of the different shape is the following: ellipse → identified proteins, rectangle → cell processes.

5.4 DISCUSSION

In this study, we demonstrate that the chemotherapeutic agent Vinblastine or the antibiotic Rapamycin drastically inhibit endothelial cell proliferation by inducing apoptosis and that the combination of the two drugs displays synergistic effects. These findings confirm previous results of Marimpietri and coworkers [13] indicating the potency of this drug combination in inhibiting Neuroblastoma-related angiogenesis. Indeed, here and in the manuscript of Marimpietri and collaborators [13], it is clearly shown the strong antiangiogenic effect of a combination of VBL and RAP at different doses against human endothelial cells. This activity was greater than the sum of the effects of the two drugs administered individually. These experimental data agree with the general understanding that antiangiogenic agents targeting a single pathway or growth factor may display preclinical efficacy, but usually confer limited clinical benefit. It is clear that combinatorial approaches are needed.

To substantiate this last statement at the molecular level, we used a proteomic approach to identify proteins differentially expressed in ECs after combined treatment with VBL and RAP. Among the proteins identified, the most notable groups were proteins thought to be involved in angiogenesis and processes related to angiogenesis such as endothelial cell proliferation, migration, and apoptosis. The role of the most important proteins detected is briefly discussed below. We will focus only on the proteins synergistically down- or up-modulated by the combined treatment with respect to the single drugs.

5.4.1 DIFFERENTIALLY DOWN-MODULATED PROTEINS

1. ATP synthase beta subunit (ATP5B down-modulated more than 10 fold by VBL+RAP)

This gene encodes a subunit of mitochondrial ATP synthase. Mitochondrial ATP synthase catalyzes ATP synthesis, utilizing an electrochemical gradient of protons across the inner membrane during oxidative phosphorylation. ATP synthase is composed of two linked multi-subunit complexes: the soluble catalytic core, F₁, and the membrane-spanning component, F₀, comprising the proton channel. This enzyme produces ATP from ADP in the presence of a proton gradient across the membrane. The beta chain is the catalytic subunit. Angiogenesis requires ATP. Blood vessel cells need chemical energy to grow and ATP synthase provides chemical energy to the cell [21].

2. Annexin A2 isoform 2 (ANXA2 down-modulated 7.1 fold by VBL+RAP)

Annexin A2 belongs to the annexin multigene family of proteins that are able to bind calcium and phospholipids in the presence of calcium. It has been suggested that these proteins may play a role in membrane trafficking events such as exocytosis, endocytosis and cell-cell adhesion. They may provide a major pathway for communication between cellular membranes and their cytoplasmic environment [22]. Annexin II is the endothelial receptor for S100A4 and their interaction triggers the functional activity directly related to pathological properties of S100A4. The S100A4 protein, also known as Mts-1 or metastasin, belongs to the S100 family of Ca_{2+} -binding proteins that are implicated in a variety of cellular events, including growth, signalling, differentiation, and motility. S100A4 and several other S100 proteins, such as S100A2, S100A6, and S100B, have attracted significant attention because of their established significance in the progression of metastatic tumors. Semov and coworkers [23] reported for the first time annexin II as the endothelial receptor of S100A4 and link the interaction between the two proteins to the t-PA-mediated plasminogen activation. Plasmin formed during this reaction is the major component of tissue fibrinolysis process that is closely linked to the known pathogenic roles of S100A4, including angiogenesis, metastasis, and inflammation. The angiogenic effects of S100A4 are induced through its interaction with annexin II (ANXA2) on the surface of endothelial cells. Extracellular S100A4 stimulates the production of ECM degrading enzymes from endothelial cells, thereby stimulating the remodeling of ECM. This could explain the angiogenic and metastasis-stimulating activity of S100A4 (mts1) [24].

Annexin II is a proposed endothelial cell coreceptor for plasminogen and t-PA. In vitro, this calcium regulated phospholipid-binding protein possesses t-PA cofactor activity and increases the catalytic efficiency (kcat/Km) of t-PA-dependent plasmin generation by 60- fold. In a recent work it was reported that postnatal angiogenesis is markedly impaired in the absence of annexin II, most likely due to a failure to localize plasmin activity on the endothelial cell surface and a failure to activate selected matrix metalloproteinases (MMPs) [25].

Finally, the primary in vivo function of Plasmin (Pm) is to regulate vascular potency by degrading fibrin-containing thrombi. However, the identification of Pg/Pm receptors and the ability of Pm to degrade other matrix proteins have implicated Pm in other functions involving degradation of protein barriers, thereby mediating cell migration, an important event in a number of normal (such as embryogenesis, wound healing, angiogenesis) and pathological (e.g., tumor growth and dissemination) processes [26]. Plasmin has broad substrate specificity and degrades several ECM components, including fibrin, fibronectin, laminin and the protein core of proteoglycans. In addition, plasmin may activate several matrix metalloproteinases such as MMP-1, MMP-3 and MMP-9.

3. Heat shock 70 kDa protein 8 (HSPA8 down-modulated 6.3 fold by VBL+RAP)

Heat shock proteins (HSPs) are molecular chaperones that are induced by various environmental and pathophysiological stimuli. Among them, the 70-kDa heat shock protein (HSP70) has been shown to play an important role in the regulation of cell growth and transformation [27]. It has been reported that various malignant tumors over-express HSP70, which relates closely to tumorigenesis, malignant development, tumor immunity, resistance to apoptosis and a poor prognosis in the clinical course. HSP70 is involved in signal transduction pathways in cooperation with HSP90 and is additionally associated with the mitotic spindle and cell proliferation [28-29].

4. Glucose-6-phosphate dehydrogenase (G6PD, down-modulated 5.2 fold by VBL+RAP)

Glucose-6-phosphate dehydrogenase, the first and rate-limiting enzyme in the pentose phosphate pathway, is the principal intracellular source of NADPH. NADPH, in turn, is utilized directly as a cofactor for eNOS, and, indirectly, to maintain levels of another important cofactor, tetrahydrobiopterin, via *de novo* synthesis and the dihydrofolate reductase salvage pathway. In this manner, G6PD regulates eNOS activity and NO• levels. In this study, we demonstrate that G6PD activity modulates endothelial cell migration, proliferation, and tube formation, by mediating NO• levels. G6PD may, therefore, serve as a novel regulatory determinant of the angiogenic phenotype. G6PD significantly influenced VEGF-mediated vascular endothelial cell proliferation, migration, and tube formation in an *in vitro* cell culture model and an *in vivo* murine model. Decreased G6PD expression and activity were associated with an impaired response to VEGF-stimulated cell proliferation, migration, and formation of tubes and networks in a Matrigel matrix. In contrast, increased G6PD expression and activity enhanced these processes. G6PD modulated these events, in part, by influencing basal and VEGF stimulated eNOS activity and NO• levels.

The mechanism(s) and/or effectors by which G6PD enhances acquisition of an angiogenic phenotype remain incompletely characterized; however, it has been demonstrated that intracellular NADPH levels, and not the production of riboses by G6PD, is the critical determinant of cell growth. This suggests that one mechanism by which G6PD regulates endothelial cell migration, proliferation, and tube formation is by synthesizing NO• via eNOS, which has an absolute requirement for NADPH as a cofactor [30].

5. Vasodilator stimulated phosphoprotein (VASP, down-modulated 3.4 fold by VBL+RAP)

Vasodilator stimulated phosphoprotein is a member of the Ena-VASP family and is a 46 kDa membrane associated protein that was first described in human platelets [31]. VASP is associated with focal adhesions and areas of dynamic membrane activity where it is thought to have an important role in actin filament assembly and cell motility. Previous studies performed in many systems have suggested that this family of proteins has a universal role in the control of cell motility and intracellular actin dynamics via a linear pathway from the receptor-ligand interaction. Moreover, VASP phosphorylation correlates well with the reversible inhibition of integrin $\alpha 2\beta 3$ (also known as fibrinogen receptor or platelet glycoprotein IIb-IIIa). In the intracellular environment, VASP is associated with filamentous actin formation and is involved in intracellular signalling pathways of the integrin-extracellular matrix (ECM) interaction. It is therefore suggested that this molecule might play a part in cell adhesion and motility. Its expression is elevated during the endothelial reorganization phase of capillary morphogenesis *in vitro* [32]. This phase of angiogenesis requires endothelial cells to migrate, involving peripheral actin assembly, as well as to undergo shape change, a process in which actin bundles would be expected to contribute to individual cell shape changes involved in lumen formation. Studies have suggested that VASP may be involved in trophoblast invasion, angiogenesis, cell proliferation and tumorigenesis. VASP participates in vasculogenesis and angiogenesis, by regulating endothelial cell migration [33].

6. Proteasome 26S non-ATPase subunits 2 and 9 and Proteasome activator subunit 2 and 3 (PROTEASOME 26S, down-modulated ~ 2.4 fold by VBL+RAP)

The 26S proteasome is a highly conserved multicatalytic protease responsible for ATP- and ubiquitin-dependent degradation of all short-lived and of 70–90% of all long-lived proteins including cyclin A, B and E, p21 and p27, p53, cJun, cFos, and I κ B. As such, the 26S proteasome controls cell cycle, signal transduction pathways, apoptosis, angiogenesis and major functions of the immune system [34]. The proteasome is a potential target for pharmacological agents. Bortezomib, the first proteasome inhibitor to reach clinical trials, has shown *in vitro* and *in vivo* activity against a variety of malignancies, including myeloma, chronic lymphocytic leukemia, prostate cancer, pancreatic cancer, and colon cancer. The proteasome inhibitor bortezomib exhibits antiproliferative, proapoptotic, antiangiogenic, and antitumor activities in several cancer models.

Proteasome inhibitors block apoptosis in thymocytes and in primary neuronal cells, conversely, they induce apoptosis in a variety of different tumor cells. Thus, the increased sensitivity of transformed cells to the transient proteasome inhibition provides a reasonable therapeutic window for the use of proteasome inhibitors in cancer therapy [35-36].

7. Tryptophanyl-tRNA synthetase (WARS, down-modulated 2.3 fold by VBL+RAP)

Aminoacyl-tRNA synthetases are key enzymes in protein biosynthesis that catalyze aminoacylation of their cognate tRNAs. Mammalian tryptophanyl-tRNA synthetases (TrpRSs) have an appended domain at their NH₂ termini that is absent from lower eukaryotic and prokaryotic TrpRSs. During their long evolution, aminoacyl-tRNA synthetases, enzymes that catalyze the first step of protein synthesis, acquired additional functions, including regulation of transcription and translation, RNA splicing, and cytokine activities in inflammatory and angiogenic signaling pathways. Recently, fragments of the closely related human tyrosyl-tRNA and tryptophanyl-tRNA synthetase (TrpRS) were discovered to regulate angiogenesis. In mammalian cells, TrpRS is activated for anti-angiogenic signalling by proteolysis or alternative splicing to generate two natural isoforms—mini-TrpRS and T2-TrpRS. While T2-TrpRS is an antagonist of vascular endothelial growth factor-induced angiogenesis, the isoform mini-TrpRS has clear an angiogenic activity [37-38].

8. Stathmin (OP18, down-modulated 2.1 fold by VBL+RAP)

Stathmin/OP18 is a 19-kDa cytoplasmic protein, whose function is related to microtubule turnover. Stathmin/OP18 promotes tubulin catastrophes by binding to a microtubule plus end, which stimulates the exposed β -subunit of the terminal tubulin dimer to hydrolyze its GTP, causing tubulin to undergo a conformational change that favors transition to depolymerization. Thus, stathmin/OP18 increases microtubule turnover and contributes to rapid reorganization of the microtubule cytoskeleton. Microtubules are essential for a wide variety of cellular functions including mitosis, cell shape, and cell motility. Stathmin/OP18 should have influences on these phenomena. The inhibition of stathmin/OP18 inhibits the migration of ECs. More importantly, the inhibition of stathmin/OP18 expression leads to the decrease of cell division and subsequent apoptosis. Because stathmin/OP18 inhibition stabilizes microtubules, stathmin/OP18 inhibits also angiogenesis [39].

5.4.2 DIFFERENTIALLY UP-MODULATED PROTEINS

1. Carbonyl reductase 1 (CBR, up-modulated more than 10 fold by VBL+RAP)

Carbonyl reductase (CBR) is a cytosolic NADPH-dependent oxidoreductase metabolizing prostaglandins, steroids, quinines, and anthracycline antibiotics. Many experimental studies have shown that CBR plays important roles in the regulation of tumor progression. Expression of CBR mRNA was inversely associated with angiogenesis and tumor progression and was a significant prognostic factor in non-small-cell lung cancer [40].

2. Rho GDP-dissociation inhibitor (GDI) alpha (ARHGDIA, up-modulated more than 10 fold by VBL+RAP)

Rho-GDI regulates the GDP/GTP exchange reaction of the Rho proteins by inhibiting the dissociation of GDP from them, and the subsequent binding of GTP to them. Rho family members are small GTPases known to regulate malignant transformation and motility of cancer cells. The increase in the amount of Rho GTPases appears to be a frequent event in different types of human tumors, and may contribute to the aggressive nature of tumor cells, as well as to tumor-induced biological processes involved in disease progression. Rho-GDI has been implicated in tumor cell apoptosis, invasion and metastases. Reduced expression of Rho-GDI in breast tumor tissues is correlated with the nodal involvement and metastasis [41].

The inhibition of angiogenesis could be related to 1) an increase in genes involved in the inhibition of cell proliferation (p19(INK4), p21(Waf/Cip1), Wnt-5a), in the inhibition of cell migration (Rho-GDI 1, alpha E-catenin) and 2) a downregulation of genes involved in angiogenesis (PAI-1, Vitronectin, HoxD3, Notch4) or in cell invasion (Semaphorin E).

3. Histone H1.0 (H1F0, up-modulated 4.9 fold by VBL+RAP)

Histones are basic nuclear proteins that are necessary for the condensation of nucleosome chains into higher order structures. The h1/h1.0 histones are found in cells that are in terminal stages of differentiation or that have low rates of cell division. They belong to the histone h1/h5 family. The H1F0 expression levels inversely correlate with cell proliferation [42].

5.5 CONCLUSIONS

As a summary of our data, we demonstrate that a combined treatment with Vinblastine and Rapamycin exhibits a synergistic effect in inhibiting the *in vitro* growth of endothelial cells by inducing apoptosis. By using two-dimensional tools and MS analysis, we identified a set of proteins that may be involved in the cooperation of Vinblastine and Rapamycin in inhibiting, among other effects, angiogenesis, which represents the primary cause of aggressive tumor growth. Interestingly, many proteins are regulated only by the combined treatment and appear to be involved in processes of proliferation, migration, and apoptosis, in addition to strong anti-angiogenic effects. A similar approach was recently attempted on a pancreatic ductal carcinoma cell line, that was subjected to a combined treatment with gemcitabine (a drug interfering with DNA synthesis) and trichostatin A (a drug interfering with histone acetylation), resulting in a marked synergistic inhibition of proliferation and induction of apoptosis [43]. Altogether, these findings might have important applications in future cancer treatment to design molecular-guided experimental therapies, especially in cases that are resistant to all currently used treatments.

5.6 BIBLIOGRAPHY

- [1] Risau, W., *Nature* 1997, **386**, 671-674.
- [2] Hanahan, D., Folkman, J., *Cell* 1996, **86**, 353-364.
- [3] Fidler, I. J., Ellis, L. M., *Cell* 1994, **79**, 185-188.
- [4] Folkman, J., *Cancer Medicine*, Williams & Wilkins, Baltimore, 1997, pp.181-204.
- [5] Folkman, J., *N. Engl. J. Med.* 1971, **285**, 1182-1186.
- [6] Kerbel, R. S., Vilorio-Petit, A., Klement, G., Rak, J., *Eur. J. Cancer* 2000, **36**, 1248-1257.
- [7] Kerbel, R. S., *J. Clin. Oncol.* 2001, **19**, 45S-51S.
- [8] Hirata, S., Matsubara, T., Saura, R., Tateishi, H., Hirohata, K., *Arthritis Rheum.* 1989, **32**, 1065-1073.
- [9] Vacca, A., Iurlaro, M., Ribatti, D., Minischetti, M., Nico, B., Ria, R., Pellegrino, A., Dammacco, F., *Blood* 1999, **94**, 4143-4155.
- [10] Wang, J., Lou, P., Lesniewski, R., Henkin, J., *Anticancer Drugs* 2003, **14**, 13-19.
- [11] Ng, S. S., Figg, W. D., Sparreboom, A., *Cancer Res.* 2004, **64**, 821-824.
- [12] Guba, M., Von Breitenbuch, P., Steinbauer, M., Koehl, G., Flegel, S., Hornung, M., Bruns, C. J., Zuelke, C., Farkas, S., Anthuber, M., Jauch, K. W., Geissler, E. K., *Nature Med.* 2002, **8**, 128-135.
- [13] Marimpietri, D., Nico, B., Vacca, A., Mangieri, D., Catarsi, P., Ponzoni, M., Ribatti, D., *Oncogene* 2005, **24**, 6785-95.
- [14] Edgell, C. J., McDonald, C. C., Graham, J. B., *Proc. Natl. Acad. Sci. U.S.A.* 1983, **80**, 3734-3737.
- [15] Ponzoni, M., Bocca, P., Chiesa, V., De Censi, A., Pistoia, V., Raffaghello, L., Rozzo, C., Montaldo, P. G., *Cancer Res.* 1995, **55**, 853-861.
- [16] Kueng, W., Silber, E., Eppenberger, U., *Anal. Biochem.* 1989, **182**, 16-19.
- [17] Herbert, B., Galvani, M., Hamdan, M., Olivieri, E., McCarthy, J., Pedersen, S., Righetti, P. G., *Electrophoresis* 2001, **22**, 2046-2057.
- [18] Gianazza, E., Giacon, P., Sahlin, B., Righetti, P. G., *Electrophoresis* 1985, **6**, 53-56.
- [19] Berggren, K. N., Schulenberg, B., Lopez, M. F., Steinberg, T. H., Bogdanova, A., Smejkal, G., Wang, A., Patton, W. F., *Proteomics* 2002, **2**, 486-498.
- [20] Medzihradzky, K. F., Campbell, J. M., Baldwin, M. A., Falick, A. M., Juhasz, P., Vestal, M. L., Burlingame, A. L., *Anal. Chem.* 2000, **72**, 552-558.
- [21] Moser, T. L., Stack, M. S., Asplin, I., Enghild, J. J., Hojrup, P., Everitt, L., Hubcnak, S., Schnaper, H. W., Pizzo, S. V., *Proc. Natl. Acad. Sci. U.S.A.* 1999, **96**, 2811-2816.
- [22] Gerke, V., Moss, S. E., *Physiol. Rev.* 2002, **82**, 331-371.
- [23] Semov, A., Moreno, M. J., Onichtchenko, A., Abulrob, A., Ball, M., Ekiel, I., Pietrzynski, G., Stanimirovic, D., Alakhov, V., *J. Biol. Chem.* 2005, **280**, 20833-20841.

- [24] Schmidt-Hansen, B., Ornas, D., Grigorian, M., Klingelhofer, J., Tulchinsky, E., Lukanidin, E., Ambartsumian, N., *Oncogene* 2004, **23**, 5487-5495.
- [25] Ling, Q., Jacovina, A. T., Deora, A., Febbraio, M., Simantov, R., Silverstein, R. L., Hempstead, B., Mark, W. H., Hajjar, K. A., *J. Clin. Invest.* 2004, **113**, 38-48.
- [26] Castellino, F. J., Ploplis, V. A., *Thromb. Haemost.* 2005, **93**, 647-654.
- [27] Kaur, J., Kaur, J., Ralhan, R., *Int. J. Cancer* 2000, **85**, 1-5.
- [28] Vargas-Roig, L. M., Fanelli, M. A., Lopez, L. A., Gago, F. E., Tello, O., Aznar, J. C., Ciocca, D. R., *Cancer Detect. Prev.* 1997, **21**, 441-451.
- [29] Diehl, J. A., Yang, W., Rimerman, R. A., Xiao, H., Emili, A., *Mol. Cell Biol.* 2003, **23**, 1764-1774.
- [30] Leopold, J. A., Walker, J., Scribner, A. W., Voetsch, B., Zhang, Y. Y., Loscalzo, A. J., Stanton, R. C., Loscalzo, J., *J. Biol. Chem.* 2003, **278**, 32100-32106.
- [31] Halbrugge, M., Walter, U., *Eur. J. Biochem.* 1989, **185**, 41-50.
- [32] Salazar, R., Bell, S. E., Davis, G. E., *Exp. Cell. Res.* 1999, **249**, 22-32.
- [33] Dertsiz, L., Ozbilim, G., Kayisli, Y., Gokhan, G. A., Demircan, A., Kayisli, U. A., *Thorax* 2005, **60**, 576-581.
- [34] Fekete, M. R., McBride, W. H., Pajonk, F., *BMC Cancer* 2005, **5**, 114.
- [35] Rajkumar, S. V., Richardson, P. G., Hideshima, T., Anderson, K. C., *J. Clin. Oncol.* 2005, **23**, 630-639.
- [36] Adams, J., *Cancer Cell* 2004, **5**, 417-421.
- [37] Wakasugi, K., Slike, B. M., Hood, J., Ewalt, K. L., Cheresch, D. A., Schimmel, P., *J. Biol. Chem.* 2002, **277**, 20124-20126.
- [38] Tzima, E., Reader, J. S., Irani-Tehrani, M., Ewalt, K. L., Schwartz, M. A., *J. Biol. Chem.* 2005, **280**, 2405-2408.
- [39] Miyashita, H., Kanemura, M., Yamazaki, T., Abe, M., Sato, Y., *Arterioscler. Thromb. Vasc. Biol.* 2004, **24**, 878-884.
- [40] Takenaka, K., Ogawa, E., Oyanagi, H., Wada, H., Tanaka, F., *Cancer Epidemiol. Biomarkers Prev.* 2005, **14**, 1972-1975.
- [41] Fryer, B. H., Field, J., *Cancer Letters* 2005, **229**, 13-23.
- [42] Bouterfa, H. L., Piedrafita, F. J., Doenecke, D., Pfahl, M., *DNA Cell Biol.* 1995, **14**, 909-919.
- [43] Cecconi, D., Donadelli, M., Scarpa, A., Milli, A., Palmieri, M., Hamdan, M., Areces, L. B., Rappsilber, J., Righetti, P. G., *J. Proteome Res.* 2005, **4**, 1909-1916.

CHAPTER 6

PRINCIPAL COMPONENT ANALYSIS AND CLASSIFICATION METHODS: APPLICATIONS TO MURINE NEUROBLASTOMA SAMPLES

6.1 INTRODUCTION

Proteomic research [1-2] has become a key tool for the study of several pathologies: the physiological state of a particular cell or tissue is related to its protein content and the onset of a particular disease may cause differences in the proteins contained in the cell. The differences which may arise can consist of both changes in the relative abundance and in the appearance/disappearance of some proteins [3-11]. 2-D gel electrophoresis can thus be applied for both diagnostic and prognostic purposes by comparing 2D-PAGE maps belonging to healthy subjects with samples belonging to individuals affected by different pathologies [3-11]. However, the comparison of different 2D-PAGE maps is not a trivial aim to achieve: the difficulty that arises during the comparison is mainly due to the high complexity of the specimen, which can produce maps with thousands of spots; the complexity is also increased by the highly articulated sample pre-treatment, often characterised by many purification/extraction steps: these experimental steps may cause the appearance of spurious spots, due to impurities, in the final 2-D maps. Moreover, the differences occurring between healthy and pathological samples may be very small, thus complicating their identification in a real complex map. The differential analysis between classes of different 2D-PAGE maps is usually achieved by mean of specific software (*i.e.* Melanie III or PDQuest) [12-15]. Recently, Zhan and Desiderio [16] determined the spatial and quantitative reproducibility between two commercial second-dimensional gel electrophoresis systems.

Two different approaches are available for the comparison of classes of 2-D maps: comparing the proteomic patterns of the investigated samples or comparing the samples on the basis of their spot volume. Both the proposed approaches produce a huge amount of information which can be investigated by multivariate techniques. Our research group has developed a new method based on fuzzy logic and on the successive application of classification methods for the comparison of the proteomic pattern of classes of 2D-PAGE maps [17, 18]; the proposed method has also been applied with Multi Dimensional Scaling (MDS) for the study of 2-D maps of control and nicotine treated rat serum samples [19]. Different proteomics patterns have also been investigated by the use of Three-way Principal Component Analysis [20]. Recently, Principal Component Analysis (PCA) [21-24] has been applied to the study of spot volume datasets of DNA and RNA fragments of several biological

systems [25-28], and to the characterisation of proteomic patterns of different classes of tissues [29-35]. PCA has been recently adopted also for the characterisation of the regulatory activity of new drugs [36-39]. In the present study, PCA is applied to a dataset composed of 6 samples of healthy adrenal nude mouse gland and adrenal nude mouse gland mass, plus primary tumors, from which twenty 2-D maps were produced (10 from the control samples and 10 from the pathological ones). The procedure was repeated three times, thus producing 3 genuine replicates for each map. The subsequent application of classification methods, like SIMCA (Soft Independent Model of Class Analogy) [40], was useful for the identification of classes of samples in the dataset. The results thus obtained have been compared with those generated by analysis of the same set of twenty 2-D maps with the commercial software PDQuest. A number of polypeptide chains, found to be up- and down-regulated, have been eluted from the gel slabs and identified *via* MALDI-TOF mass spectrometry (see chapter 3).

6.2 THEORY

6.2.1 PRINCIPAL COMPONENT ANALYSIS

PCA [21-24] is a multivariate statistical method which allows the representation of the original dataset in a new reference system characterised by new variables, called Principal Components (PCs). Each PC has the property of explaining the maximum possible amount of variance contained in the original dataset. The PCs, which are expressed as linear combinations of the original variables, are orthogonal to each other and can be used for an effective representation of the system under investigation, with a lower number of variables than in the original case. The co-ordinates of the samples in the new reference system are called *scores* while the coefficient of the linear combination describing each PC, (*i.e.* the weights of the original variables on each PC), are called *loadings*.

6.2.1 CLASSIFICATION: SIMCA MODEL

The Soft-Independent Model of Class Analogy (SIMCA) [40, 41] is a classification method based on Principal Component Analysis. This method builds boxes containing the samples of each class, on the basis of a class-dependent PC model. Each object is then assigned to a class by calculating its distance from the different PCs model assessed on the samples of each class. Since this method is based on PCs and not on the original variables, the main advantage of its application is that the objects are assigned to each class by considering only the systematic and useful information contained in the dataset, overlooking the unsupportive

ones. Moreover the method is suitable also for investigating small datasets, like the ones objects of the present study. This depends on the possibility of performing a substantial size reduction of the descriptors through the calculation of the principal components. In the comparison of sets of gels from control and treated samples, SIMCA allows for the identification of modelling and discriminating spots, *i.e.* of those spots which are differentially expressed in the two sets of samples under comparison.

6.3 EXPERIMENTAL PROCEDURES

Six samples, divided in two classes, constituted the dataset:

- 6 samples belonging to healthy adrenal nude mouse gland (the pool of which was used to perform ten 2-D maps - figure 1(a));
- 6 samples belonging to adrenal nude mouse gland mass, plus metastasis at other key organs (the pool of which was used to obtain ten 2-D maps - figure 1(b)).

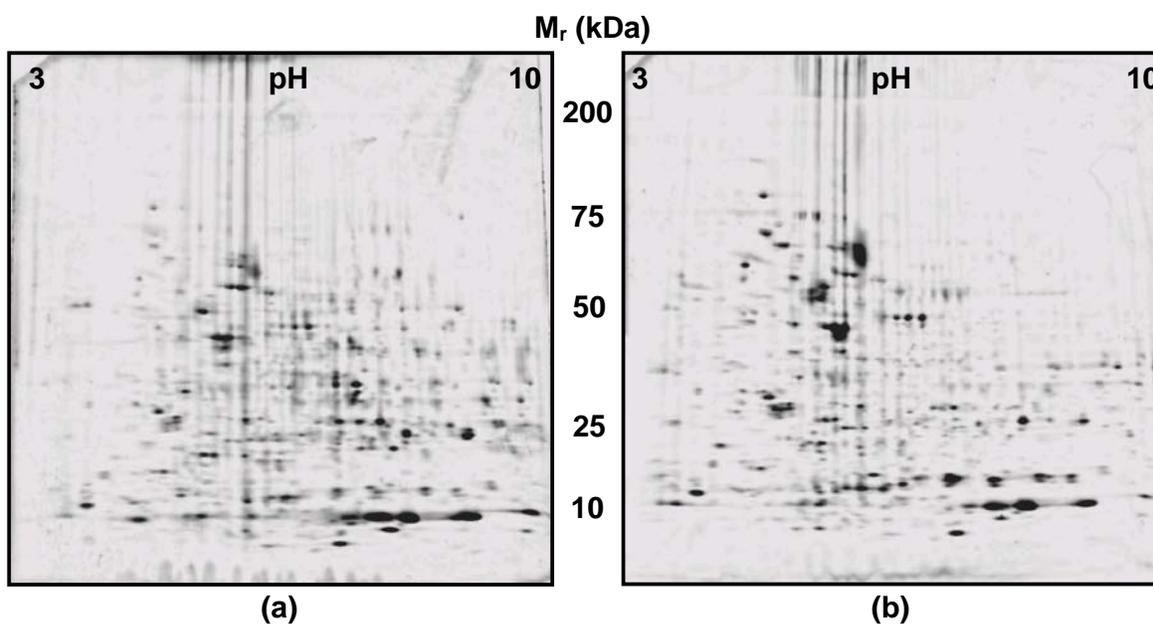


Figure 1. Examples of the obtained 2D-PAGE maps: control (a) and diseased (b) samples.

6.3.1 2D- PAGE

6.3.1.1 SAMPLE TREATMENT

Healthy and tumoral adrenal glands from nude mice (nu/nu) were stored at -80°C . All samples were homogenized (5% homogenate) with a lysis solution containing 7 M urea, 2 M thiourea, 3% CHAPS, 40 mM Tris, 5 mM TBP, 0.5% carrier ampholytes pH 3-10, 5 mM EDTA, 1 mM PMSF, 50 U/mL DNase. After 1 hour lysis, samples were centrifuged at 4°C at 6.000 rpm for eliminating all residual particles. The collected supernatant was alkylated for 1 hour with 10 mM acrylamide (1 M stock solution) and the reaction was blocked with 10 mM DTT (from 1 M stock solution). Interfering substances (lipids, salts) were removed by a precipitation step. A cold mix of acetone and methanol (8:1) was added; after 2 hours at -20°C , the solution was centrifuged at 13.000 rpm for 30 minutes. The pellets were finally resuspended in sample solution containing urea, thiourea, CHAPS and Tris.

6.3.1.2 ELECTROPHORETIC SEPARATION

All the necessary steps for producing high-resolution 2-D maps, data analysis and protein identification by mass spectrometry were carried out as described in section 3.2 of chapter 3.

6.3.1.3 SOFTWARE

PCA was performed with UNSCRAMBLER (Camo Inc., ver. 7.6), together with classification (SIMCA method). Graphical representations were performed with both UNSCRAMBLER and STATISTICA (Statsoft Inc., ver. 5.1). 2D-PAGE maps were scanned with a Versa Doc (Bio-Rad), and analyzed with the software PDQuest Version 6.2 (Bio-Rad, Laboratories, Hercules, CA, USA).

6.4 RESULTS AND DISCUSSION

6.4.1 PRINCIPAL COMPONENT ANALYSIS

The differential analysis performed by PDQuest on the 60 samples permitted the identification of 532 spots. The matching procedure produced a dataset constituted by 532 variables (the optical density of each matched spot) and 60 objects (the 60 samples) thus giving a data matrix 60 x 532. Since the 3 genuine replicates of each map proved to have a good reproducibility, their results were averaged, thus reducing the data matrix to 20 x 532, where the 20 samples are constituted by 10 averaged healthy samples and 10 averaged diseased ones. All variables were autoscaled before performing PCA, which gave the results reported in table 1.

	% Explained variance	% Cumulative explained variance
PC₁	47,3	47,3
PC₂	14,5	61,8
PC₃	10,6	72,4
PC₄	8,3	80,7

Table 1. Results of PCA (for the first four PCs) performed on the overall data set.

The first three PCs explain more than 72% of the total variance contained in the original dataset, but only the first PC was considered for the successive analysis, since it showed to be able to clearly separate the two classes of samples. Figure 2 represents the score plot of the first two principal components. The score plot shows that the first PC allows an effective separation of the samples in the two classes: the pathological samples (ILL1-ILL10) lay at large negative scores on PC1, while the control ones (HEA1-HEA10) are grouped at large positive values on the same PC.

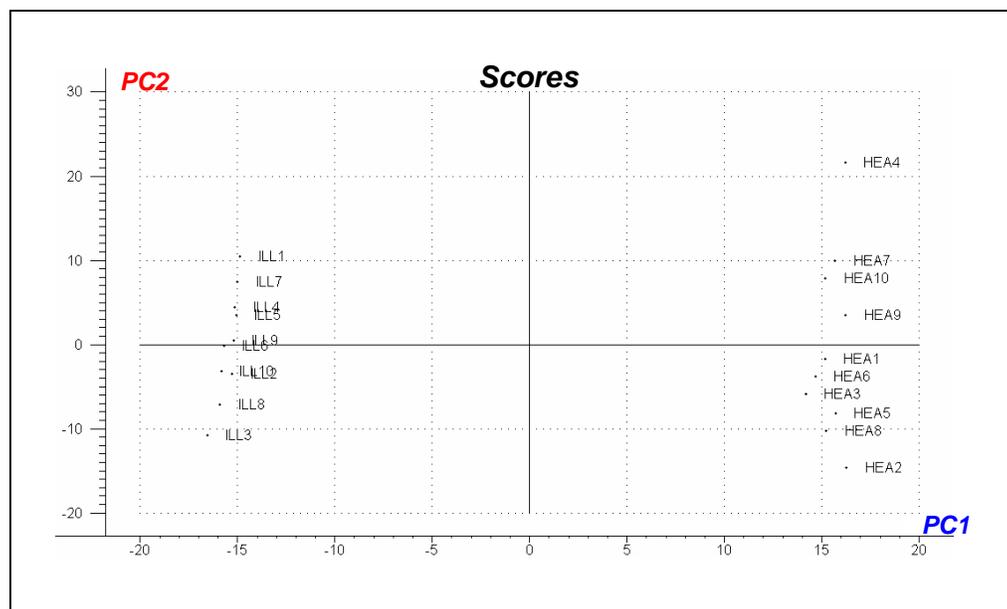


Figure 2. Score plot of PC₂ vs PC₁.

The loading plots of PC1 (figure 3) represents the spots as circles centred in the x-y position identified by mean of PDQuest analysis. The red coloured spots identify those spots showing a large positive loading on the first PC, while the blue coloured ones identify the spots showing large negative values on the same PC. The spots are represented in a colour scale ranging from light to dark red or blue, where the increasing tone is proportional to the corresponding loading. The spots showing a small influence are light red (small positive loading) or light blue (small negative loading) coloured, while those showing a large influence are dark red (large positive loading) or dark blue (large negative loadings) coloured. The spots marked as a black circle do not have a relevant loading on the first PC. The loading plots (figure 3) allow the identification of the spots which show an increase in the optical density with the onset of the disease (blue coloured spots, large negative loadings on PC1) or which are absent in the control samples. The red spots show instead the proteins which decrease or are absent after the appearance of the disease.

6.4.2 CLASSIFICATION

Since the PCA is able to identify two groups of samples in the data set, a classification method (SIMCA) was applied in order to verify the presence of the two classes and identify the modelling variables (the spots which describe each class) and the discriminating ones (the spots which are responsible for the differences between the two classes). SIMCA was selected since it is able to perform well, even when small data sets are available. PCA was then applied to the two classes (healthy and pathological samples) separately and a model based on principal components was calculated for each class. For both classes, the samples are exhaustively described by a model which includes only the first principal component; the results of PCA performed on the two separate data sets are reported in Table 2.

	Control class		Diseased class	
	% Explained variance	% Cumulative explained variance	% Explained variance	% Cumulative explained variance
PC₁	43,3	43,3	46,0	46,0
PC₂	29,5	72,8	31,7	77,7

Table 2. Results of PCA (for the first two PCs) performed on the control and diseased classes separately.

The SIMCA method, performed on the overall data set, by considering the two models just calculated, allowed the correct classification of the samples in control and pathological ones, with a subsequent non-error rate (NER%) of 100%. The SIMCA model, as just pointed out, allows the identification of both the modelling and the discriminating variables. The first ones are represented in Figure 4 for both healthy and pathological samples; the spots are represented in a colour scale, according to their modelling power (ranging from 0 to 1): spots from light to dark red show an increasing modelling power, while spots marked as a black circle do not play a relevant role in the definition of the SIMCA model built for each class.

The spots showing the largest modelling power for each class are those which describe the class more effectively; the comparison between the modelling spots of Figure 4 and the loadings plots of Figure 3 confirms the conclusions just driven by the performance of PCA.

The modelling spots for the healthy class show positive loadings on PC₁, while the modelling ones for the diseased class show negative loadings on the same PC.

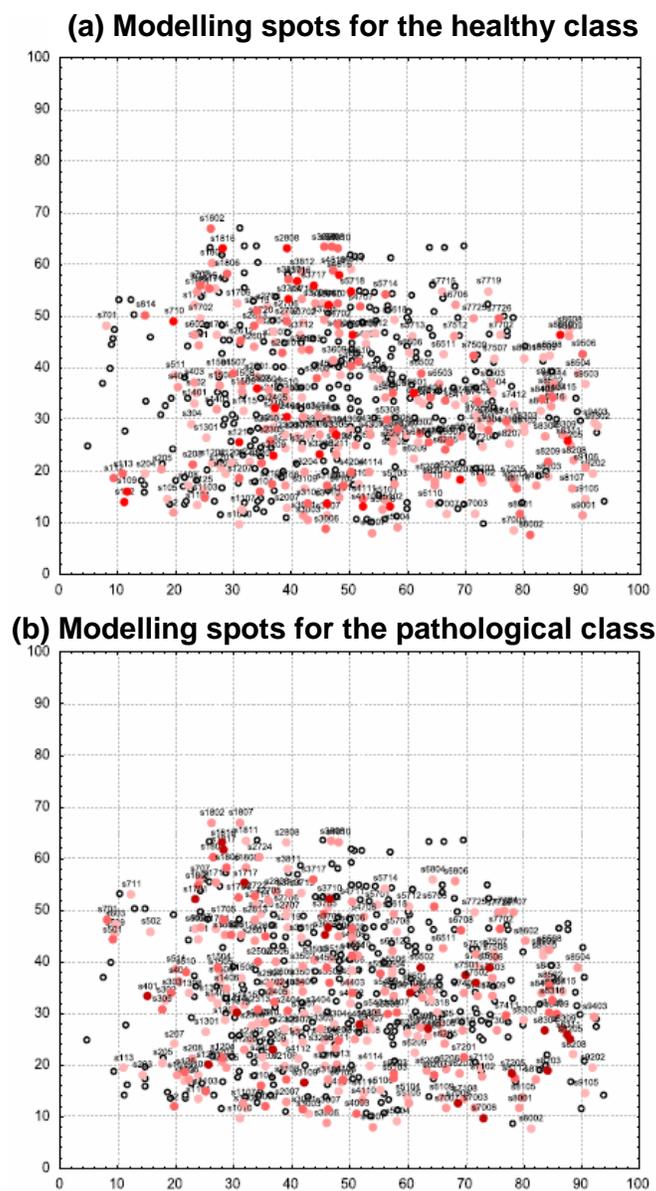


Figure 4. Modelling variables obtained by the calculation of the SIMCA model for the two classes of samples: modelling variables for the healthy (a) and pathological (b) classes. Spots are identified by their SSP value.

In order to check the validity of the results obtained by the PCA data treatment, the same set of gels was analysed *via* a commercially available software, namely the PDQuest, a software which has been continuously refined and improved over the years since its inception in 1979. The results of such data handling are shown in Figure 6, which displays, highlighted, a total of 84 spots differentially expressed in normal vs pathological tissues. A number of these spots were eluted from the gels, digested and analysed by MALDI-TOF mass spectrometry. Fourteen of these polipeptide chains could be identified and are labelled in Figure 6 by their respective SSP numbers. Two of these spots (No. 15 and 16) gave good spectra, but could not be found in any of the databases available. All the other unidentified spots were too faint and gave poor quality spectra.

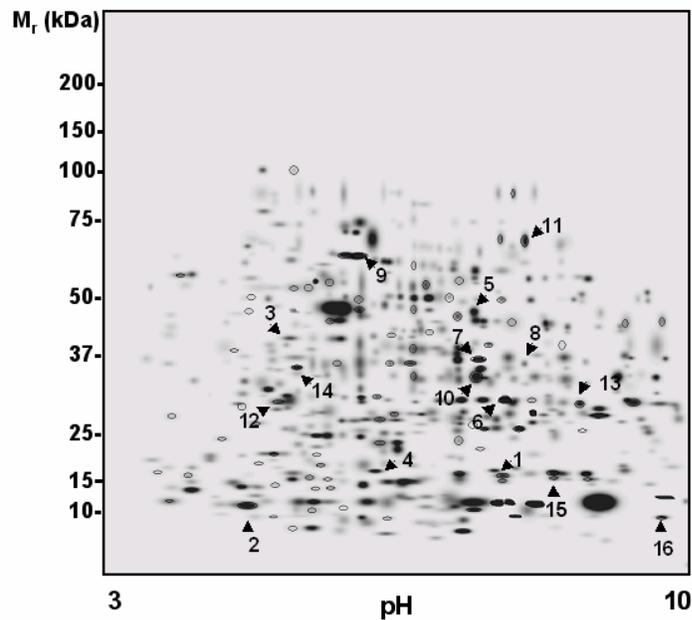


Figure 6. Master map of Neuroblastoma tissue excised from the adrenal glands of nude mice. The 84 spots differentially expressed are highlighted with circles and the identified spots are numbered.

6.5 CONCLUSIONS

PCA is applied here to a data set composed by 20 samples belonging to control and diseased nude mouse adrenal cells. The performance of PCA on the overall data set turned out to be an effective tool for the identification of both classes of samples by means of only the first PC. The analysis of the loadings of PC_1 allowed us to evaluate the spots characterising each class of samples; the separation of the samples in the two classes (healthy and pathological) was also confirmed *via* the calculation of the SIMCA classification model. Moreover, this classification method, by means of one PC for each class, allowed the identification of the modelling spots and the discriminating ones. This analysis was in good agreement with the conclusions just driven by PCA. The results of PCA and SIMCA analysis were additionally compared to those obtained *via* dedicated commercial softwares, such as PDQuest, and found to be in very good agreement. Fourteen up and down-regulated proteins, of relevance in tumor development, could be excised from the gels and identified *via* mass spectrometry and database searching. The use of 2-D gel-electrophoresis coupled to multivariate statistical tools, *e.g.*, PCA and classification methods, is certainly an interesting tool in the field of clinical chemical research, allowing the identification of the proteins responsible for the differences occurring between healthy and pathological samples.

6.6 BIBLIOGRAPHY

- [1] Righetti, P. G., Stoyanov, A., Zhukov, M., *The Proteome revisited: Theory and Practice of the Relevant Electrophoretic Steps*, Elsevier, Amsterdam, 2001.
- [2] Wilkins, M. R., Williams, K. L., Appel, R. D., Hochstrasser, D. F., *Proteome Research: New Frontiers in Functional Genomics*, Springer-Verlag, Berlin, Germany, 1997.
- [3] Fountoulakis, M., Schlaeger, E. J., *Electrophoresis* 2003, **24**, 260-275.
- [4] Kim, J., Kim, S. H., Lee, S. U., Ha, G. H., Kang, D. G., Ha, N. Y., Ahn, J. S., Cho, H. Y., Kang, S. J., Lee, Y. J., Hong, S. C., Ha, W. S., Bae, J. M., Lee, C. W., Kim, J. W., *Electrophoresis* 2002, **23**, 4142-4156.
- [5] Gromov, P. S., Ostergaard, M., Gromova, I., Celis, J. E., *Prog. Biophys. Mol. Bio.* 2002, **80**, 3-22.
- [6] Dwek, M. V., Rawlings, S. L., *Mol. Biotechnol.* 2002, **22**, 139-152.
- [7] Kostanjevecki, V., Lannoy, L., Daniels, A., *Neurobiol. Aging* 2002, **23**, 1383.
- [8] Castegna, A., Aksenov, M., Thongboonkerd, V., Klein, J. B., Pierce, W. M., Booze, R., Markesbery, W. R., Butterfield, D. A., *J. Neurochem.* 2002, **82**, 1524-1532.
- [9] Sickmann, A., Marcus, K., Schafer, H., Butt-Dorje, E., Lehr, S., Herkner, A., Suer, S., Bahr, I., Meyer, H. E., *Electrophoresis* 2001, **22**, 1669-1676.
- [10] Sinha, P., Kohl, S., Fischer, J., Hutter, G., Kern, M., Kottgen, E., Dietel, M., Lage, H., Schnolzer, M., Schadendorf, D., *Electrophoresis* 2000, **21**, 3048-3057.
- [11] Tsuji, T., Shiozaki, A., Kohno, R., Yoshizato, K., Shimohama, S., *Neurochem. Res.* 2002, **27**, 1245-1253.
- [12] Westergren-Thorsson, G., Malmstrom, J., Marko-Varga, G., *J. Pharmaceut. Biomed. Anal.* 2001, **24**, 815-824.
- [13] Bhatia, K., Lord, R., Stanton, P., *Eur. J. Cancer* 2002, **38**, S156.
- [14] Hoffmann, F., Kriegel, K., Wenk, C., *Discrete Appl. Math.* 1999, **93**, 75-88.
- [15] Lee, K. H., *Trends Biotechnol.* 2001, **19**, 217-222.
- [16] Zhan, X., Desiderio, D. M., *Electrophoresis* 2003, **24**, 1834-1846.
- [17] Marengo, E., Robotti, E., Gianotti, V., Righetti, P. G., *Ann. Chim. (Rome)* 2003, **93**, 105-116.
- [18] Marengo, E., Robotti, E., Righetti, P. G., Antonucci, F., *J. Chromatogr. A* 2003, **1004**, 13-28.
- [19] Marengo, E., Robotti, E., Gianotti, V., Righetti, P. G., Domenici, E., Cecconi, D., *Electrophoresis* 2003, **24**, 225-236.
- [20] Marengo, E., Leardi, R., Robotti, E., Righetti, P. G., Antonucci, F., Cecconi, D., *J. Proteome Research* 2003, **2**, 351-360.

- [21] Davis, J. C., *Statistics and Data Analysis in Geology*, John Wiley & Sons., New York, 1986.
- [22] Brereton, R. G., *Chemometrics: Application of Mathematics and Statistics to Laboratory Systems*, Ellis Norwood, New York, 1990.
- [23] Box, G. E. P., Hunter, W. G., Hunter, J. S., *Statistics for Experimenters*, Wiley, New York, 1978.
- [24] Massart, D. L., Vandeginste, B. G. M., Deming, S. N., Michette, Y., Kaufman, Y., *Chemometrics: a Textbook*, Elsevier, Amsterdam, 1988.
- [25] Leonard, R. B., Mayer, J., Sasser, M., Woods, M. L., Mooney, B. R., Brinton, B. G., Newcombgayman, P. L., Carroll, K.C., *J. Clin. Microbiol.* 1995, **33**, 2723-2727.
- [26] Johansson, M. L., Quednau, M., Ahrne, S., Molin, G., *Int. J. Syst. Bacteriol.* 1995, **45**, 670-675.
- [27] Couto, M. M. B., Vogels. J. T. W. E., Hofstra, H., Husiintveld, J. H. J., Vandervossen J. M. B. M., *J. Appl. Bacteriol.* 1995, **79**, 525-535.
- [28] Boon, N., De Windt, W., Verstraete, W., Top, E. M., *FEMS Microbiol. Ecol.* 2002, **39**, 101-112.
- [29] Alike, J. E., Akenova, M. E., Fatokun, C. A., *Genet. Res. Crop Evol.* 1995, **42**, 393-399.
- [30] Anderson, N. L., EsquerBlasco, R., Richardson, F., Foxworthy, P., Eacho, P., *Toxicol. Appl. Pharm.* 1996, **137**, 75-89.
- [31] Vohradsky, J., *Electrophoresis* 1997, **18**, 2749-2754.
- [32] Kovarova, H., Radzioch, D., Hajdich, M., Sirova, M., Blaha, V., Macela, A., Stulik, J., Hernychova, L., *Electrophoresis* 1998, **19**, 1325-1331.
- [33] Lacroix-Desmazes, S., Bayry, J., Misra, N., Horn, M. P., Villard, S., Pashov, A., Stieltjes, N., d'Oiron, R., Saint-Remy, J., Hoebeke, J., Kazatchkine, M. D., Kaveri, S. V., *New Engl. J. Med.* 2002, **34**, 662-667.
- [34] Dewettinck, K., Dierckx, S., Eichwalder, P., Huyghebaert, A., *Lait* 1997, **77**, 77-89.
- [35] Kovarova, H., Hajdich, M., Korinkova, G., Halada, P., Krupickova, S., Gouldsworthy, A., Zhelev, N., Strnad, M., *Electrophoresis* 2000, **21**, 3757-3764.
- [36] Ryan, T. E., Patterson, S. D., *Trends Biotechnol.* 2002, **20**, S45-S51.
- [37] Cacabelos, R., *Ann. Med.* 2002, **34**, 357-379.
- [38] Poirier, F., Pontet, M., Labas, V., le Caer, J. P., Sghiouar-Imam, N., Raphael, M., Caron, M., Joubert-Caron, R., *Electrophoresis* 2001, **22**, 1867-1877.
- [39] Steiner, S., Witzmann, F. A., *Electrophoresis* 2000, **21**, 2099-2104.
- [40] Wold, S., *Kem. Kemi.* 1982, **9**, 401.
- [41] Frank, I. E., Lanteri, S., *Chemometr. Intell. Lab.* 1989, **5**, 247.

CHAPTER 7

SPOT OVERLAPPING IN TWO-DIMENSIONAL MAPS

7.1 INTRODUCTION

The central theme of proteomics is the identification of proteins that are either involved in specific cellular process (cell map proteomics) or exhibit an altered expression profile as the result of some changed physiological condition (expression proteomics) [1]. Moreover, large-scale analysis of proteomes should also take into account post-translational modifications (PTMs), since altered protein profiles of tissues or cells are the results of altered protein modification rather than altered gene expression. 2D-PAGE separation is the method of choice in proteome studies. The final goal of proteomic projects is the identification of each protein, with particular interest to low abundance proteins: it requires a complete separation of all the proteins present in the sample, *i.e.* each spot occupying a specific position in the map contains one single component (SC) [2]. Although at the first glance the resolution of 2D-PAGE seems impressive, it is still not sufficient compared to the enormous diversity of cellular proteins: separation of all the proteins present in the sample is still far from being achieved and comigrating proteins in the same spot are not uncommon [3]. More worrying limitations of 2-D electrophoresis are linked to: (i) the impressive number of poly-peptide chains present in the sample, presumptively tens of thousands: an enormous resolution power is required for separating all these species, *i.e.* for displaying pure spots, formed by only one poly-peptide. (ii) The enormous chemical diversity of proteins. As a consequence, the range of pI and M_r of proteins is extremely wide and exceeds what can be routinely analysed on 2-D gels even in very simple organism [4]. (iii) The very divergent expression of proteins in cells or tissues resulting in a wide range of protein abundance. For example, in some biological fluids such as serum, albumin is present at 60 mg/mL and cytokines at pg/mL levels: this concentration interval is clearly out of the dynamic range of 2-D electrophoresis. A multicomponent mixture always gives rise to a complex map, which is characterized by a random crowd of spots allocated along the available separation space: some regions of the complex map are nearly empty and others are crowded with spots, determining spot overlapping (Figure 1a) [5]. This real separation is different from the ideal one, where spots of identical size are evenly allocated in the separation space, so that they are all resolved (as exemplified in Figure 1b).

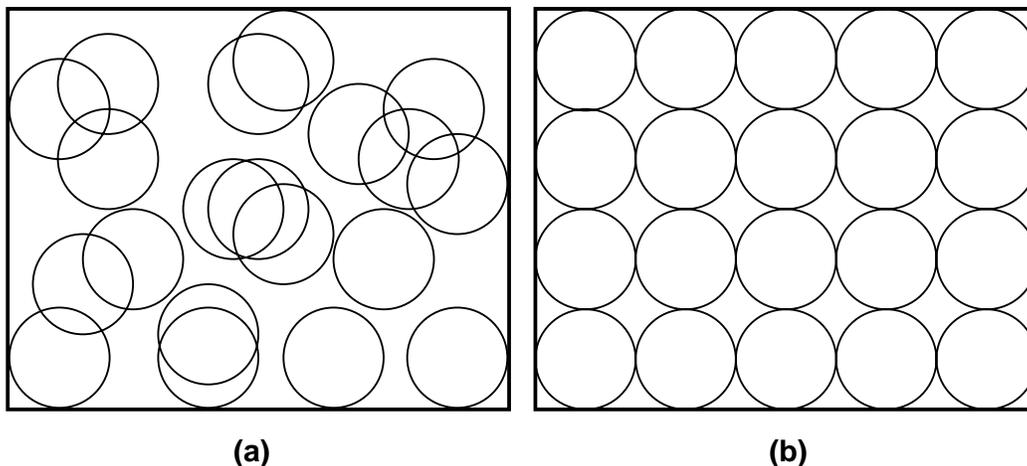


Figure 1. Examples of complex maps. (a) Representation of a real map: a random series of spots with casual distributions of positions resulting in spot overlapping; (b) ideal map: spots are evenly allocated in the separation space resulting in a perfect separation of all the component.

Many efforts have been spent for achieving the highest resolution in protein separation, especially for crowded areas in 2-D gels: narrow immobilized pH gradients have been used, *i.e.*, one strip 18 cm long covering 1 pH range [6] or the separation distance in the first dimension has been increased, *i.e.* from 8 cm to 11, 18, or even 39 cm [7]. In both these conditions many more proteins are detectable than in conventional maps and several apparently single spots are split into two or more peaks. For these reasons, it is fundamental to study overcrowding in protein maps in order to quantify the degree of overlapping present in the separation space.

Some consequences of spot overlapping are: (i) the number of spots, p , detectable in the map is always lower than the number m of components present in the mixture. The number of spots, p , does not represent the real complexity of the mixture: the loss of analytical information is represented by the separation degree $\gamma = p/m$ and is proportional to the degree of overlapping present in the mixture [5]; (ii) some spots are pure (singlets, s), *i.e.*, formed by 1 single component per spot; (iii) some spots derive by overlapping of 2 (doublets, d) or 3 (triplets, t) or more (n -test) single proteins.

Therefore, the total number of detectable spots, p , is given by:

$$p = s + d + t + \dots$$

While the total number of single components (SCs), m , is:

$$m = s + 2d + 3t + \dots$$

that means that for computing the real number of proteins present in a sample a quantitative estimation of the spot overlapping is required.

Many statistic theories have been developed for describing spot overlapping in 2-D separations [5, 8-13]: they all have been tested extensively by computer-simulated data as well as by experimental data. Up to recent times no applications of such statistic theories to 2D-PAGE separations of protein mixtures had been reported. A previous paper [11] demonstrated that the Quantitative Theory of Peak Overlapping, a procedure previously developed for 1-D chromatograms, could be successfully extended to 2-D separations. The whole map is divided into many strips in order to obtain 1-D separations on which the statistic procedure is applied. The 2-D map is regarded as a superimposition of several 1-D strips into which the map is divided. The procedure allows an accurate and precise estimation of the number of components, m . In the present study, the statistical theory of spot overlapping is applied to Neuroblastoma samples and it is here demonstrated that the experimental data on the abundance of overlapping spots are in excellent agreement with theoretical data calculated on the basis of this theory.

7.2 EXPERIMENTAL PROCEDURES

7.2.1 XENOGRAFT NEUROBLASTOMA ANIMAL MODEL

NB cells were injected in the adrenal glands of mice, as reported by Pastorino and coworkers [14]. Briefly, 5-week-old female SCID mice purchased from Harlan Laboratories (Udine, Italy) were anaesthetised and injected with 2×10^6 SH-SY5Y NB cells in 20 mL of HEPES buffer in the capsule of the left adrenal gland after laparotomy. Mice were monitored at least two times weekly for evidence of tumor development, quantification of tumor size and evidence of tumor-associated morbidity. Mice were sacrificed and organs were frozen and stored in liquid nitrogen after being washed in PBS. All experiments involving animals were reviewed and approved by the licensing and ethical committee of the National Cancer Research Institute and by the Italian Ministry of Health.

7.2.2 2D- PAGE, MS and DATA ANALYSIS

All the necessary steps for producing high-resolution 2D maps, MS and data analysis were carried out as described in section 4.2 of chapter 4.

7.2.3 APPLICATION OF THE SPOT OVERLAPPING THEORY

The statistical procedure “quantitative theory of peak overlapping” [12, 13] was applied to experimental data in order to statistically estimate the overlapping degree present in the 2D-PAGE separation. A map was constructed by using the p_i and M_i coordinates (p_i and M_i of the barycentre of each of the 74 spots detected). The third intensity coordinate is described by the abundance of each spot as retrieved from the PDQuest software. The 2D-PAGE map was divided into a proper number of strips (4) so that each strip contained at least 20 single components (SCs). On each 1-D separation, the statistical procedure was applied for estimating the number of proteins present in each strip and their sum yielded the total number in the map. Data concerning each strip are presented in Table 1.

p	m	α
21	28	0,26
18	21	0,20
19	31	0,33
16	23	0,19

Table 1. p , m , α values computed for each strip.

In Table 1 p is the total number of detectable spots (comprising singlets, s , doublets, d , triplets, t , etc.) and m the total number of SCs (thus, $m = s + 2d + 3t + \dots$). Their sum yielded the total number in the map: 103 components, whereas 74 spots were detected. This means that the degree of overlapping γ , *i.e.*, p/m (in %), present in the map was 71% (note that in an ideal map where a complete separation has been achieved, $p/m\%$ will be 100, since $p = m$, *i.e.*, all proteins will appear in the map as singlets). Starting from the estimated m values and using the statistical degree of overlapping model (SDO, originally developed by Davis [5, 8]), the degree of spot overlapping present in the 2D-PAGE map can be statistically estimated, *i.e.*, the number of spots formed by one, two, three or more proteins. Computations are based on the critical interdistance value, x_0 , between two adjacent, fully resolved spots along the pI axis. A value of $x_0 = 0.05$ pH was assumed to represent the experimental conditions. This value was calculated by knowing that the total IPG gel length was 18 cm, over a total pH span of 7 units, which would give a slope of 0.04 pH units/mm along the pI axis. By assuming an average spot to be an ideal circle of 1 mm diameter, and allowing a minimum of empty space in between, an $x_0 = 0.05$ pH was deemed to be a reasonable distance for baseline resolution. The basic equations for estimating the number of singlets, doublets and the number of detectable peaks as their sum are the following:

$$s = me^{-2\alpha}; d = me^{-2\alpha} (1 - e^{-\alpha}); \dots p = me^{-\alpha} \quad (1)$$

where α , the saturation factor, is given by

$$\alpha = mx_0/x \quad (2)$$

X being the total strip length. Specific α values were computed for each strip (see Table 1) and used for the computation.

7.3 RESULTS

By a lucky coincidence, since we had minute amounts of biopsies to be loaded onto the 2-D gel, the standard size map (18 x 20 cm) exhibited a total of 264 spots, instead of the usual 750 revealed when loading standard amounts. Such a level should be regarded as a minimal number, since we routinely stain our gels with colloidal Coomassie [15], a stain that is *ca.* five to ten times less sensitive than a silver staining protocol [16]. Of those, 85 spots, found to be up- or down-regulated in tumor tissue, were eluted and subjected to MS analysis by LC-MS/MS (see Figure 1).

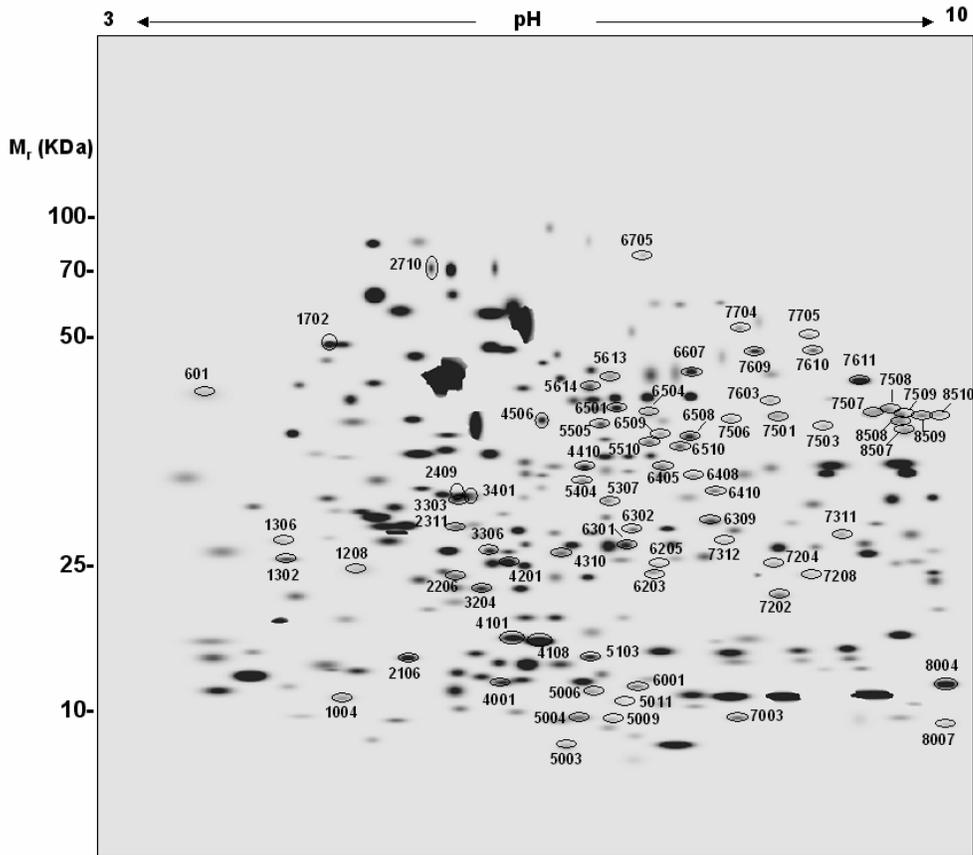


Figure 1. 2D map of the Neuroblastoma cells. The circles indicate the spots identified by MS analysis.

Among them, 74 could be identified, as listed in Table 2. Of these, 52 were found to be singlets (*i.e.*, spots comprising only a single polypeptide chain); 14 were doublets, 6 triplets, 1 quadruplet and 1 quintuplet.

Table 2. Total number of spots identified by MS and divided into singlets, doublets, triplets, quadruplets and quintuplets.

SINGLET (52 SPOTS)					
SSP number	2D Gel Liver Metastasis -Primary Tumor		pI, Mr (Da) (Swiss Prot)	Δ pI, Δ Mr	Sample
8004	P12710 FABL_MOUSE	Fatty acid-binding protein, liver	8,59 - 14246		Liver metastasis
8007	P31786 ACBP_MOUSE	Acyl-CoA-binding protein (ACBP)	8,78 - 9869		Liver metastasis
7003	P26772 CH10_RAT	10 kDa heat shock protein, mitochondrial (Hsp10)	8,91 - 10770		Liver metastasis
5009	P80254 DOPD_RAT	D-dopachrome tautomerase	6,15 - 13002		Liver metastasis
5011	P16436 HNT1_BOVIN	Histidine triad nucleotide-binding protein 1	6,31 - 13648		Liver metastasis
4410	Q9QXD6 F16P_MOUSE	(Q9QXD6) Fructose-1,6-bisphosphatase (EC 3.1.3.11)	6,18 - 36781		Liver metastasis
7204	P19157 GTP2_MOUSE	Glutathione S-transferase P 2	8,13 - 23478		Liver metastasis
7208	P35700 PDX1_MOUSE	Peroxisomal protein 1 (EC 1.11.1.-)	8,26 - 22176		Liver metastasis
6205	P24472 GTA4_MOUSE	Glutathione S-transferase 5.7 (EC 2.5.1.18)	6,77 - 25564		Liver metastasis
6203	P35700 PDX1_MOUSE	Peroxisomal protein 1 (EC 1.11.1.-)	8,26 - 22176		Liver metastasis
4310	O08709 PDX6_MOUSE	Peroxisomal protein 6 (EC 1.11.1.-)	5,72 - 24739		Liver metastasis
1208	O00264 PGC1_HUMAN	Membrane associated progesterone receptor component	4,56 - 21540		Liver metastasis
601	P16991 YB1_HUMAN	Nuclease sensitive element binding protein 1	9,87 - 35924		Liver metastasis
6309	P13255 GLMT_RAT	Glycine N-methyltransferase	7,21 - 32418		Liver metastasis
7603	P54869 HMCM_MOUSE	Hydroxymethylglutaryl-CoA synthase	8,02 - 53786		Liver metastasis
6408	P00480 OTC_HUMAN	Ornithine carbamoyltransferase, mitochondrial precursor	8,75 - 39901		Liver metastasis
5613	P47738 DHAM_MOUSE	Aldehyde dehydrogenase, mitochondrial precursor	7,53 - 56537		Liver metastasis
5006	P70349 HNT1_MOUSE	Histidine triad nucleotide-binding protein 1	6,31 - 13648		Liver metastasis
1306	P34022 RANG_MOUSE	Ran-specific GTPase-activating protein	5,15 - 23596		Liver metastasis
2206	Q9CPU0 LGUL_MOUSE	Lactoylglutathione lyase	5,25 - 20678		Liver metastasis
2409	Q64374 SM30_MOUSE	Senescence marker protein-30	5,16 - 33407		Liver metastasis
2311	Q64374 SM30_MOUSE	Senescence marker protein-30	5,16 - 33407		Liver metastasis
7610	P04762 CATA_RAT	Catalase (EC 1.11.1.6)	7,15 - 59626		Liver metastasis
6504	Q99J08 S142_MOUSE	SEC14-like protein 2	6,68 - 46300		Liver metastasis
7704	P30038 PUT2_HUMAN	Delta-1-pyrroline-5-carboxylate dehydrogenase	8,25 - 61751		Liver metastasis
6607	P10860 DHE3_RAT	Glutamate dehydrogenase, mitochondrial precursor	8,05 - 61428		Liver metastasis
6508	P35505 FAAA_MOUSE	Fumarylacetoacetase	6,92 - 46103		Liver metastasis
7503	P17764 THIL_RAT	Acetyl-CoA acetyltransferase, mitochondrial precursor	8,92 - 44695		Liver metastasis
5404	P14152 MDHC_MOUSE	Malate dehydrogenase, cytoplasmic (EC 1.1.1.37)	6,16 - 36346		Liver metastasis
5614	P47738 DHAM_MOUSE	Aldehyde dehydrogenase, mitochondrial precursor	7,53 - 56537		Liver metastasis
6410	P13255 GLMT_RAT	Glycine N-methyltransferase (EC 2.1.1.20)	7,21 - 32418		Liver metastasis
7312	P29410 KAD2_RAT	Adenylate kinase isoenzyme 2, mitochondrial	6,32 - 26248		Liver metastasis
7705	P30038 PUT2_HUMAN	Delta-1-pyrroline-5-carboxylate dehydrogenase	8,25 - 61751		Liver metastasis
7508	O09171 BHMT_RAT	Betaine-homocysteine S-methyltransferase	8,02 - 44976		Liver metastasis
7509	P13437 THIM_RAT	3-ketoacyl-CoA thiolase, mitochondrial	8,09 - 41871		Liver metastasis
6705	Q63342 M2GD_RAT	Dimethylglycine dehydrogenase, mitochondrial precursor	6,91 - 96047		Liver metastasis

7609	P24270 CATA_MOUSE Catalase (EC 1.11.1.6)	7,72 – 59626	Liver metastasis
4201	Q00623 Apolipoprotein A-I precursor	5,64 – 30587	Primary tumor
3306	P09936 UBL1_HUMAN Ubiquitin carboxyl-terminal hydrolase isozyme	5,33 – 24824	Primary tumor
7202	P30086 PEBP_HUMAN Phosphatidylethanolamine-binding protein (PEBP)	7,43 – 20925	Primary tumor
5003	P02248 UBIQ_HUMAN (P02248) Ubiquitin	6,56 – 8565	Primary tumor
5004	P31949 S111_HUMAN Calgizzarin (S100C protein)	6,56 – 11740	Primary tumor
1004	P46656 ADX_MOUSE Adrenodoxin, mitochondrial precursor	5,37 – 20123	Primary tumor
1302	P17080 RAN_HUMAN GTP-binding nuclear protein RAN (TC4)	7,01 – 23596	Primary tumor
4001	P02695 RET3_MOUSE Retinoic acid-binding protein I	5,31 – 15460	Primary tumor
6302	O70250 PMG2_MOUSE Phosphoglycerate mutase	8,65 – 28696	Primary tumor
1702	P27797 CRTC_HUMAN (P27797) Calreticulin precursor (CRP55) (Calregulin)	4,29 - 48111	Primary tumor
3401	Q15691 MAE1_HUMAN (Q15691) Microtubule-associated protein RP/EB family member	5,02 – 29999	Primary tumor
8507	P09411 PGK1_MOUSE Phosphoglycerate kinase 1	7,52 – 44405	Primary tumor
5505	Q15019 SEP2_HUMAN Septin 2 (NEDD5 protein homolog)	6,15 – 41487	Primary tumor
4101	P13668 STN1_RAT Stathmin (Phosphoprotein p19)	6 – 17117	Primary tumor
4108	P13668 STN1_RAT Stathmin (Phosphoprotein p19) (pp19)	6 – 17117	Primary tumor

DOUBLETS (14 SPOTS)

SSP number	2D Gel Liver Metastasis -Primary Tumor	pI, M _r (Da) (Swiss Prot)	Δ pI, Δ M _r	Sample
5103	P08228 SODC_MOUSE Superoxide dismutase [Cu-Zn] (EC 1.15.1.1)	6,03 – 15811		Liver metastasis
	P02278 H2BA_HUMAN (P02278) Histone H2B.a/g/k	10,32 – 13755	4,02 - 2056	
3303	Q64374 SM30_MOUSE Senescence marker protein-30 (SMP-30)	5,16 – 33407		Liver metastasis
	Q99426 TBCB_HUMAN (Q99426) Tubulin-specific chaperone B	5,06 – 27325	0,1 – 6082	
6405	Q61176 ARG1_MOUSE Arginase 1 (Liver-type arginase)	6,52 – 34808		Liver metastasis
	P00481 OTC_RAT Ornithine carbamoyltransferase, mitochondrial	9,12 – 39886	2,6 - 5078	
7506	Q92524 PRSX_HUMAN 26S protease regulatory subunit S10B	7,09 – 44173		Liver metastasis
	P16460 ASSY_MOUSE Argininosuccinate synthase	8,36 – 46584	1,27 – 2411	
7501	P16460 ASSY_MOUSE Argininosuccinate synthase	8,36 – 46584	0,27 – 4713	
	P13437 THIM_RAT 3-ketoacyl-CoA thiolase, mitochondrial	8,09 - 41871		Liver metastasis
7311	P14604 ECHM_RAT Enoyl-CoA hydratase, mitochondrial precursor	8,40 – 31516		Liver metastasis
	Q8R164 BPHL_MOUSE Valacyclovir hydrolase	9,06 – 32851	0,66 – 1335	
7611	Q03265 ATPA_MOUSE ATP synthase alpha chain, mitochondrial precursor	9,22 – 59752		Liver metastasis
	P24549 DHA1_MOUSE Aldehyde dehydrogenase 1A1	7,91 – 54336	1,31 – 5416	
6510	Q07417 ACDS_MOUSE Acyl-CoA dehydrogenase, short-chain	8,96 – 44946		Liver metastasis
	Q35423 SPYA_MOUSE Serine--pyruvate aminotransferase	8,57 – 45815	0,4 – 869	
8510	P13437 THIM_RAT 3-ketoacyl-CoA thiolase, mitochondrial	8,09 – 41871		Liver metastasis
	Q9DB77 UCR2_MOUSE Ubiquinol-cytochrome C reductase complex core	9,26 – 48234	1,17 – 6363	
8509	P16460 ASSY_MOUSE Argininosuccinate synthase	8,36 – 46584		Liver metastasis
	P00558 PGK1_HUMAN Phosphoglycerate kinase	8,3 – 44596	0,06 – 1988	
4506	P55262 ADK_CRIGR Adenosine kinase (EC 2.7.1.20)	6,12 – 40285		Liver metastasis
	P13444 METL_RAT S-adenosylmethionine synthetase alpha and beta	5,61 – 43698	0,51 – 3413	

6501	P42024 ACTZ_HUMAN	Alpha-centractin (Centractin)	6,19 – 42613		
	P26641 EF1G_HUMAN	Elongation factor 1-gamma (EF-1-gamma)	6,25 – 50118	0,06 – 7505	Primary tumor
2710	P08238 HS9B_HUMAN	Heat shock protein HSP 90-beta	4,97 – 83133	0,19 – 33033	
	P05218 TBB5_HUMAN	Tubulin beta-5	4,78 – 50100		Primary tumor
6001	P16436 HNT1_BOVIN	Histidine triad nucleotide-binding protein 1	6,31 – 13648		
	P02088 HBB1_MOUSE	Hemoglobin beta-1 chain (B1)	7,26 - 15709	0,95 - 2061	Primary tumor

TRIPLETS (6 SPOTS)

SSP number	2D Gel Liver Metastasis -Primary Tumor	pI, M _r (Da) (Swiss Prot)	Δ pI, Δ M _r	Sample
7507	P16460 ASSY_MOUSE	Argininosuccinate synthase	8,36 – 46584	
	P09411 PGK1_MOUSE	Phosphoglycerate kinase	7,52 – 44405	0,84 - 2179
	P00558 PGK1_HUMAN (P00558)	Phosphoglycerate kinase	8,30 – 44596	
2106	P00173 CYB5_RAT	Cytochrome b5	4,9 – 15224	
	P02278 H2BA_HUMAN	Histone H2B.a/g/k	10,32 – 13960	5,42 - 1264
	P02261 H2AC_HUMAN	Histone H2A.c/d/l/n/p	10,32 - 13960	
5510	Q64442 DHSO_MOUSE	Sorbitol dehydrogenase	6,60 – 40091	
	P32754 HPPD_HUMAN	4-hydroxyphenylpyruvate dioxygenase	6,50 – 44803	
	P26516 PSD7_MOUSE	26S proteasome non-ATPase regulatory subunit	6,29 – 36540	0,31 - 8263
6509	P15650 ACDL_RAT	Acyl-CoA Dehydrogenase	7,63 – 47873	
	P09606 GLNA_RAT	Glutamine synthetase	6,64 – 42268	0,99 – 5605
	P35505 FAAA_MOUSE	Fumarylacetoacetase	6,92 – 46103	
3204	O75947 ATPQ_HUMAN	ATP synthase D chain, mitochondrial	5,22 – 18360	
	Q2362 RS7_XENLA	40S ribosomal protein S7	10,09 – 22184	4,87 – 3824
	P32119 PDX2_HUMAN	Peroxisomal protein 2	5,66 – 21892	
6301	P60174 TPIS_HUMAN	Triosephosphate isomerase (EC 5.3.1.1) (TIM)	6,51 – 26538	
	P30041 PDX6_HUMAN	Peroxisomal protein 6	6,02 – 24904	
	P34062 PSA6_HUMAN	Proteasome subunit alpha type 6	6,35 - 27399	0,49 - 2495

QUADRUPLETS (1 SPOT)

SSP number	2D Gel Liver Metastasis -Primary Tumor	pI, M _r (Da) (Swiss Prot)	Δ pI, Δ M _r	Sample
5307	P46953 3HAO_RAT	3-hydroxyanthranilate 3,4-dioxygenase	5,57 – 32582	
	P82979 HCC1_HUMAN	Nuclear protein Hcc-1	6,10 – 23671	
	P97532 THTM_RAT	3-mercaptopyruvate sulfurtransferase	5,88 – 32809	0,53 - 9138
	P46952 3HAO_HUMAN (P46952)	3-hydroxyanthranilate 3,4-dioxygenase	5,62 - 32542	

QUINTUPLETS (1 SPOT)

SSP number	2D Gel Liver Metastasis -Primary Tumor	pI, M _r (Da) (Swiss Prot)	Δ pI, Δ M _r	Sample
8508	P17764 THIL_RAT	Acetyl-CoA acetyltransferase	8,92 – 44695	
	P24752 THIL_HUMAN		8,98 – 45199	
	P13437 THIM_RAT	3-ketoacyl-CoA thiolase	8,09 – 41871	0,96 - 5988
	P00883 ALFA_RABIT	Fructose-bisphosphate aldolase	8,40 – 39211	
	O09171 BHMT_RAT	Betaine-homocysteine S-methyltransferase	8,02 - 44976	

In Figure 2 a graphic representation of these data is shown. Table 3 gives the results derived from the SDO model [5, 8]. *i.e.*, the number of pure peaks, singlets, doublets, triplets, quadruplets and quintuplets.

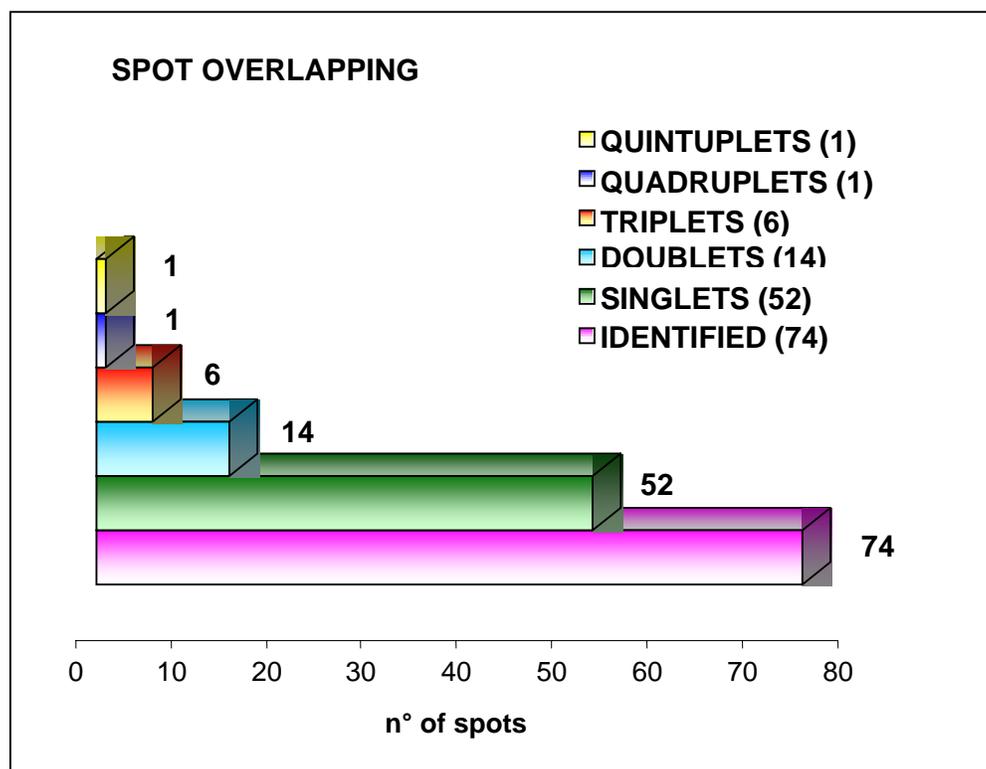


Figure 2. Graphic representation of the number of singlets, doublets, triplets, quadruplets and quintuplets found in the spots analysed by MS (the first bar indicates the total number of proteins identified).

What is comforting is the excellent agreement between the theoretical and experimentally observed peak population (singlets, doublets . . . ; cf. the two bottom lines “total map” and “experimental” in Table 3). It should be noted here that the results thus obtained strongly depend on: (i) the critical interdistance x_0 values, related to the pH resolution obtained in different experimental conditions; (ii) resolution in terms of ΔpI range, which depends on the experimental conditions of the first-dimension IEF separation, *i.e.*, the range of the IPGs adopted (in this case pH 3–10); (iii) scanner resolution, *i.e.*, standard (1 mm) or high (0.5 mm); (iv) sensitivity of the detection system; the ability of the MS analytical system to identify the presence of two or more components in one spot, *i.e.*, the contamination level of the major component in a spot. It is evident that any information on the purity degree of each spot depends on these experimental data.

	Counted spots (p)	Number of proteins (m)	Number of detectable peaks (p)	s : number of singlets	d : number of doublets	t : number of triplets	q : number of quadruplets	q : number of quintuplets
1 st strip	21	20	20,3	14,7	4,0	1,1	0,3	0,1
2 nd strip	18	21	15,7	11,8	2,9	0,7	0,2	0,0
3 rd strip	19	31	20,5	13,6	4,6	1,5	0,5	0,2
4 th strip	16	23	18,2	14,4	3,0	0,6	0,1	0,0
Total map	74	103	74,7	54,5	14,6	4,0	1,1	0,3
Experimental			74	52	14	6	1	1

Table 3. Results obtained from the Quantitative Theory of Peak Overlapping (m values) and SDO models (s, d, t, q, q values): comparison between estimated and experimental data. The number of pure peaks, singlets, doublets, triplets, quadruplets, quintuplets were computed for a critical interdistance $x_0 = 0.05$ pH along the pI axis.

7.4 DISCUSSION

7.4.1 THE THEORY OF SPOT OVERLAPPING AT WORK

This is the first instance in which our theory of spot overlapping has been applied to a real experimental 2D-PAGE gel, after its development and validation on computer-generated maps [5, 8]. The excellent agreement with data obtained from MS analysis of a real 2-D map proves that it is a powerful and robust method for accurately predicting the overlapping degree present in a map. It must be emphasized that the present data represent more of an exception than the rule, due to the very minute sample applied and the presence of only 264 total spots detectable. Thus, this is one of the few cases in which the number of singlets would be in the majority over the other spots (71%). In crowded conditions, as is typical of a tissue homogenate under normal loading conditions (*ca.* 1 mg total protein) and standard gel sizes (18 x 20 cm, IEF x SDS-PAGE), the singlets would be by far the least abundant species. The other surprising result is the fact that the only quintuplet found was in a pI region (pH *ca.* 9) less crowded with protein spots (statistical analysis has shown that most proteins crowd around the pH 4-6 region, *ca.* 60% of the total) [17].

7.4.2 HOW TO MINIMIZE SPOT OVERLAP

There are three major ways to increase the total number of spots seen, while diminishing spot overlapping: either increase the gel size, or exploit narrower and narrower pH gradients over the same separation distance, or both. The first avenue was explored long ago by Young [18, 19], who first adopted giant gels (37 cm in the first, 39 cm in the second dimension) out of frustration in his inability to detect protein changes in cells stimulated with adrenal steroid hormones. He had been using conventional size 2-D gels and had been able to resolve and detect only 700 to 900 spots. When increasing the gel area by a factor of 6, and loading concomitantly an amount of protein *ca.* 100-fold greater, he could resolve and detect a total of *ca.* 5,000 polypeptide chains and was thus able to monitor many hormone-induced changes in protein levels. Young [19] stated that “many of the major spots visible on the smaller gels, and especially the streaks, were resolved into multiple spots on the larger gels. Thus, spots presumed to represent single proteins on smaller gels often in fact represent several”. The other outstanding result came from the work of Klose and Zeindl [20], who reported the resolution and detection of at least 10,000 different polypeptides in an epithelial-like human larynx carcinoma cell line. Here too the data were obtained after substantial experimental effort: the cell line had been fully labelled with ¹⁴C-amino acids, the gels used were of giant size (30 cm in the first, 42 cm in the second dimension) and

detection had to be done sequentially by using multiple exposures for revealing the faintest spots. The other remedy to adopt is to run the so-called “zoom gels”, *i.e.*, a series of narrow-range IPG strips (covering no more than 1 pH unit) [21, 22]. The third path was explored by Oguri *et al.* [23], in what they call “cybergel” technology. Not only did they run “zoom gels” by using seven different narrow-range (1pH unit) IPG gels, covering the entire pH 3-10 range, over the standard 18-cm length of the pH axis, but they also greatly increased the length of the SDS-PAGE dimension by running gels of either 35 cm or even 55 cm length, instead of the standard 20 cm length. All the gel images were then “stitched” electronically into a 70 x 67 cm “cybergel”. The increment in spot detection (and resolution) was impressive: from a total of 853 spots in a standard-size gel up to 6677 spots, an increase of 783% in total spot count. However, even these remedies do not seem immune from some surprising results. A case in point is represented by a paper of Gygi *et al.* [3], who, in search of the “missing” low-abundance proteins (those with a codon bias value ,0.1) in a total yeast lysate, prepared a narrow IPG range (4.5-5.5) over a 25 cm gel length (resulting in a pH slope of 0.004/mm, 1 order of magnitude shallower than the pH gradients reported here!). By loading high sample levels (0.5 mg total protein), they could visualise no fewer than 1,500 spots, a truly remarkable resolution/detection capability. Yet, when excising a 4 cm² gel area and analysing 50 proteins therein, they found a spot to be composed of a “sextuplet”.

7.4.3 ON THE ΔpI and ΔM_r SPREAD IN OVERLAPPING SPOTS

Some other interesting conclusions can be drawn by a close inspection of Table 1, where in the subsections listing doublets, triplets, quadruplets and quintuplets, there are listed also the ΔpI and ΔM_r values of all proteins found in a “single spot”. Considering that when cutting out a gel segment for MS analysis, generally an ellipsoid of *ca.* 3 mm pH axis) x 2 mm (M_r dimension) is excised, one would expect to collect on average, in each single gel segment, a pI interval of 0.2 pH units maximum and 1,000 Da maximum, respectively. This is respected in only a few cases, *e.g.*, spots no. 8509, 3303, 7501, 2710 and 6501 among the doublets. In the vast majority of cases, the pI spread is as wide as 1 pH unit. What does this mean, considering the high precision of the IPG technology and the excellent correlation between theoretical pIs of proteins and their expected position along the IPG gradient, as demonstrated by Bjellqvist and coworkers [24]? In some cases it might simply mean that the theoretically predicted pI is wrong, due to some erroneous sequences stored in databases (as in fact elegantly found in [24]). But in the majority of cases it might simply mean that the proteins of diverging pI values collected in a single spot might represent post-translationally modified proteins, such as glycosylated, phosphorylated and deamidated proteins and any modification altering the surface charge. The other surprise comes when analysing the ΔM_r

data. Here too the interval very rarely encompasses only 1,000 Da; in most cases the M_r spread goes from 2,000 to 7,000 Da and even higher. In a few cases, one might consider that some of the spots with the largest ΔM_r values could represent cleaved products. But it cannot be excluded that in a number of other cases, this large spread might be due to an aberrant binding of SDS. Although Pitt-Rivers and Impiombato [25] demonstrated long ago that the classic SDS/protein ratio is 1.4:1 mg, which accounts for the fact that there is a very good relationship between migration behaviour and M_r , quite a few exceptions have been found, notably for histones [26], glycoproteins [27] and others [28].

7.4.4 ON THE SIGNIFICANCE OF APPARENTLY “ABERRANT” SPOTS

Some proteins, however, were found in spots that did not match the predicted M_r or pI by a divergence that cannot be explained by either of the ways discussed so far. Table 4 presents these proteins together with others identified in the respective spots.

Table 4. Spots that did not match the predicted M_r or pI

SSP number	pID	MASCOT score	peptide	M_r (kDa)	pI	comment
2710	P08238	869	25	83,4	4,97	
	P05218	161	6	50,1	4,78	Mr low
2106	P00173	571	15	15,2	4,9	
	P02278	105	3	13,7	10	pI high
	P02261	65	2	13,9	10,9	pI high
	gil16877384	304	9	44 (23,7)	8,8 (6,1)	(gi32129199)
3204	P32821	239	8	22,1	10	pI high
	O75947	270	7	18,4	5,2	
	P32119	152	4	22	5,6	
5103	P08228	319	5	15,9	6	
	P02278	55	2	13,1	10,3	pI high

As can be seen from the number of sequenced peptides and the resulting protein MASCOT score, all proteins were identified unambiguously. When investigating the M_r and pI , however, it is obvious that tubulin has too low an M_r for spot 2710, while histone 2A and 2B have too high a pI for spot 2106. The same is true for the ribosomal protein S7 in spot 3204 and histone H2B in spot 5103. The best explanation may come from the fact that they only have odd behaviour in one dimension, either M_r or pI , and that they are very abundant proteins: tubulin, ribosomal protein S7, and histones H2A and H2B. While we prepared our gels with utmost care, we still had remnants of streaking visible in Figure 1. It is therefore conceivable that the sensitivity was so high that we started picking up background in the 2-D gel. In line with this, we found histone H2A in two spots, 5103 and 2106, which migrated at nearly identical M_r . We encountered one other error when using NCBI, a richer database than Swiss-Prot, but also less annotated. In spot 2106 we found an additional protein gi16877384 that has a high M_r and pI for spot 2106. When reinvestigating our data we realised that for gi16877384 all peptides match the C-terminal region of this protein. We then actually found another database entry, gi32129199, to which all peptides match and that covers only the C-terminal part of gi16877384. This entry fits with pI 6.1, in good agreement with the observed value. It has a slightly lower mass than expected from the position of the spot. This may indicate that yet another isoform of the protein is present that has not yet been deposited in the database. Thus, paradoxically, careful examination of the types of proteins found in overlapping spots can yield some information on the validity of spot detection by MS (in our case, the error seems to be of the order of only 1-2%). By the same token, however, the identity of at least one protein found in the sextuplet of Gygi and coworkers [3] must also be discussed: protein VMA1 (M_r 67.7 kDa) cannot simply be found in an envelope of spots all having M_r values between 48 and 50 kDa.

7.4.5 PROTEIN ANALYSIS BY MASS SPECTROMETRY

Spots excised from the gel were analysed by LC-MS to get a better coverage of potential mixtures than is possible by MALDI-TOF PMF. LC allows for separation of peptides and therefore simplifies spectra and furthermore extends analysis time. From the peptides that had overlapping elution from the column, peaks were selected for fragmentation by intensity, *i.e.*, the most intense first. Nevertheless, this did not bias the analysis towards more abundant factors at the expense of less abundant proteins above the sensitivity threshold of this technology. During LC-MS analysis there was ample time to sequence a lot more peptides than actually was done. This means that the analysis was not limited by time, *i.e.*, exhaustive with respect to how many proteins could be identified with this technology. Using better equipment or improved analysis procedures would allow improvement of the detection limits and hence would likely result in more proteins to be identified. If consequently thought through, one easily realises that singlets do not exist. In fact, even “empty” spaces on the gel are likely to contain many proteins, just below the detection limit of the staining procedure. Today, MS supersedes the sensitivity of staining procedures such as Coomassie or silver staining. However, radioactive labelling allows proteins to be spotted even when nothing can be detected by MS.

7.5 BIBLIOGRAPHY

- [1] Righetti, P. G., Stoyanov, A., Zhukov, M. Y., *The Proteome Revisited: Theory and Practice of All Relevant Electrophoretic Steps*, Elsevier, Amsterdam, 2001, pp. 275-378.
- [2] Righetti, P. G., Castagna, A., Herbert, B., *Anal. Chem.* 2001, **73**, 230A-326A.
- [3] Gygi, S. P., Corthals, G. L., Zhang, Y., Rochon, Y., Aebersold, R., *Proc. Natl. Acad. Sci. U.S.A.* 2000, **97**, 9390-9395.
- [4] Corthals, G. L., Wasinger, V. C., Hochstrasser, D. F., Sanchez, J. C., *Electrophoresis* 2000, **21**, 1144-1115.
- [5] Davis, J. M., *Anal. Chem.* 1991, **63**, 2141-2152.
- [6] Zuo, X., Speicher, D. W., *Proteomics* 2002, **2**, 58-68.
- [7] Voris, B. P., Young, D. A., *Anal. Biochem.* 1980, **104**, 478-484.
- [8] Davis, J. M., *Anal. Chem.* 1993, **65**, 2014-2023.
- [9] Felinger, A., *Data Analysis and Signal Processing in Chromatography*, Elsevier, Amsterdam 1998, p. 366.
- [10] Dondi, F., Bassi, A., Cavazzini, A., Pietrogrande, M. C., *Anal. Chem.*, 1998, **70**, 766-773.
- [11] Pietrogrande, M. C., Cavazzini, A., Dondi, F., *Rev. Anal. Chem.* 2000, **19**, 123-155.
- [12] Pietrogrande, M. C., Marchetti, N., Dondi, F., Righetti, P. G., *Electrophoresis* 2002, **23**, 283-291.
- [13] Pietrogrande, M. C., Marchetti, N., Dondi, F., Righetti, P. G., *Electrophoresis* 2003, **24**, 217-224.
- [14] Pastorino, F., Brignole, C., Marimpietri, D., Cilli, M., Gambini, C., Ribatti, D., Longhi, R., Allen, T. M., Corti, A., Ponzoni, M., *Cancer Res.* 2003, **63**, 7400-7409.
- [15] Neuhoff, V., Arold, N., Taube, D., Ehrhardt, W., *Electrophoresis* 1988, **9**, 255-262.
- [16] Merril, C. R., Washart, K. M., *Gel Electrophoresis of Proteins: A Practical Approach*, Oxford University Press, Oxford, 1998, pp. 53-91.
- [17] Gianazza, E., Righetti, P. G., *J. Chromatogr.* 1980, **193**, 1-8.
- [18] Voris, B. P., Young, D. A., *Anal. Biochem.* 1980, **104**, 478-484.
- [19] Young, D. A., *Clin. Chem.* 1984, **30**, 2104-2108.
- [20] Klose, J., Zeindl, E., *Clin. Chem.* 1984, **30**, 2014-2020.
- [21] Hoving, S., Voshol, H., van Oostrum, J., *Electrophoresis* 2000, **21**, 2617-2621.
- [22] Westbrook, J. A., Yan, J. X., Wait, R., Welson, S. Y., Dunn, M. J., *Electrophoresis* 2001, **22**, 2865-2871.
- [23] Oguri, T., Takahata, I., Katsuta, K., Nomura, E., *Proteomics* 2002, **2**, 666-672.
- [24] Bjellqvist, B., Hughes, G., Pasquali, C., Paquet, N., Ravier, F., Sanchez, J. C., Frutiger, S., Hochstrasser, D., *Electrophoresis* 1993, **14**, 1023-1031.
- [25] Pitt-Rivers, R., Impiombato, F. S. A., *Biochem. J.* 1968, **109**, 825-833.

- [26] Panyim, S., Chalkley, R., *Arch. Biochem. Biophys.* 1969, **130**, 337-346.
- [27] Segrest, J. P., Jackson, R. L., Andrews, E. P., Marchesi, V. T., *Biochem. Biophys. Res. Commun.* 1971, **44**, 390-395.
- [28] Nelson, C. A., *J. Biol. Chem.* 1971, **246**, 3895-3901.

CHAPTER 8

GENERAL CONCLUSIONS

8.1 CONCLUSIONS

Cancer remains a major public health challenge despite progress in detection and therapy. From biotechnology to chemistry, from applied physics to software, increasing resources are being brought to bear on the goals of prevention and reducing mortality. Innovations and applications of biotechnology have allowed the exploitation of biological process in an effort to study pathogenesis at the molecular level. Novel technologies that are designed to advance the molecular analysis of healthy and diseased human cells are poised to revolutionize the field of health and disease. Advances in the field of genomics and proteomics are hoped to provide insights into the molecular complexity of the disease process and thus enable the development of tools to help in treatment as well as in detection and prevention. Proteomic technologies allow for identification of the protein changes caused by the disease process in a relatively accurate manner. The inherent advantage afforded to proteomics is that the identified protein is itself the biological endpoint. At the protein level, distinct changes occur during the transformation of a healthy cell into a neoplastic cell, including altered expression, differential protein modification, changes in specific activity, and aberrant localization, all of which may affect cellular function. Identifying and understanding these changes is the underlying theme in cancer proteomics [1].

In the present thesis work the differential proteomic approach based on 2-D gels coupled with mass spectrometry was applied for investigating Neuroblastoma growth and progression in two different animal models. The complexity of the interaction between NB tumors and their regional microenvironment emphasizes the importance of an *in vivo* model for such investigation. We started our study with a murine NB cell line orthotopically grafted into immunocompromised mice (nude mice) [2]. Mouse adrenal glands and adrenal gland carrying NB primary tumor were subjected to a comparative 2-D analysis and 84 spots were found to be differentially expressed. Among the MS identified proteins we could find some really interesting ones, which are, for example, related to apoptosis processes and tumorigenesis [3]. The information provided with this study could be very valuable in the evaluation of proteins associated with tumor development.

We were also interested in the comprehension of metastatic mechanisms and another animal model was chosen. A human NB cell line was orthotopically xenografted into SCID mice [4]. This model best reflects the typical growth pattern of human NB. In order to

evaluate NB progression and the mechanism underlying the metastatic process and to elucidate possible therapeutic targets, as well as putative molecular targets regulating metastasis, we performed proteomic analysis on primary adrenal tumors vs liver metastasis. We were able to identify a lot of proteins involved in the process of metastasis. Among them there were some of particular interest: at least three potential therapeutic targets have been identified, namely the 26S proteasome regulatory chain [5], the Eukaryotic translation initiation factor 4E [6], and the Y box binding protein-1 [7]. Moreover, at least two putative molecular targets able to control the metastatic process have been also identified in the Septin-2 [8] and the Raf kinase inhibitor protein [9].

There is no doubt that biomarker discovery in cancer is one of the most objectives of current research activities. However, it is reasonable to hypothesize that another major task is to predict response patterns to existing therapies, which will in turn help to design more efficient and less damaging therapeutics. With this idea in mind, we started a differentially proteomic analysis on an endothelial cell line untreated or treated with two chemotherapeutics (Vinblastine and Rapamycin). We were able to demonstrate that a combined treatment with VBL and RAP exhibited a synergistic effect in inhibiting the *in vitro* growth of endothelial cells by inducing apoptosis. By using two-dimensional tools and MS analysis, we identified a set of proteins that may be involved in the cooperation of VBL and RAP in inhibiting, among other effects, angiogenesis, which represents the primary cause of aggressive tumor growth. These findings might have important applications in future cancer treatment to design molecular-guided experimental therapies, especially in cases that are resistant to all currently used treatments. We think that, with the present thesis work, we improved the knowledge on proteins with important roles in relation to tumor development and metastasis as well as the anti-angiogenic effect of two drugs.

Recent technological advances in proteomics could revolutionize the way of doing research in medicine by providing researchers with formidable high-throughput laboratory tools with which to study protein expression profiles and functions in health and disease. Although in its infancy, oncoproteomics holds the promise of dissecting the biology of cancer, thereby improving cancer management from early detection to the development of disease specific drugs and patient-tailored therapeutic strategies. This unique opportunity to analyze the dynamics of the proteome on a genome-wide scale is based on different biochemistry techniques such as MS and protein arrays. The integration of proteomic and genomic data sets through powerful bioinformatics will yield a comprehensive database of protein properties that will serve as an invaluable tool for researchers to build and to test scientific hypotheses aimed at dissecting the process of tumorigenesis and, ultimately, at unveiling the Achilles heel of tumors [10].

While proteomics technologies certainly serve as useful tools for clinical investigation, it is important to remember that proper study design is the key to any kind of success derived from using these tools. It is possible that systemic errors can be introduced in the study that will artifactually discriminate disease from non-disease including differences in sample collection (*e.g.*, arterial vs venous blood; plasma vs serum) or using nonmatched patient groups (*e.g.*, everyone in the disease group is over 60 years and everyone in the control group is under 40 years). Possible diurnal variation of protein expression must be accounted for, so the time of sample collection should be controlled as closely as possible. In addition, it is important that patients with benign conditions, or with disease not related to the clinical condition of interest be included in the control group so that the study is truly examining differences between those with a specific disease condition and those without, rather than measuring differences between generally sick and generally healthy patients. These types of details can be difficult to manage, particularly in a retrospective study, but careful attention to study design will yield a greater chance of success for proteomic comparisons.

A bigger challenge will be to integrate proteomics with all the information from the other ‘-omics’, in particular genomics and metabolomics. This will allow us to obtain the holistic molecular view of pathogenetic processes, which enables their early recognition and the best selection of therapy tailored to the individual patient.

8.2 BIBLIOGRAPHY

- [1] Srinivas, P. R., Verma, M., Zhao, Y., Srivastava, S., *Clin. Chem.* 2002, **48**, 1160-1169.
- [2] Pastorino, F., Brignole, C., Marimpietri, D., Sapra, P., Moase, E. H., Allen, T. M., Ponzoni, M., *Cancer Res.* 2003, **63**, 86-92.
- [3] Camprostrini, N., Pascali, J., Hamdan, M., Astner, H., Marimpietri, D., Ponzoni, M., Righetti, P. G., *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 2004, **808**, 279-286.
- [4] Pastorino, F., Brignole, C., Marimpietri, D., Cilli, M., Gambini, C., Ribatti, D., Longhi, R., Allen, T. M., Corti, A., Ponzoni, M., *Cancer Res.* 2003, **63**, 7400-7409.
- [5] Rajkumar, S. V., Richardson, P. G., Hideshima, T., Anderson, K. C., *J. Clin. Oncol.* 2005, **23**, 630-639.
- [6] Mamane, Y., Petroulakis, E., Rong, L., Yoshida, K., Ler, L. W., Sonenberg, N., *Oncogene* 2004, **23**, 3172-3179.
- [7] Huang, J., Tan, P. H., Li, K. B., Matsumoto, K., Tsujimoto, M., Bay, B. H., *Int. J. Oncol.* 2005, **26**, 607-613.
- [8] Russell, S. E., Hall, P. A., *Br. J. Cancer.* 2005, **93**, 499-503.
- [9] Keller, E. T., Fu, Z., Yeung, K., Brennan, M., *Cancer Lett.* 2004, **207**, 131-137.
- [10] Mocellin, S., Rossi, C. R., Traldi, P., Nitti, D., Lise, M., *Trends Mol. Med.* 2004, **10**, 24-32.

PAPERS PUBLISHED DURING THE PhD STUDIES

The Proteome: *Anno Domini* 2002

Pier Giorgio Righetti^{1*}, Annalisa Castagna¹, Francesca Antonucci¹, Chiara Piubelli¹, Daniela Cecconi¹, Natascia Campostrini¹, Gianluigi Zanusso² and Salvatore Monaco²

¹ Department of Agricultural and Industrial Biotechnologies, University of Verona, Verona, Italy

² Section of Clinical Neurology, Department of Neurological and Visual Sciences, University of Verona, Verona, Italy

We present some current definitions related to functional and structural proteomics and the human proteome, and we review the following aspects of proteome analysis: Classical 2-D map analysis (isoelectric focusing (IEF) followed by SDS-PAGE); Quantitative proteomics (isotope-coded affinity tag (ICAT), fluorescent stains) and their use in e.g., tumor analysis and identification of new target proteins for drug development; Electrophoretic pre-fractionation (how to see the hidden proteome!); Multidimensional separations, such as: (a) coupled size-exclusion and reverse-phase (RP)-HPLC; (b) coupled ion-exchange and RP-HPLC; (c) coupled RP-HPLC and RP-HPLC at 25/60 °C; (d) coupled RP-HPLC and capillary electrophoresis (CE); (e) metal affinity chromatography coupled with CE; Protein chips.

Some general conclusions are drawn on proteome analysis and we end this review by trying to decode the glass ball of the aruspex and answer the question: “*Quo vadis, proteome*”? Clin Chem Lab Med 2003; 41(4):425–438

Key words: Proteome; Two-dimensional maps; Protein chips; Mass spectrometry; Multidimensional chromatography.

Abbreviations: 2-D, two-dimensional; CA, carrier ampholyte; CBI, codon bias index; CE, capillary electrophoresis; Cy3, 1-(5-carboxypentyl)-1'-propylindocarbocyanine halide N-hydroxysuccinimidyl ester; Cy5, 1-(5-carboxypentyl)-1'-methylindodicarbocyanine halide N-hydroxysuccinimidyl ester; DIGE, differential in-gel electrophoresis; DTE, dithioerythritol; DTT, dithiothreitol; HPLC, high-performance liquid chromatography; ICAT, isotope-coded affinity tag; IEF, isoelectric focusing; IPG, immobilized pH gradients; MALDI-TOF, matrix-assisted laser desorption ionization time of flight; MCAT, mass-coded abundance tagging; MCE, multicompartiment electrolyzers; MS, mass spectrometry; RP, reverse-phase; SCX, strong cation exchanger; SEC, size exclusion chromatography; SELDI, surface-enhanced laser desorption/ionization; TBP, tributyl phosphine.

“There are quite a few altar-fires flickering in the temple of proteome complexity”

Henry James, Washington Square, 1880, Bantam Books, NY, 1989:116.

Introduction

To be honest, James never prophesied and never quite entered the “temple of proteome complexity”, as quoted above: if he had done so, he would have been smarter than Tiresias, the blind soothsayer of Thebes, who in turn could not predict his assassination by the order of Creon. Nevertheless, it is quite sure that today there are more than a few altar-fires alight in the proteome field: in fact the entire arena is ablaze with fires, with diggers from all corners of science excavating deeper and deeper in search of the mother lode. This is today's scenario in science: nobody does it for science's sake, but most people do it for the sake of the green Dollar, the polychromatic Euro, the Yen, you name it.

Figure 1 gives an explanation for this furious digging: proteins are drug targets, so not only the pharmaceutical industry, but just about any research fellow in the field looks for proteins with key activities, whose modulation could be affected by active site-directed drugs. Among those, up to 52% could be cell membrane receptors and this gives an idea how complex are the investigations in the field: in a total cell lysate, the receptors would be among the least abundant species; thus, their detection requires special skills and techniques, as described below. The damnation of proteome research is that we keep seeing only the tip of

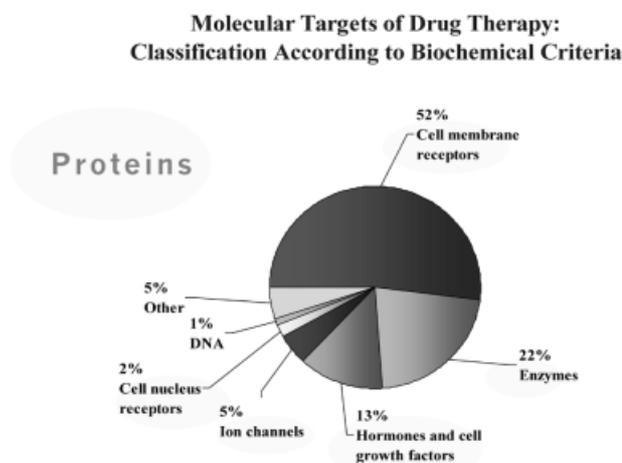


Figure 1 The distribution of important proteins acting as drug targets. Approximately 52% of them are cell membrane receptors, the second most abundant category being enzymes.

*E-mail of the corresponding author: righetti@sci.univr.it

Table 1 The sloppy vernacular of biology, or how too many "omics" could lead us to "comics".

GOD ALMIGHTY, LET US HAVE OUR DAILY "OMICS":

Genomics: all the genes you could have find out

Transcriptomics: all the messengers on the loose (perhaps a 100000 strong)

Proteomics: all the king's men and all the king's horses ... (an army: perhaps 1000000!)

Metabolomics: the few survivors of the wrath of all previous macromolecules (a handful, perhaps only 2000)

Systemomics: a cocktail of all the above!

Sed libera nos a ...

Comics!

the iceberg, the vast majority of precious and rare proteins being still hidden from sight. This can be easily appreciated in the vast body of papers published weekly, if not daily: when the spots are analyzed and identified by mass spectrometry, it is quite clear that the various laboratories are "re-discovering", over and over again, the very same set of proteins, *i.e.*, the most abundant ones, which are in general the house-keeping and structural proteins. The rare (and precious) ones are missing the roll-call.

It is not quite possible here to thoroughly review methodologies and the enormous literature in the field. We will discuss only some recent hot topics. The smart reader can find now quite a bit of information in the new journals: *Proteomics* (Wiley), which started publishing in the year 2000, *Journal of Proteome Research* (American Chemical Society) and *Molecular and Cellular Proteomics* (American Biochemical Society), both funded in 2001. Of course, plenty of papers will still appear in *Electrophoresis* and *Analytical Chemistry*, the two leading journals in separation science. Among recent books dedicated to this topic, one could recommend those by Wilkins *et al.* (1), Kellner *et al.* (2), Rabilouid (3), James (4), and Righetti *et al.* (5). Among book chapters, Westermeier (6) and Hanash (7), to name just a few.

Some Definitions

Before entering the proteome arena, some definitions are necessary, given the fact the today's science jargon is becoming more and more unintelligible. We will start with those that have surreptitiously penetrated the fortress of modern science, as depicted in Table 1: were it not enough to have to digest such odd terms as genomics, transcriptomics, proteomics, and metabolomics, we have to gobble down such an aberrant term as "systemomics". Fear not though, a whole host of them is appearing on the horizon: there are rumours of "glycomics", "complexomics" and the like. And, just to increase the entropy in the Babel tower, here come more hair-splitting definitions: Jeremy Nicholson (personal communication) insists on a distinction between "metaboLomics" (cellular metabolism) and "metabo-

Table 2 The concept of functional proteomics.

The description of changes in protein expression during differentiation, proliferation and signalling of cells, both in qualitative and quantitative terms, falls under the field of functional proteomics. It includes also studies of co-ordinated expression of genes, as well as elucidation of the sequence of regulatory events during all stages which a cell or an organism undergoes during its entire life span (8).

Table 3 The definition of structural proteomics.

The aim of structural proteomics is to identify the molecular structure, *i.e.* the amino acid sequence of the protein entities involved in a given process and to relate this information to the database of identified genes. The most powerful method, one that has revolutionized proteomics, is mass spectrometry analysis (4).

Table 4 A guess estimate of the proteome in humans.

Proteome of a cell: 5000 polypeptides;

Proteome of an individual as a snap-shot: 10^6 polypeptides;

Proteome of an individual during the entire life span: 10^7 polypeptides;

Proteome of a species: 10^8 polypeptides.

Based on an estimate of 45000 genes in the human genome (9)

Table 5 Some questions awaiting answers in modern genomic/proteomic analysis.

We do not need to know everything about everything, but we might want to be able to answer to at least some of the following questions:

1) How many and which genes are activated in a cell?

2) How many polypeptides, and in what amounts, are synthesized?

3) What post-translational modifications occur on a synthesized protein?

4) What is the half-life of protein molecules and how does this affect their function?

5) What are the rules for moving some proteins in and out of any cellular compartment?

Nomics" (intact systems metabolism). They might all be right, but who will save us from "comics"? Some definitions can be found in Tables 2 (8) and 3 (4), which refer to such clear terms as functional and structural proteomics. Table 4 offers a glimpse of the tremendous complexity of the proteome: assuming a total of 45000 genes, the human race could display as many as 10^8 polypeptides, quite a large number, indeed (9). In Table 5 we pose some intriguing, as yet unanswered, questions. Whereas some of them might be too far-fetched, questions 2, 3, and 4 definitely fall within the domain of proteome research.

Table 6 A (partial) list of methods available today for tackling the proteome complexity.

- Classical 2-D map analysis (IEF followed by SDS-PAGE);
 - Chromatographic and electrophoretic pre-fractionation;
 - Multidimensional separations, such as:
 - (a) coupled size-exclusion and RP-HPLC;
 - (b) coupled ion-exchange and RP-HPLC;
 - (c) coupled RP-HPLC and RP-HPLC at 25/60 °C;
 - (d) coupled RP-HPLC and capillary electrophoresis (CE);
 - (e) metal affinity chromatography coupled with CE.
- Protein chips

What Methods Exist to Tackle the Proteome Complexity?

Given the outstanding complexity of the proteome field, as illustrated in Table 4, it would appear that, to pursue such a problem, we need a panoply of tools, and these are listed in Table 6.

It is evident that, in addition to the classical electrophoretic approach, *i.e.*, the two-dimensional (2-D) map analysis, today we have a number of competing chromatographic techniques, trying to replace the good old O'Farrell approach (10), incriminated on the grounds of its complexity and labor-intensive manipulations. Curiously, chromatographers do not seem to be contented with just a 2-D approach, but they boast a "multidimensional" procedures, although all the papers we screened appear to offer, at best, the 2-D approach. To our reckoning, a multidimensional scheme was only put together once, in Jorgenson's laboratory. It was, in fact, a three-dimensional protocol, size-exclusion chromatography, coupled with reverse-phase (RP) liquid chromatography, followed by high-speed capillary zone electrophoresis (11). This was not successful and led these authors to conclude that "the increased peak capacity of this system may not be worth the extra effort and added complexity that is entailed".

Standard 2-D Map Analysis

Although the power of 2-D electrophoresis as a biochemical separation technique had been well recognized since its introduction, its application, nevertheless, became particularly significant in the past few years, as a result of a number of developments:

- The 2-D technique has been tremendously improved to generate 2-D maps that are superior in terms of resolution and reproducibility. This new technique utilizes a unique first-dimension, that replaces the carrier ampholyte-generated pH gradients with immobilized pH gradients (IPG) and replaces the tube gels with gel strips supported by a plastic film backing (12).
- Methods for rapid analysis of proteins have been improved so that single spots eluted or transferred from single 2-D gels can be rapidly identified. Mass spectroscopic techniques have been developed that

allow analysis of very small quantities of proteins and peptides (13–15). Chemical microsequencing and amino acid analysis can be performed on increasingly smaller samples (16). Immunochemical identification is now possible with a wide assortment of available antibodies.

- More powerful, less expensive computers and software are now available, allowing routine computerized evaluation of highly complex 2-D patterns.
- Data on entire genomes (or their substantial fractions) for a number of organisms are now available, allowing rapid identification of genes encoding a protein separated by 2-D electrophoresis.
- The World Wide Web (WWW) provides simple, direct access to spot pattern databases for the comparison of electrophoretic results, and to genome sequence databases for sequence information.

In 2-D PAGE, one of the critical steps is the initial sample solubilization. For decades, the most popular lysis solution has been the O'Farrell cocktail (9 M urea, 2% Nonidet P-40, 2% β -mercaptoethanol, and 2% carrier ampholytes, at any desired pH interval). Although much in vogue also in present times, over the years, new, even more powerful, solubilizing mixtures were devised. Great efforts were put into such developments, especially because many authors noted that hydrophobic proteins were largely absent from 2-D maps (17). They observed that, quite strikingly, in three different analyzed species (*Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*), all proteins above a given hydrophobicity value were completely missing, independently from the isoelectric focusing (IEF) mode (soluble carrier ampholytes (CAs) or IPG). This suggested that the initial sample solubilization was the primary cause for loss of such hydrophobic proteins. The development of solubilizing cocktails can be summarized as follows (see also reviews by Molloy (18) and Rabilloud and Chevallet (19)):

- *Chaotropes*. Although for decades urea (up to 9.5 M) has been the only chaotrope used in IEF, recently thiourea has been found to further improve solubilization, especially of membrane proteins (20). The inclusion of thiourea is recommended for use with IPGs, which are prone to adsorptive losses of hydrophobic and isoelectrically neutral proteins. Typically, thiourea is added at concentrations of 2 M in conjunction with 5–7 M urea. The high concentration of urea is essential for solvating thiourea, which is poorly water soluble. Among all substituted ureas (alkyl ureas, both symmetric and asymmetric) Rabilloud *et al.* (21) found thiourea to be still the best additive. However, it seems that thiourea at >2 M concentrations inhibits binding of SDS in the equilibration step between the 1st and 2nd dimension, thus leading to poor transfer of proteins into the 2-D gel. Therefore, it would appear that not much higher amounts of thiourea can be added to the IPG gel strip. It should also be remembered that urea in water exists in equilibrium with ammonium cyanate, whose level increases with increasing pH and temperature. Since cyanate can react with amino groups

in proteins, such as the N-terminus α -amino or the ϵ -amino groups of lysine, these reactions should be avoided since they will produce a false sample heterogeneity and give wrong M_r values upon peptide analysis by matrix-assisted laser desorption ionization time of flight (MALDI-TOF) mass spectrometry (MS). Thus, fresh solutions of pure grade urea should be used, concomitant with low temperatures and the use of scavengers of cyanate (such as the primary amines of carrier ampholytes or other suitable amines). In addition, protein mixtures solubilized in high-concentration urea should be subjected to separation in the electric field as soon as possible: in the presence of the high voltage gradients typical of the IEF protocol, cyanate ions are quickly removed and no carbamylation can possibly take place (22), whereas it will if the protein/urea solution is left standing on the bench.

- **Surfactants.** These molecules are always included in solubilizing cocktails to act synergistically with chaotropes. Surfactants are important in preventing hydrophobic interactions due to exposure of protein hydrophobic domains induced by chaotropes. Both the hydrophobic tails and the polar head groups of detergents play an important role in protein solubilization. The surfactant tail binds to hydrophobic residues, allowing dispersal of these domains in an aqueous medium, while the polar head groups of detergents can disrupt ionic and hydrogen bonds, aiding in dispersion. Detergents typically used in the past included Nonidet P-40 or Triton X-100, in concentrations ranging from 0.5 to 4%. More and more, zwitterionic surfactants, such as CHAPS, are replacing those neutral detergents (23), often in combination with low levels (0.5%) of Triton X-100. In addition, small amounts of CAs (<1%) are added, since they appear to reduce protein-matrix hydrophobic interactions and overcome detrimental effects caused by salt boundaries (24). Linear sulfobetaines are now emerging as perhaps the most powerful surfactants, especially those with at least a 12-carbon tail (SB 3–12). The inclusion of an amido group along the hydrophobic tail was found to improve urea tolerance up to 8.5 M, leading to the synthesis of ASB 14, an amidosulfobetaine containing a 14-C linear alkyl tail (25). This reagent has since been used successfully in combination with urea and thiourea to solubilize integral membrane proteins of both *E. coli* (26) and *Arabidopsis thaliana* (27).
- **Reducing agents.** Thiol agents are typically used to break intramolecular and intermolecular disulfide bridges. Cyclic reducing agents, such as dithiothreitol (DTT) or dithioerythritol (DTE) are the reagents most commonly admixed to solubilizing cocktails. These chemicals are used in large excess (*e.g.*, 20 to 40 mM) so as to shift the equilibrium toward oxidation of the reducing agent with concomitant reduction of protein disulfides. Because this is an equilibrium reaction, loss of the reducing agent through migration of proteins away from the sample application zone can permit reoxidation of free cysteine to

disulfides in proteins, which would result not only in horizontal streaking but also, possibly, in formation of spurious extra bands due to scrambled -S-S-bridges and their cross-linking of different polypeptide chains. Even if sample is directly reswollen in the dried IPG strip, as it is customary today, the excess DTT or DTE will not remain in the gel at a constant concentration, since, due to their weakly acidic character, both compounds will migrate above pH 7 and be depleted from the alkaline gel region. Thus, this will aggravate focusing of alkaline proteins and be one of the multifactorial factors responsible for poor focusing in the alkaline pH. According to recent data of Bordini *et al.* (28), when probing the alkylation by acrylamide of -SH groups in proteins by MALDI-TOF, it was found that the primary site of attack, even in proteins having both disulfide bridges and free -SH groups, was not the free -SH residues, as it should, but it was systematically one of the -SH residues engaged in the disulfide bridges. This can only be explained by assuming that, at alkaline pH (the incubation was carried out at approximate pH 10), disulfide bridges are weakened and are probably constantly snapped broken and re-formed. The situation would be aggravated using β -mercaptoethanol, since it has an even lower pK value, thus it is more depleted in the alkaline region and will form a concentration gradient towards pH 7, with a distribution in the gel following its degree of ionization at any given pH value along the IEF strip (29). This is probably the reason for the dramatic loss of pH gradient above pH 7.5, lamented by most users of conventional IEF in CAs, when generating 2-D maps. The most modern solution to all the above problems appears to be the use of phosphines as alternative reducing agents. Phosphines operate in a stoichiometric reaction, thus allowing the use of low concentrations (barely 2 mM). The use of tributyl phosphine (TBP) was recently proposed by Herbert *et al.* (30), who reported much improved protein solubility for both Chinese hamster ovary cell lysates and intractable, highly disulfide cross-linked, wool keratins. TBP thus offers two main advantages: it can be used at much reduced levels as compared to other thiol reagents (at a concentration which is at least one order of magnitude lower) and, additionally, it can distribute uniformly in the IPG gel strip (when rehydrated with the entire sample solution) since, being uncharged, it will not migrate in the electric field. A major drawback of TBP, though, is that it is volatile, toxic, and rather flammable in concentrated stock. No matter which strategy is used, it is imperative that all samples, prior to the IEF/IPG step, are reduced and alkylated; failure to implement this will result in a large number of spurious spots in the final 2-D map, due to mixed disulfide bridges among like and unlike polypeptide chains (31–33).

Quantitative Proteomics

When analyzing normal or pathological tissues undergoing changes in protein expression profiles (*e.g.*, during growth, differentiation, drug exposure, development of tumors) it is important to be able to quantify such changes, to detect polypeptides undergoing up-regulation or down-regulation. Such proteins might be key targets for drug action, thus their detection would be important for medical treatment and pharmaceutical industry. One of the earliest ways to achieve that has been the proper use of computer programs (*e.g.*, Melanie, PD Quest) able to compare standard maps generated from, *e.g.*, normal and pathological tissues. The minimum statistical requirement for such standard maps would be to collect three to five samples in each state, combine them, and generate, simultaneously, a minimum of five 2-D maps side by side under the same experimental conditions to minimize errors. Sophisticated computer programs would then combine the five different maps and generate one reference and one "pathological" map, which could then be overlapped with confidence. The thus obtained synthetic comparative map would detect proteins being upregulated or downregulated through variations in uptake of the Coomassie Blue stain. An example of such "differential maps" is given in Figure 2A and B: these maps refer to a peculiar category of tumors, called mantle cell lymphomas (non-Hodgkins).

Figure 2A displays nine spots exhibiting increments of spot density from 500 to 1000%, whereas Figure 2B discloses a set of seven spots with stain density decrements of 500 to 1300% (34). Is this system reliable? When the small spot marked 1010 in Figure 2A was eluted, digested with trypsin and analyzed by MALDI-TOF MS, it was found to be the "T cell leukemia/lymphoma protein 1A", heavily upregulated in lymphomas (in our case by one order of magnitude). Also all the other spots identified were found to be proteins either upregulated or downregulated in a number of tumors (34). It would thus appear that this method is truly reliable, although extremely lengthy and labor-intensive. Thus, in recent years, other approaches have emerged, based on the principle of differential isotope labeling of certain residues in proteins. The first one to appear, termed isotope-coded affinity tags (ICATs), is used for modifying the -SH groups of cysteine residues (35, 36). The structure of this chemical is shown in Figure 3: it is composed of three regions, a terminal iodine tail for adding onto -SH groups; an intermediate or linker region, coded either with d_0 or d_8 (d = deuterium; thus imparting a difference of 8 Da to the two differentially labeled protein populations), and an affinity-capture extremity, containing a biotin bait.

Its use is illustrated in Figure 4: the two samples to be compared are labeled either with the light or heavy reagent, mixed in a 1:1 ratio and digested with trypsin. This results in an extremely heterogeneous mixture of peptides, of the order of a few hundred thousands in a total cell lysate, and such mixture is not amenable to direct MS analysis. In order to reduce the complexity of

the sample, the peptide mixture is subjected to affinity purification on an avidin column. Only the peptide population containing cysteine residues (ca. 8–9% of the total) will be captured and analyzed by MS. Each peptide will be divided into two peaks, spaced apart by 8 Da, representing the light and heavy label, respectively. The ratio of these two peaks will give their ratio in the original sample mixture, and thus offer an immediate insight into upregulated or downregulated polypeptides. Since the ICAT chemical is very expensive, recently Sechi (37) and Gehanne *et al.* (38) have proposed an inexpensive set of chemicals, namely d_0/d_3 acrylamide, which add with the same efficiency to -SH groups and perform the same task as ICAT (with the added benefit that they can be used in standard 2-D maps, *i.e.*, in the analysis of intact proteins, not of their peptide digests). Although the mass difference between the two isotope-coded, cysteine-bearing peptides is only 3 Da, this does not pose any problem with modern mass spectrometers, whose mass accuracy allows for baseline resolution of adjacent peaks spaced apart even by this minute mass difference.

Figure 5 outlines a typical strategic scheme for their use. This brings about an intriguing question: how are proteins identified by MS techniques? It should be noted here that MS has become such an important tool in proteome analysis that today it would be impossible to properly fund a laboratory devoted to 2-D mapping without this equipment. MS analysis, for most practical purposes, has eliminated lengthy and cumbersome identification procedures based on micro-sequencing (Edman degradation). How protein identification is achieved *via* MS protocols is illustrated in Figure 6: a protein eluted from a 2-D map is digested with trypsin and the entire lysate subjected, typically, to time of flight (TOF) analysis. The mass spectrometer fully resolves such a mixture, no matter how complex, and gives precise mass values. Interrogating databases with this set of values only might not give an unequivocal answer, though: it would be highly desirable to select a single peptide (*e.g.*, one showing a strong peak) and subject it to fragmentation *via*, *e.g.*, post-source decay or MS/MS. Computer algorithms, *via* assessment of the size of each fragment, are able to reconstruct the sequence of up to 20 residues (most often a 10–12 amino acid sequence will univocally identify a protein). With these data (namely, the total number of fragments generated, their precise mass values, and a lead sequence) one should be able to kill the Minotaur, like Theseus, and exit unscathed from the labyrinth.

There are other drawbacks in the use of the ICAT protocol, reported by Zhang and Regnier (39), which apply to capture techniques that are different from those reported by the inventors of the ICAT, based on avidin-biotin affinity. When dealing with such complex mixtures arising from a tryptic digest of a total cell lysate, one might want to use different chromatographic protocols. For example, isotopically labeled peptides could be separated by ion-exchange chromatography followed by a RP column, or by RP chromatography followed by ion-mobility separation. It turns out that, if ICAT-labeled

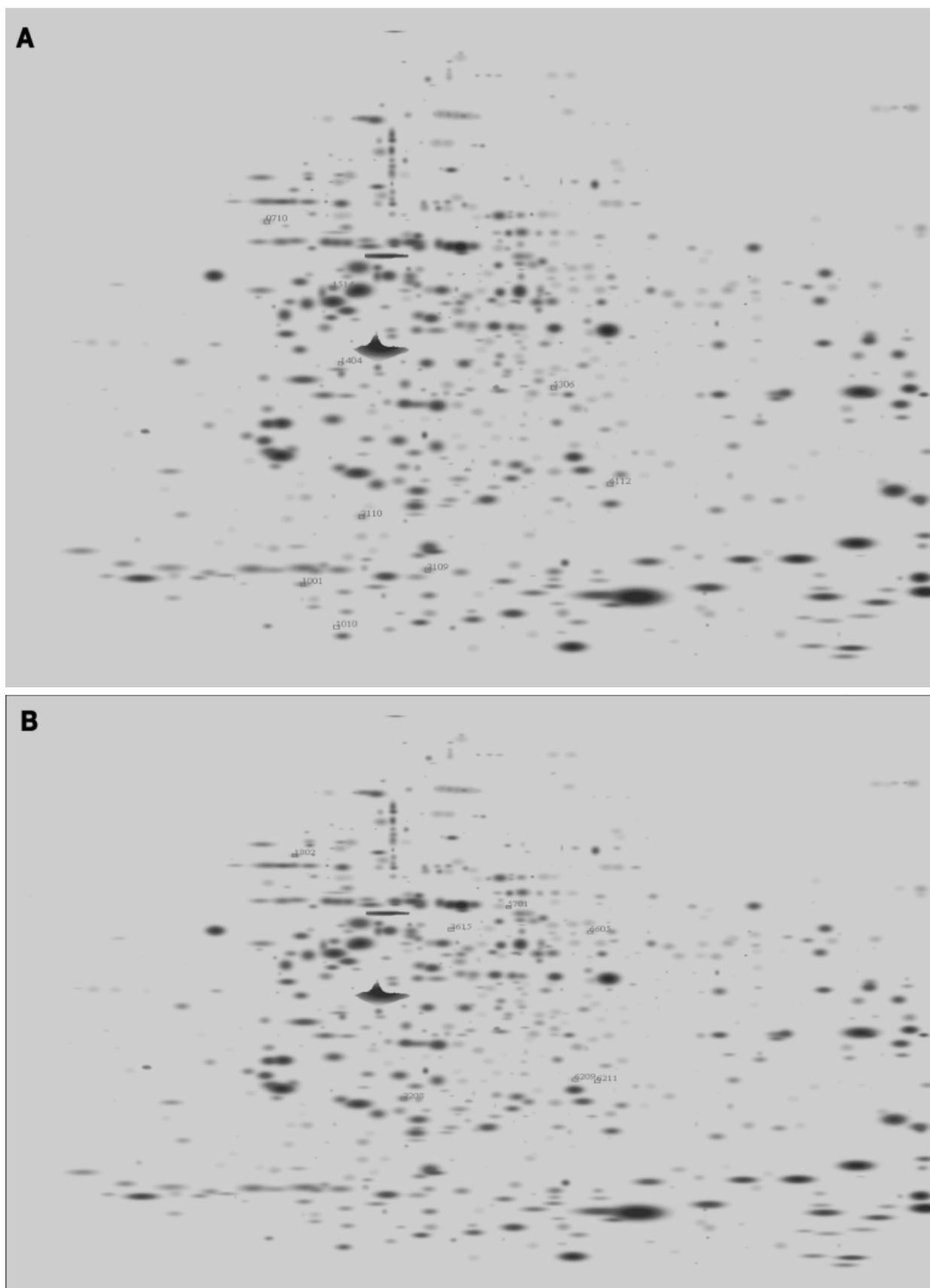
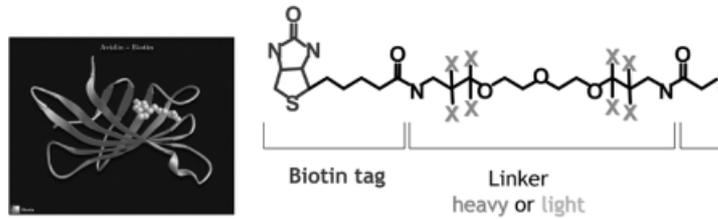


Figure 2 2-D pattern of Mantle Cell Lymphoma (MCL) tissue (A) and reactive lymph node (B). The 2-D electrophoresis was performed on a linear Immobiline gradient pH 3–10, followed by 8–18% SDS-PAGE. Proteins were detected by colloidal Coomassie staining. Image analysis of scanned gels was carried out using the PDQUEST software. The 2-D gels were

matched to find out quantitative differences of spot patterns. In panel (A) are highlighted nine proteins which are over-expressed on MCL tissue map. In panel (B) are highlighted seven proteins which are over-expressed on a reactive lymph node map (reprinted, with permission, from ref. 34).

ICAT - Isotope Coded Affinity Tags

Reagents: Heavy reagent **d8-ICAT** (X=deuterium)
Light reagent **d0-ICAT** (X = hydrogen)



Aebersold, R *et al.* *Nat Biotech* 2001; 19:946–51.

Figure 3 Chemical structure of the ICAT reagent. The reactive iodine atom at the chain terminus on the right is the one responsible for covalent binding to the Cys-SH groups (reprinted, with permission, from ref. 35).

Protein ID & Quantitation by ICAT

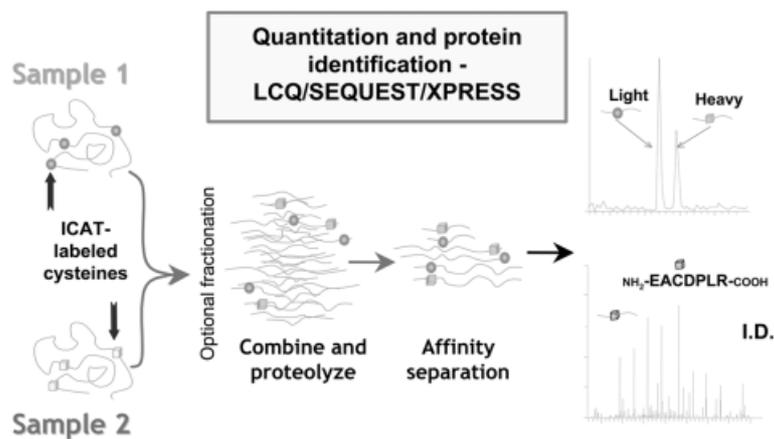


Figure 4 Protocol for the use of the ICAT label and affinity capture purification in quantitative proteomics (reprinted, with permission, from ref. 35).

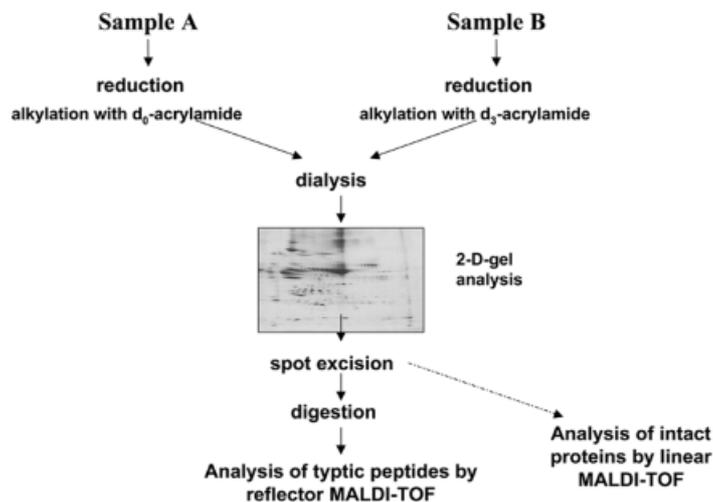


Figure 5 Protocol for the use of light/heavy acrylamide in quantitative proteomics of intact polypeptide chains (reprinted, with permission, from ref. 38).

Protein ID with Uninterpreted MS/MS

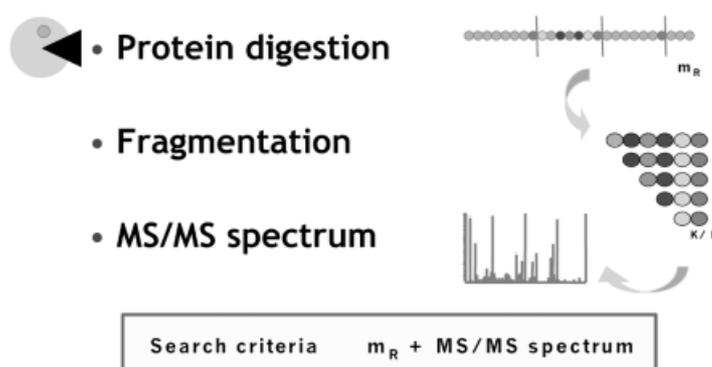


Figure 6 Protocol for identification of proteins extracted from a 2-D map *via* tryptic digestion and MALDI-TOF MS analysis, followed, if needed by an MS/MS sequencing procedure.

peptides are separated on a RP column, the deuterated peptide elutes earlier than its non-deuterated counterpart; that separation causes an enormous variation in isotope ratio across the two different elution profiles of the isoforms. The effect is more pronounced the smaller the tagged peptide. For example, in the case of a simple, cysteine-bearing octapeptide the chromatographic resolution between the light/heavy species was as high as $R_s = 0.74$ (note that an $R_s = 1.2$ means just baseline resolved peaks). Because the column eluate is continuously analyzed by electrospray ionization (ESI)-MS, obtaining the correct quantitative peak ratio based on the isotopic ratio becomes extremely difficult. Zhang and Regnier (39) reported that the resolution of the isoforms, in a C_{18} column, exceeds 0.5 with 20% of the peptides in the digest. Conversely, the same authors reported the complete absence of such isotope effect in the case of peptides differentially labeled with ^{13}C - and ^{12}C succinate, and they strongly recommended this type of peptide-coding when attempting separations on C_{18} column. A number of other approaches have also been reported. For example, Goodlett *et al.* (40) described methyl esterification (using d_0 - or d_3 -methanol) of peptides, a procedure which converts carboxylic acids on the side chains of aspartic and glutamic acids, as well as the carboxyl terminus, to their corresponding methyl esters. Derivatization of lysine to homoarginine was proposed by several groups (41–45) to enhance their intensity in the investigation of proteins/peptides by MALDI-TOF MS. The same derivatization procedure was later adapted by Cagney and Emili (46), who termed the approach “mass-coded abundance tagging” (MCAT), where C-terminal lysine residues of tryptic peptides are modified through differential guanidination (resulting in a mass difference of 42 Da, as opposed to 8 Da in the case of ICAT).

All of the above methods are indeed similar, in that they exploit mass differences among polypeptides for inferring their quantitative expression ratio. A totally different approach to quantitative proteomics, fluorescent labeling, has been recently described. The first report came from Unlu *et al.* (47), and the technique was

aptly termed differential in-gel electrophoresis (DIGE). In fluorescence 2-D DIGE, each sample is covalently labeled with a different mass- and charge-matched fluorophore, Cy3 [1-(5-carboxypentyl)-1'-propylindocarbocyanine halide N-hydroxysuccinimidyl ester] and Cy5 [1-(5-carboxypentyl)-1'-methylindocarbocyanine halide N-hydroxysuccinimidyl ester], before mixing the samples and analysis performed on the same 2-D gel. These structurally similar, but spectrally different (Cy3: $\lambda_{em} = 569$, orange color; Cy5, $\lambda_{em} = 645$, far red color), fluorophores undergo nucleophilic substitution reaction with the lysine ϵ -amino groups on proteins to form an amide. The dyes have very similar molecular masses and are positively charged to match the charge on the lysine group. Charge matching ensures that there is little shift in the pI over the unlabeled protein, although it would appear that >95% of the protein can significantly shift away from the unlabeled protein in the mass dimension, particularly at lower protein masses. When the two Cy3/Cy5 mixed samples are analyzed on the same 2-D gel, and the latter is monitored by different fluorescent gel imaging, the following scenario takes place: all matched spots, expressed in a 1:1 ratio in control and experimental tissues will appear in a violet color; those in which the Cy5 predominates will have progressively more pronounced red hues, related to the quantitative Cy5/Cy3 ratio, and those in which the Cy3 marker is more abundant will tend towards progressively bluish color (special filters transform the orange fluorescence of Cy3 into blue; that is why all articles deal in terms of blue/red fluorescence). Special image analysis software matches the images, quantifies the spots, normalizes the signals, and shows the difference in expression of any set of two proteins by comparison (48–50). More on this topic can be found in a review by Hamdan and Righetti (51).

Pre-Fractionation in Proteome Analysis

When analyzing protein spots from 2-D maps by MS, it appears that only generally abundant proteins (codon

bias >0.2) can be properly identified. Thus, the number of spots on a 2-D gel is not representative of the overall number or classes of expressed genes that can be analyzed. Gygi *et al.* (52) have calculated that, when loading only 40 μg of total yeast lysate, as done in the early days of 2-D mapping, only polypeptides with an abundance of at least 51000 copies/cell could be detected. With 0.5 mg of starting protein, proteins present at 1000 copies/cell could now be visualized by silver staining, but those present at 100 and 10 copies/cell could not be revealed. These authors thus concluded that the large range of protein expression levels limits the ability of the 2-D MS approach to analyze proteins of medium to low abundance, and thus the potential of this technique for proteome analysis is limited. This is indeed a severe limitation, since it is quite likely that the portion of proteomes we are presently missing are the most interesting one from the point of view of understanding cellular and regulatory proteins, since such low-abundance polypeptide chains will typically be regulatory proteins. Thus, presently available techniques of IEF/IPG-SDS-PAGE, coupled to MS or MS-MS, although being the best ones, do not appear to be suitable for global detection of proteins expressed by cells. As a corollary, the construction of complete, quantitative protein maps based on this approach will be challenging, even for relatively simple, unicellular organisms.

A way out of this impasse would be pre-fractionation. At present, two major approaches exist: chromatographic and electrophoretic. Fountoulakis' group has extensively developed this approach. First, Fountoulakis *et al.* (53) and Fountoulakis and Takács (54) adopted affinity chromatography on heparin gels as a pre-fractionation step for enriching certain protein fractions from the bacterium *Haemophilus influenzae*. In a second approach (55), the same lysate of *H. influenzae* was pre-fractionated by chromatofocusing on Polybuffer exchanger. In yet another procedure, Fountoulakis *et al.* (56) pre-fractionated the cytosolic soluble proteins of *H. influenzae* by hydrophobic interaction chromatography (HIC) on a TSK Phenyl column. In yet another variant, Fountoulakis *et al.* (57) reported enrichment of low abundance proteins of *E. coli* by hydroxyapatite chromatography. All these different chromatographic steps allowed the discovery and characterization of several hundred new polypeptide chains, which were present in the original, unfractionated lysate, at too low levels to be detected.

In terms of electrophoretic pre-fractionation, the most efficient procedure is that based on multicompartment electrolyzers (MCE), as devised by Righetti *et al.* (58), whose scheme is depicted in Figure 7. This method relies on isoelectric membranes, fabricated with the same Immobililine chemicals as these used in IPG fractionation (59). The advantages of such a procedure are immediately apparent:

- It is fully compatible with the subsequent first dimension separation, a focusing step based on the Immobililine technology. Such a pre-fractionation protocol is based on precisely the same concept of

immobilized pH gradients and thus protein mixtures harvested from the various chambers of this apparatus can be loaded onto IPG strips without a need for further treatment, in that they are isoelectric and devoid of any non-amphoteric ionic contaminant.

- It permits harvesting a population of proteins having pI values matching precisely the pH gradient of any narrow or wide IPG strip.
- The above leads to much reduced chances of protein precipitation. In fact, when an entire cell lysate is analyzed in a wide gradient, there are fewer risks of protein precipitation; on the contrary, when the same mixture is analyzed in a narrow gradient, massive precipitation of all non-isoelectric proteins could occur, with a strong risk of co-precipitation of proteins that would otherwise focus in the narrow pH interval.
- Due to the fact that only the proteins co-focusing in the same IPG interval will be present, much higher sample loads can be used, permitting detection of low-abundance proteins.
- Finally, in samples containing extreme ranges of protein concentrations (such as human serum, where a single protein, albumin, represents $>60\%$ of the total), one could assemble an isoelectric trap narrow enough to just eliminate the unwanted protein from the entire complex, this also permitting much higher sample loads without interference from the most abundant species.

The apparatus depicted in Figure 7 has been miniaturized by Herbert and Righetti (60, 61). In this particular case, a set up made of five chambers is divided by four membranes, with pIs of 3.0, 5.0, 6.0, and 10.5, selected to trap albumin in the central, narrow pI chamber. This permits concentration of other fractions and detection of many more, dilute species in other regions on the pH scale. By properly exploiting this pre-fractionation device, we have been able to capture and detect much

Multicompartment electrolyzer fractionation

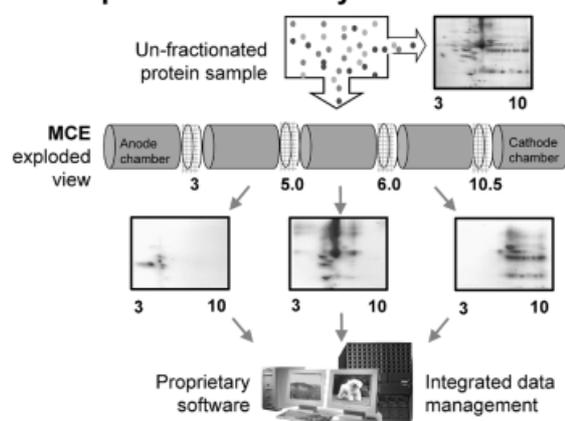


Figure 7 Scheme for sample pre-fractionation based on multicompartment electrolyzer. The upper right panel shows a 2-D map of unfractionated human serum vs. three different 2-D maps (lower panels) of three isoelectric fractions, captured into traps having membranes with pI 3.0–5.0; 5.0–6.0, and 6.0–10.5 (Herbert and Righetti, unpublished).

more of the "unseen" yeast membrane proteome (62).

The following were identified:

- 780 protein isoforms
- 323 unique proteins (genes)
- 105 integral membrane proteins (33%)
- 54 membrane associated proteins (17%)
- 159 total membrane/associated proteins (50%)
- 90 proteins with codon bias index (CBI) <0.2 (27%).

The importance of some of these findings is highlighted below: integral membrane proteins are rarely seen in 2-D maps; proteins with CBI <0.2 represent low abundance proteins and are scarcely detected in 2-D maps unless enriched by some pre-fractionation protocol. The power of our methodology, which not only relies on pre-fractionation steps but also deals with analysis of intact proteins, rather than proteolytic digests (as customary in most protocols exploiting coupled chromatographic processes; see below), can be appreciated also in Figure 8.

In yeast, there exist two forms of NADH-cytochrome b5 reductase, one called p34 (pI 8.7, M_r 34 kDa) and the other p32 (pI 7.8, M_r 30 kDa). The first species is an integral outer membrane protein, which mediates the reduction of cytochrome b5; the second one is derived from the first one by an *in vivo* proteolytic cleavage, it is soluble and resides into the intermembrane space. Our enrichment protocol, coupled to 2-D analysis of all intact polypeptide chains, detects both species. In addition, we can easily observe two isoforms of the p34 chain, one with pI 8.7 and the other one more alkaline, pI 9.1. Chromatographic techniques, which rely on the presence of just one (or a few) peptides in a total cell digest, would not have been able to detect any of these, biologically relevant, forms.

Two-Dimensional Chromatography

In chromatographic approaches, the proteins, in most cases, are digested into peptides prior to separation. The advantage is that peptides (especially from membrane proteins) are more soluble in a wide variety of solvents and hence easier to separate than their parent proteins. The disadvantage is the tremendous increment in the number of species to be resolved. Since, after fractionation in coupled columns, the eluate is sent

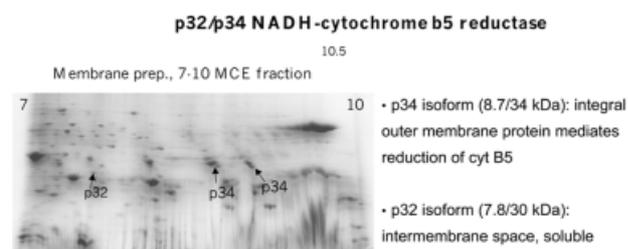


Figure 8 Example of pre-fractionation/enrichment of yeast membrane proteins and detection of truncated and isoforms of the p34 NADH-cytochrome b5 reductase (reprinted, with permission, from ref. 62).

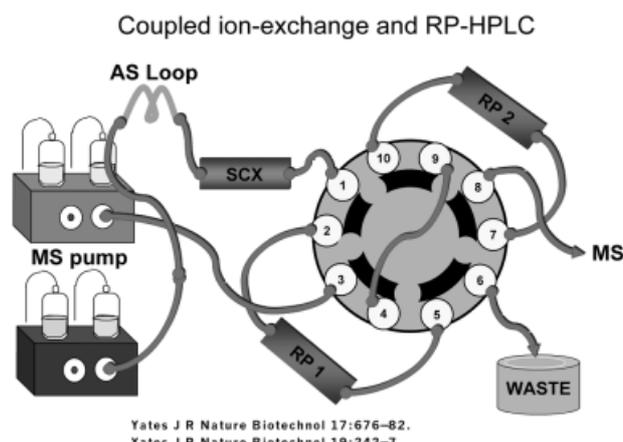


Figure 9 Scheme for coupled ion exchange and RP-HPLC in proteome analysis. SCX: strong cation exchanger column; AS loop: autosampler; MS: column eluate pumped to a mass spectrometer. It should be noted that, in most cases, such coupled columns are used for peptide fragment analysis.

directly to MS instrumentation, sample complexity might still be too high for proper analysis. While MS technology is rapidly improving, its dynamic measurement range is still 2–3 orders of magnitude less than the range of protein expression found within mammalian cells. In one of the earliest reports, Opiteck *et al.* (63, 64) described a 2-D HPLC system that used size exclusion chromatography (SEC) followed by RP-HPLC for mapping of *E. coli* proteins. Perhaps one of the most successful approaches, though, was the one of Yates and co-workers (65–67), who developed an on line 2-D ion-exchange column coupled to RP-HPLC, termed MudPit, for separating tryptic digests of 80S ribosomes from yeast. The acidified peptide mixture was loaded onto a strong cation exchanger (SCX) column; discrete eluate fractions were then fed onto a RP column, whose effluent was fed directly into a mass spectrometer, according to the scheme shown in Figure 9.

This iterative process was repeated twelve times using increasing salt gradient elution from the SCX bed and an increasing organic solvent concentration from the RP beads (typically a C₁₈ phase). In a total yeast lysate, the MudPit strategy allowed the identification of almost 1500 proteins (67). A similar procedure was used for separation of proteins and peptides in human plasma filtrates (68), plasma (69), blood ultrafiltrates (70), and human urines (71). The same set-up (SCX followed by C₁₈-RP) was followed by Davis *et al.* (72) for resolving a protein digest derived from conditioned medium from human lung fibroblasts. Perhaps the most sophisticated instrumentation was the one devised by Unger and co-workers (73) for processing proteins of M_r less than 20 kDa. The set-up consisted of two gradient HPLC instruments, two UV detectors, an isocratic pump, four RP columns, and ion-exchange column, four ten-port valves, an injection valve, two fraction collector stations, and a work station to control this fully automated system. This system was applied

ProteinChip® Arrays

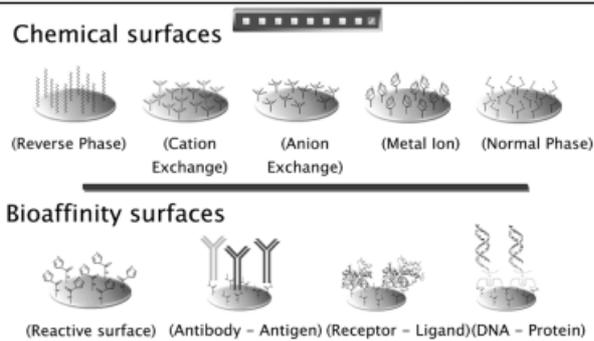


Figure 10 Scheme of different surfaces adopted in protein chip arrays (courtesy of CIPHERGEN).

to mapping of human hemofiltrates as well as lysates from human fetal fibroblasts. There are also other, hybrid systems, consisting in coupling, *e.g.*, an HPLC column to electrophoretic instrumentation, notably capillary electrophoresis. For those, we refer the reader to two recent reviews (74, 75).

Protein Chip Arrays

One of the latest technology platforms that has been developed for proteomics includes chip-based arrays. Such arrays have been elaborated for separating proteins based on surface chemistries or known protein ligands (*e.g.*, antibodies), with subsequent identification by MS (76–78). Other promising applications of protein chip microarrays have been differential profiling (79) and high-throughput functional analysis (80, 81). An example of such protein chip arrays is shown in Figure 10. Such arrays can be divided into chemical and bioaffinity surfaces. In the first case (see upper row in Figure 10), such surfaces function essentially like mini-chromatographic columns, in that they capture a given protein population by, *e.g.*, hydrophobic interaction (reverse phase), metal chelation, and different types of ion exchangers. Such chemical surfaces are not highly selective; nevertheless, they can be used in a cascade fashion, *e.g.*, proteins captured on ion-exchangers can be eluted and re-adsorbed onto a reverse phase, or on metal chelators so as to further sub-fractionate a most heterogeneous protein mixture as a total cell lysate. The bioaffinity surfaces (see lower row in Figure 10) clearly work on a much higher selectivity principle and allow capture of a narrow and well defined protein population, immediately ready for MS analysis. How the latter will be performed is shown schematically in Figure 11: once a selected protein population has been captured, the surface of the chip is bombarded with a laser beam, which will desorb and ionize the different proteins, ultimately identified by their precise molecular mass. The main difference between this kind of structural analysis and MALDI-TOF desorption ionization is that, in the latter case, the sample has to be manually transferred to a microwell plate,

Principle of ProteinChip® Process

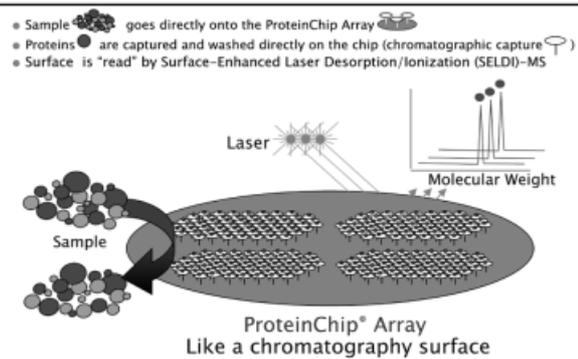


Figure 11 Scheme of processing/detection of protein species captured by different chemical surfaces (chromatographic adsorption; courtesy of CIPHERGEN).

admixed with a special matrix (such as sinapinic acid), and then desorbed/ionized *via* pulse laser shots. Conversely, in the protein chips illustrated here (Figure 11), a stick containing 8 (or multiples of 8) such affinity surfaces with the captured protein is directly inserted into the MS instrument and desorbed by surface-enhanced laser desorption/ionization (SELDI), in the absence of added matrix (the surface of such arrays being, in fact, already coated with energy absorbing polymers) (82). Such instrumentation is rapidly being adopted in many clinical chemistry laboratories worldwide and might soon become part of the standard instrumentation. We refer the reader to several recently published papers on the arrays exploiting SELDI technology (83–90).

Conclusions: Quo vadis, Proteome?

At the beginning of the 20th century, an astronomer would probably have described the universe as a flattened disk, 30000 light years in diameter and 6500 light years thick, comprising just a single galaxy, the Milky Way, with our solar system located near the center; a pictorial representation known as Kapteyn universe, after the Dutch astronomer Jacobus Cornelius Kapteyn. Three major discoveries helped us to profoundly reshape our universe. The first one came in 1924, when Edwin Hubble proved that, as our sun is just one of billions of stars in our home galaxy, the Milky Way was just one of billions of Galaxies (91). The second event was the discovery of the expansion of our universe, again by Hubble (92), which gave to our universe fantastic dimensions: not just a mere 30000 light years across (a gigantic size, mind you!) but enormously more, 30 billion light years across! The final major explosion came with the discovery by Arno Penzias and Robert Wilson (93) of the cosmic microwave background radiation, which paved the way to the Big Bang model on the birth of our universe.

By analogy with these cosmological discoveries, our proteome picture changed dramatically too during the 20th century. Up to the early seventies', scientists could

only separate a handful of proteins when attempting various electrophoretic approaches. The major event was surely the O'Farrell 2-D map protocol (10), which showed that even a simple microorganism such as an *E. coli* is endowed with at least 1100 polypeptide chains. Later, when working with human tissues (a pure cell line), Klose and Zeindl showed that eukaryotes displayed at least 10000 spots in a 2-D map (94). The second major event in proteome analysis came with the demonstration, in the late eighties', that proteins and peptides could be softly ionized and be amenable to analysis by MS, such as MALDI-TOF. This paved the way for protein recognition on a large scale. The third major event, that produced the Big Bang in proteomics, was the explosion of informatics, the building of genome and proteome databases, and the development of new, powerful algorithms allowing normalization and comparison of various 2-D maps for spotting events of upregulation and downregulation. Like in astronomy, we can now look at the starry sky represented by a 2-D map with a few thousand protein spots, and count each one of them, give them tri-dimensional (at least) coordinates, such as pI, M_r and spot volume, and hopefully assign to them first and last name, together with (ideally) a biological function. Like Saint Peter, who was reproached by an angel when trying to escape martyrdom by leaving Rome, and was made to turn back to meet his fate, we now have no excuse for abandoning the battle-field: we must face proteome and conquer it. The tools are there, and they are being constantly refined and made more powerful.

Acknowledgements

Supported in part by grants from ASI (Agenzia Spaziale Italiana, grant No. I/R/167/01) and by a European Community grant (No. QL2-CT-2001-01903).

References

- Wilkins MR, Williams KL, Appel RD, Hochstrasser DF, editors. Proteome research: new frontiers in functional genomics. Berlin: Springer, 1997:243 pp.
- Kellner R, Lottspeich F, Meyer HE, editors. Microcharacterization of proteins. Weinheim: Wiley-VCH, 1999:399 pp.
- Rabilloud T, editor. Proteome research: two-dimensional gel electrophoresis and identification methods. Berlin: Springer, 2000:248 pp.
- James P, editor. Proteome research: mass spectrometry. Berlin: Springer, 2001:274 pp.
- Righetti PG, Stoyanov AV, Zhukov MY. The proteome revisited: theory and practice of all relevant electrophoretic steps. Amsterdam: Elsevier, 2001:395 pp.
- Westermeier R. High resolution 2D electrophoresis. In: Westermeier R., editor. Electrophoresis in practice. Weinheim: VCH, 1997:213–28.
- Hanash S. Two dimensional gel electrophoresis. In: Hames BD, editor. Gel electrophoresis of proteins. Oxford: Oxford University Press, 1998:189–212.
- Celis JE, Ostergaard M, Jensen NA, Gromova I, Rasmussen HH. Human and mouse proteomic databases: novel resources in the protein universe. FEBS Letters 1998; 430: 64–72.
- Lander ES, Weinberg RA. Genomics: journey to the center of biology. Science 2000; 287:1777–82.
- O'Farrell P. High resolution two dimensional electrophoresis of proteins. J Biol Chem 1975; 250:4007–21.
- Hooker TF, Jeffrey DJ, Jorgenson JW. Two dimensional separations in high-performance capillary electrophoresis. In: Khaledi MG, editor. High performance capillary electrophoresis: theory, techniques and applications. New York: Wiley, 1998:581–612.
- Bjellqvist B, Ek K, Righetti PG, Gianazza E, Görg A, Postel W, *et al.* Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications. J Biochem Biophys Methods 1982; 6:317–39.
- Aebersold R, Leavitt J. Sequence analysis of proteins separated by polyacrylamide gel electrophoresis: towards an integrated protein database. Electrophoresis 1990; 11:517–27.
- Patterson SD, Aebersold R. Mass spectrometric approaches for the identification of gel-separated proteins. Electrophoresis 1995; 16:1791–814.
- Lahm HW, Langen H. Mass spectrometry: a tool for the identification of proteins separated by gels. Electrophoresis 2000; 21:2105–14.
- Lottspeich F, Houthaeve T, Kellner R. Chemical methods for protein synthesis analysis. In: Kellner R, Lottspeich F, Meyer HE, editors. Microcharacterization of proteins. Weinheim: Wiley-VCH, 1999:141–58.
- Wilkins MR, Gasteiger E, Sanchez JC, Bairoch A, Hochstrasser DF. Two-dimensional gel electrophoresis for proteome projects: the effects of protein hydrophobicity and copy number. Electrophoresis 1998; 19:1501–05.
- Molloy MP. Two dimensional electrophoresis of membrane proteins using immobilized pH gradients. Anal Biochem 2000; 280:1–10.
- Rabilloud T, Chevallet M. Solubilization of proteins in two dimensional electrophoresis. In: Rabilloud T, editor. Proteome research: two-dimensional gel electrophoresis and identification methods, Berlin: Springer, 2000:9–29.
- Musante L, Candiano G, Ghiggeri GM. Resolution of fibronectin and other characterized proteins by 2D-PAGE with thiourea. J Chromatogr B 1998; 705:351–6.
- Rabilloud T, Adessi C, Giraudel A, Lunardi J. Improvement of the solubilization of proteins in two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis 1997; 18:307–16.
- McCarthy J, Hopwood F, Oxley D, Laver D, Castagna A, Righetti PG, *et al.* Carbamylation of proteins in 2D electrophoresis – myth or reality? J Proteome Res 2003. In press.
- Hochstrasser DF, Harrington MG, Hochstrasser AC, Miller MJ, Merrill CR. Methods for increasing the resolution of two-dimensional protein electrophoresis. Anal Biochem 1988; 173:424–35.
- Rimpilainen M, Righetti PG. Membrane protein analysis by isoelectric focusing in immobilized pH gradients. Electrophoresis 1985; 6:419–22.
- Chevallet M, Santoni V, Poinas A, Rouquie D, Fuchs A, Kieffer S, *et al.* New zwitterionic detergents improve the analysis of membrane proteins by two-dimensional electrophoresis. Electrophoresis 1998; 19:1901–9.
- Molloy MP, Herbert BR, Slade MB, Rabilloud T, Nouwens AS, Williams KL, *et al.* Proteomic analysis of the *Escherichia coli* outer membrane. Eur J Biochem 2000; 267:1–12.
- Santoni V, Kieffer S, Desclaux D, Masson F, Rabilloud T. Membrane proteomics: use of additive main effects with

- multiplicative interaction model to classify plasma membrane proteins according to their solubility and electrophoretic properties. *Electrophoresis* 2000; 21:3329–44.
28. Bordini E, Hamdan M, Righetti PG. Probing the reactivity of S-S bridges to acrylamide in some proteins under high pH conditions by matrix-assisted laser-desorption/ionization. *Rapid Commun Mass Spectrom* 1999; 13:1818–27.
29. Righetti PG, Tudor G, Gianazza E. Effect of 2-mercaptoethanol on pH gradients in isoelectric focusing. *J Biochem Biophys Methods* 1982; 6:219–27.
30. Herbert BR, Molloy MP, Gooley BAA, Walsh J, Bryson WG, Williams KL. Improved protein solubility in two-dimensional electrophoresis using tributyl phosphine as reducing agent. *Electrophoresis* 1998; 19:845–51.
31. Herbert B, Galvani M, Hamdan M, Olivieri E, McCarthy J, Pedersen S, *et al.* Reduction and alkylation of proteins in preparation of two-dimensional map analysis: why, when and how? *Electrophoresis* 2001; 22:2046–57.
32. Galvani M, Hamdan M, Herbert B, Righetti PG. Alkylation kinetics of proteins in preparation for two-dimensional maps: a MALDI-TOF mass spectrometry investigation. *Electrophoresis* 2001; 22:2058–65.
33. Galvani M, Rovatti L, Hamdan M, Herbert B, Righetti PG. Proteins alkylation in presence/absence of thiourea in proteome analysis: a MALDI-TOF mass spectrometry investigation. *Electrophoresis* 2001; 22:2066–74.
34. Antonucci F, Chilosi M, Santacatterina M, Hamdan M, Astner H, Righetti PG. Two-dimensional molecular profiling of mantle cell lymphoma. *Proteomics* 2003. In press.
35. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotech* 1999; 17:994–9.
36. Han DK, Eng J, Zhou H, Aebersold R. Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nature Biotech* 2001; 19:946–51.
37. Sechi S. A method to identify and simultaneously determine the relative quantities of proteins isolated by gel electrophoresis. *Rapid Commun Mass Spectrom* 2002; 16:1416–24.
38. Gehanne S, Cecconi D, Carboni L, Righetti PG, Domenici E, Hamdan M. Quantitative analysis of two-dimensional gel-separated proteins using isotopically marked alkylating agents and matrix-assisted laser desorption/ionisation mass spectrometry. *Rapid Commun Mass Spectrom* 2002; 16:1692–8.
39. Zhang R, Regnier FJ. Minimizing resolution of isotopically coded peptides in comparative proteomics. *J Proteom Res* 2002; 1:139–47.
40. Goodlett DR, Keller A, Watts JD, Newitt R, Yi EC, Purvine S, *et al.* Differential stable isotope labelling of peptides for quantitation and de novo sequence derivation. *Rapid Commun Mass Spectrom* 2001; 15:1214–21.
41. Hale JE, Butler JP, Knierman MD, Becker GW. Increased sensitivity of tryptic peptides detection by MALDI-TOF mass spectrometry is achieved by conversion of lysine to homoarginine. *Anal Biochem* 2000; 287:110–7.
42. Keough T, Lacey MP, Youngquist RS. Derivatization procedures to facilitate de novo sequencing of lysine-terminated tryptic peptides using post-source decay matrix assisted laser desorption/ionisation mass spectrometry. *Rapid Commun Mass Spectrom* 2000; 14:2348–56.
43. Beardsley RL, Karty JA, Reilly JP. Enhancing the intensities of lysine-terminated tryptic peptide ions in matrix assisted laser desorption/ionisation mass spectrometry. *Rapid Commun Mass Spectrom* 2000; 14:2147–53.
44. Brancia F, Oliver, SG, Gaskell SJ. Improved matrix-assisted laser desorption/ionisation mass spectrometric analysis of tryptic hydrolysates of proteins following guanidination of lysine-containing peptides. *Rapid Commun Mass Spectrom* 2000; 14:2070–3.
45. Peters EC, Horn DM, Tully DC, Brock A. A novel multifunctional labelling reagent for enhanced protein characterisation with mass spectrometry. *Rapid Commun Mass Spectrom* 2001; 15:2387–92.
46. Cagney G, Emili A. De novo peptide sequencing and quantitative profiling of complex protein mixtures using mass-coded abundance tagging. *Nat Biotechnol* 2002; 20:163–70.
47. Unlu M, Morgan ME, Minden JS. Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* 1997; 18:2071–7.
48. Tonge R, Shaw J, Middleton B, Rowlinson R, Rayner S, Young J, *et al.* Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. *Proteomics* 2001; 1:377–96.
49. Zhou G, Li H, DeCamps D, Chen S, Shu H, Gong Y, *et al.* 2D differential in-gel electrophoresis for the identification of esophageal scans cell cancer-specific protein markers. *Proteomics* 2002; 2:117–24.
50. Gharbi S, Gaffney P, Yang A, Zvelebil MJ, Cramer R, Waterfield MD, *et al.* Evaluation of two-dimensional differential gel electrophoresis for proteomic expression analysis of a model breast cancer cell system. *Mol Cell Proteomics* 2002; 2:91–8.
51. Hamdan M, Righetti PG. Modern strategies for protein quantification in proteome analysis: advantages and limitations. *Mass Spectrom Reviews* 2002; 21:287–302.
52. Gygi SP, Corthals GL, Zhang Y, Rochon Y, Aebersold R. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc Natl Acad Sci USA* 2000; 97:9390–5.
53. Fountoulakis M, Langen H, Evers S, Gray C, Takacs B. Two-dimensional map of *Haemophilus influenzae* following protein enrichment by heparin chromatography. *Electrophoresis* 1997; 18:1193–202.
54. Fountoulakis M, Takacs B. Design of protein purification pathways: application to the proteome of *Haemophilus influenzae* using heparin chromatography. *Protein Expr Purif* 1998; 14:113–9.
55. Fountoulakis M, Langen H, Gray C, Takacs B. Enrichment and purification of proteins of *Haemophilus influenzae* by chromatofocusing. *J Chromatogr A* 1998; 806:279–91.
56. Fountoulakis M, Takacs MF, Takacs B. Enrichment of low-copy-number gene products by hydrophobic interaction chromatography. *J Chromatogr A* 1999; 833:157–68.
57. Fountoulakis M, Takacs MF, Berndt P, Langen H, Takacs B. Enrichment of low abundance proteins of *Escherichia coli* by hydroxyapatite chromatography. *Electrophoresis* 1999; 20:2181–95.
58. Righetti PG, Wenisch E, Jungbauer A, Katinger H, Faupel M. Preparative purification of human monoclonal antibody isoforms in a multicompartement electrolyzer with Immobiline membranes. *J Chromatogr* 1990; 500:681–96.
59. Righetti PG. Immobilized pH gradients: theory and methodology. Amsterdam: Elsevier, 1990:397 pp.
60. Herbert B, Righetti PG. A turning point in proteome analysis: sample pre-fractionation via multicompartement electrolyzers with isoelectric membranes. *Electrophoresis* 2000; 21:3639–48.
61. Righetti PG, Castagna A, Herbert B. Prefractionation techniques in proteome analysis. *Anal Chem* 2001; 73:320A–6A.
62. Pedersen SK, Harry JL, Sebastian L, Baker J, Traini MD,

- McCarthy J, *et al.* The unseen proteome: mining below the tip of the iceberg to find low abundance and membrane proteins. *J Proteome Res* 2003. In press.
63. Opiteck GJ, Lewis KC, Jorgenson JW. Comprehensive LC/LC/MS of proteins. *Anal Chem* 1997; 69:1518–24.
64. Opiteck GJ, Ramirez SM, Jorgenson JW, Moseley III MA. Comprehensive two-dimensional high-performance liquid chromatography for the isolation of overexpressed proteins and proteome mapping. *Anal Biochem* 1998; 258: 349–61.
65. Yates III JR 3rd, McCormack AL, Schieltz D, Carmack E, Link A. Direct analysis of protein mixtures by tandem mass spectrometry. *J Protein Chem* 1997; 16:495–7.
66. Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, *et al.* Direct analysis of protein complexes using mass spectrometry. *Nature Biotech* 1997; 17:676–82.
67. Washburn MP, Wolters D, Yates JR. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotech* 2001; 19:242–7.
68. Raida M, Schultz-Knape P, Heine G, Forssmann WG. Liquid chromatography and electrospray mass spectrometric mapping of peptides from human plasma filtrate. *J Amer Mass Spectrom* 1999; 10:45–54.
69. Richter R, Schultz-Knape P, Schrader M, Standker L, Jurgens M, Tammen H, *et al.* Composition of the peptide fraction in human blood plasma: database of circulating human peptides. *J Chromatogr B* 1999; 726:25–35.
70. Schrader M, Jurgens M, Hess R, Schultz-Knappe P, Raida M, Forssmann WG. Matrix-assisted laser desorption/ionisation mass spectrometry guided purification of human guanylin from blood ultrafiltrate. *J Chromatogr A* 1997; 766:139–45.
71. Heine G, Raida M, Forssmann WG. Mapping of peptides and protein fragments in human urine using liquid chromatography-mass spectrometry. *J Chromatogr A* 1997; 766:117–24.
72. Davis MT, Beierle J, Bures ET, McGinley MD, Mort J, Robinson JH, *et al.* Automated LC-LC-MS-MS platform using binary ion-exchange and gradient reversed-phase chromatography for improved proteomic analyses. *J Chromatogr B* 2001; 752:281–91.
73. Wagner K, Miliotis T, Marko-Varga G, Bischoff R, Unger KK. An automated on-line multidimensional HPLC system for protein and peptide mapping with integrated sample preparation. *Anal Chem* 2002; 74:809–20.
74. Issaq HJ, Conrads TP, Janini GM, Veenstra TD. Methods for fractionation, separation and profiling of proteins and peptides. *Electrophoresis* 2002; 23:3048–61.
75. Issaq HJ. The role of separation science in proteomics research. *Electrophoresis* 2001; 22:3629–38.
76. Borrebaeck CAK. Antibodies in diagnostics: from immunoassays to protein chips. *Immunol Today* 2000; 21:379–82.
77. Lueking AM, Horn H, Eickhoff H, Lehrqach H, Walter G. Protein microarrays for gene expression and antibody screening. *Anal Biochem* 1999; 270:103–11.
78. Borrebaeck CAK, Ekstrom S, Molmborg Hager AC, Nilson J, Laurell T, Marko-Varga G. Protein chips based on recombinant antibody fragments: a highly sensitive approach as detected by mass spectrometry. *BioTechniques* 2001; 30: 1126–32.
79. MacBeath G, Schreiber SL. Printing proteins as microarrays for high-throughput function determination. *Science* 2000; 289:1760–3.
80. Fung ET, Thulasiraman V, Weibernger SR, Dalmaso EA. Protein chips for differential profiling. *Curr Opin Biotechnol* 2001; 12:65–9.
81. Sawyer TK. Proteomics – structure and function. *BioTechniques* 2001; 31:156–60.
82. Voivodov KI, Ching J, Hutchens TW. Surface arrays of energy absorbing polymers enabling covalent attachment of biomolecules for subsequent laser-induced uncoupling/desorption. *Tetrahedron Lett* 1996; 37:5669–72.
83. Lin S, Tornatore P, King D, Orlando R, Weinberger SR. Limited acid hydrolysis as a means for fragmenting proteins isolated upon ProteinChip array surfaces. *Proteomics* 2001; 1:1172–84.
84. Paweletz CP, Liotta LA, Petricoin III EF. New technologies for biomarker analysis of prostate cancer progression: laser capture microdissection and tissue proteomics. *Urology* 2001; 57:160–4.
85. Boyle MD, Romer TG, Meeker AK, Sledjeski DD. Use of surface-enhanced desorption ionization protein chip system to analyze streptococcal exotoxin B activity secreted by streptococcus pyogenes. *J Microbiol Methods* 2001; 46: 87–97.
86. Wang S, Diamond DL, Hass GM, Sokoloff R, Vessella RL. Identification of prostate-specific membrane antigen (PMSA) as the target of monoclonal antibody 107–1A4 by protein chip array surface-enhanced laser desorption/ionization (SELDI) technology. *Int J Cancer* 2001; 92:871–6.
87. Thulasiraman V, McCutchen-Maloney SL, Motin VL, Garcia E. Detection and identification of virulence factors in *Yersinia pestis* using SELDI ProteinChip systems. *BioTechniques* 2000; 30:428–32.
88. Austen BM, Frears ER, Davies H. The use of SELDI ProteinChip arrays to monitor production of Alzheimer's β -amyloid in transfected cells. *J Pept Sep* 2000; 6:459–69.
89. Von Eggeling F, Davies H, Lomas L, Fiedler W, Junker K, Claussen U, *et al.* Tissue-specific microdissection coupled with ProteinChip array technologies: application in cancer research. *BioTechniques* 2000; 29:1066–70.
90. Merchant M, Weinberger SR. Recent developments in surface-enhanced laser desorption/ionization (SELDI)-time of flight mass spectrometry. *Electrophoresis* 2000; 21:1164–7.
91. Hubble E. The Milky Way and other galaxies in our universe. *Publ Amer Astron Soc* 1929; 5:261–70.
92. Hubble E. The red shift of stars and the expansion of our universe. *Proc Natl Acad Sci USA* 1929; 15:168–76.
93. Penzias AA, Wilson RW. The cosmic microwave background radiation and its implications in cosmogony. *Astrophys J* 1965; 142:419–26.
94. Klose J, Zeindl E. An attempt to resolve all the various proteins in a single human cell type by two-dimensional electrophoresis: I. extraction of all cell proteins. *Clin Chem* 1984; 30:2014–20.

Received 9 October 2002, revised 7 February 2003,
accepted 10 February 2003

Corresponding author: Prof. Pier Giorgio Righetti, University of Verona, Faculty of Sciences, Strada Le Grazie No. 15, 37134 Verona, Italy
Phone: +39-045-8027901, Fax: +39-045-8027929,
E-mail: righetti@sci.univr.it

Study of proteomic changes associated with healthy and tumoral murine samples in neuroblastoma by principal component analysis and classification methods

Emilio Marengo^{a,*}, Elisa Robotti^a, Pier Giorgio Righetti^b, Natascia Campostrini^b, Jennifer Pascali^b, Mirko Ponzoni^c, Mahmoud Hamdan^d, Hubert Astner^b

^aDepartment of Environmental and Life Sciences, University of Eastern Piedmont, Spalto Marengo 33-15100 Alessandria, Italy

^bScientific and Technological Department, University of Verona, Strada le Grazie 15-37134 Verona, Italy

^cLaboratory of Oncology, "G. Gaslini" Children's Hospital, Genoa, Italy

^dComputational, Analytical and Structural Sciences, GlaxoSmithKline, Via Fleming 4-37135 Verona, Italy

Received 19 December 2003; received in revised form 15 February 2004; accepted 16 February 2004

Abstract

Background: The adrenal gland is the election organ forming primary neuroblastoma (NB) tumours, the most common extracranial solid tumours of infancy and childhood. **Methods:** Samples of adrenal gland belonging to healthy and diseased nude mouse were analysed by 2D gel-electrophoresis. The resulting 2D-PAGE maps were digitized by PDQuest and investigated by principal component analysis (PCA). **Results:** The analysis of the loadings of the first principal component (PC) permitted the evaluation of the spots characterising each class of samples. Moreover, the soft-independent model of class analogy (SIMCA) method confirmed the separation of the samples in the two classes and allowed the identification of the modelling and discriminating spots. Very good correlation was found between the data obtained by analysis of 2D maps via the commercial software PDQuest and the present PCA analysis. In both cases, the comparison between such maps showed up- and down-regulation of 84 polypeptide chains, out of a total of 700 spots detected by a fluorescent stain, Sypro Ruby. Spots that were differentially expressed between the two groups were analysed by matrix-assisted laser desorption time-of-flight (MALDI-TOF) mass spectrometry and 14 of these spots were identified so far.

© 2004 Elsevier B.V. All rights reserved.

Keywords: 2D gel electrophoresis; Principal component analysis; Neuroblastoma; Proteomics; Multivariate analysis

1. Introduction

Proteomic research [1,2] has recently become a key tool for the study of several pathologies: the physio-

logical state of a particular cell or tissue is related to its protein content and the onset of a particular disease may cause differences in the proteins contained in the cell. The differences, which may arise, can consist of both changes in the relative abundance and in the appearance/disappearance of some proteins [3–11]. 2D gel electrophoresis can thus be applied for both diagnostic and prognostic purposes by comparing 2D-

* Corresponding author. Tel.: +39-131-287428; fax: +39-131-287416.

E-mail address: emilio.marengo@mf.unipmn.it (E. Marengo).

PAGE maps belonging to healthy subjects with samples belonging to individuals affected by different pathologies [3–11].

However, the comparison of different 2D-PAGE maps is not a trivial process to achieve: the difficulty which arises during the comparison is above all due to the high complexity of the specimen, which can produce maps with thousands of spots; the complexity is also increased by the highly articulated sample pre-treatment, often characterised by many purification/extraction steps. These experimental steps may cause the appearance of spurious spots due to impurities in the final 2D-maps. Moreover, the differences occurring between healthy and pathological samples may be very small, thus complicating their identification in a real complex map.

The differential analysis between classes of different 2D-PAGE maps is usually achieved by means of a specific software (i.e., Melanie III or PDQuest) [12–15]. Recently, Zhan and Desiderio [16] determined the spatial and quantitative reproducibility between two commercial 2D gel electrophoresis systems.

Two different approaches are available for the comparison of classes of 2D maps: comparing the proteomic pattern of the investigated samples or comparing the samples based on their spot volume. Both the proposed approaches produce a huge amount of information, which can be investigated by multivariate techniques. Our research group has developed a new method based on fuzzy logic and on the successive application of classification methods for the comparison of the proteomic pattern of classes of 2D-PAGE maps [17,18]; the proposed method has also been applied with multidimensional scaling (MDS) for the study of 2D-maps of control and nicotine-treated rat serum samples [19]. Different proteomic patterns have also been investigated by the use of three-way principal component analysis (PCA) [20]. Recently, (PCA) [21–24] has been applied to the study of spot volume data sets of DNA and RNA fragments of several biological systems [25–28], and the characterisation of proteomic patterns of different classes of tissues [29–35]. PCA has been recently adopted also for the characterisation of the regulatory activity of new drugs [36–39].

In the present paper, a data set constituted by adrenal nude mouse gland is investigated. The adrenal gland is the election organ forming primary neuroblas-

toma (NB) tumours. Neuroblastoma, together with lymphoma, osteosarcoma, Ewing's tumours, rhabdomyosarcoma and lymphoblastic leukemia, belongs to a group of undifferentiated paediatric malignancies known as the small round-cell tumours of childhood. NB is the most common extracranial solid tumour of infancy and childhood. It arises from primitive neuro-epithelial cells of the neural crest and occurs in the adrenal medulla or paraspinal sympathetic ganglia of the abdomen, chest or neck.

Neuroblastoma accounts for approximately 9% of all childhood cancers, occurring once out of 8000 live births. This results in an annual incidence of approximately 1 in 100,000 children less than 15 years of age worldwide. The median age at diagnosis is approximately 22 months with over one-third diagnosed at less than 1 year and over 88% diagnosed by the age of 5 years.

Despite various therapeutical treatments, including radiotherapy, immunotherapy and chemotherapy at elevated doses, with or without bone marrow transplant, childhood cancers arising from neural crest cells are still characterised by a high percentage of relapse and a high degree of mortality. Starting from these considerations, there is a great need of designing novel therapeutic strategies.

To be maximally effective, cancer therapy has to be individualized. This requires a detailed understanding of the properties of each tumour with regard to metastasising propensity and drug sensitivity, information that is difficult or even impossible to gain from microscopical analysis of hematoxylin-stained tissue sections. Detailed information regarding mutations in growth-controlling genes, expression of genes controlling growth, metastasis and drug resistance must be obtained. Proteome analysis is one of the available approaches for defining markers useful for tumour diagnosis. In addition, proteomics has the potential to unravel basic tumour biological questions regarding mechanisms involved in the pathogenesis of cancer.

PCA is here applied to a data set composed by six samples of healthy adrenal nude mouse gland and adrenal nude mouse gland mass, plus primary tumours, from which 20 2D maps were produced (10 from the control samples and 10 from the pathological ones). The successive application of classification methods, like soft-independent model of class analogy (SIMCA) [40], is useful for the identification

of classes of samples in the data set. The results thus obtained have been compared with those generated by the analysis of the same set of 20 2D maps with the commercial software PDQuest. A number of polypeptide chains, found to be up- and down-regulated, have been eluted from the gel slabs and identified via mass spectrometry in the matrix-assisted laser desorption time-of-flight (MALDI-TOF) mode.

2. Theory

2.1. Principal component analysis

PCA [21–24] is a multivariate statistical method which allows the representation of the original data set in a new reference system characterised by new variables, called principal components (PCs). Each PC has the property of explaining the maximum possible amount of variance contained in the original data set. The PCs, which are expressed as linear combinations of the original variables, are orthogonal to each other and can be used for an effective representation of the system under investigation, with a lower number of variables than in the original case. The co-ordinates of the samples in the new reference system are called *scores*, while the coefficient of the linear combination describing each PC, i.e. the weights of the original variables on each PC, are called *loadings*.

2.2. Classification: SIMCA model

SIMCA [40,41] is a classification method based on principal component analysis. This method builds boxes containing the samples of each class, based on a class-dependent PC model. Each object is then assigned to a class by calculating its distance from the different PCs model calculated on the samples of each class. As this method is based on PCs and not on the original variables, the main advantage of its application is that the objects are assigned to each class considering only the systematic and useful information contained in the data set, overlooking the uninformative one. Moreover, the method is also suitable for investigating small data set, like the objects of the present study. This depends on the possibility to perform a substantial size reduction of

the descriptors through the calculation of the principal components.

3. Materials and methods

3.1. Chemicals and materials

Tris, mineral oil, DL-dithiothreitol (DTT) and EDTA were purchased from Sigma-Aldrich (Steinheim, Germany). Glycerol, methanol, ethanol, acetic acid and acetone were from Merck (Darmstadt, Germany). Acrylamide/Bis solution (40%), *N,N,N',N'*-tetramethylethylenediamine (TEMED), ammonium persulfate (APS) and IPG strips 17 cm long, pH 3–10, were obtained from Bio-Rad Labs (Hercules, CA, USA); glycine, sodium dodecyl sulfate (SDS), iodoacetamide, urea, thiourea, tributylphosphine (TBP), CHAPS were from Fluka. Bromophenol blue, carrier ampholytes and agarose were purchased from Pharmacia Biotech (Uppsala, Sweden). All reagents were analytical grade.

3.2. Orthotopic neuroblastoma animal model

Five-week-old female nude (nu/nu) mice were purchased from Harlan Laboratories (Harlan Italy, S. Pietro al Natisone, UD). Mice were anaesthetised and injected with 5×10^4 cultured murine NXS2 neuroblastoma cells [42] (kindly provided by Dr. Reisfield, Scripps Clinic, La Jolla, CA, USA) in 20 μ l of HEPES buffer, after laparotomy, in the capsule of the left adrenal gland. The lethality of the method was 0%. Mice were monitored at least two times weekly for evidence of tumour development, quantification of tumour size and evidence of tumour-associated morbidity. Mice were sacrificed and organs, after washing in PBS, were frozen and stored in liquid nitrogen. All experiments involving animals have been reviewed and approved by the licensing and ethical committee of the National Cancer Research Institute and by the Italian Ministry of Health.

Six samples, divided into two classes, constituted the data set:

- six samples belonging to healthy adrenal nude mouse gland (the pool of which was used to perform 10 2D maps, Fig. 1a);

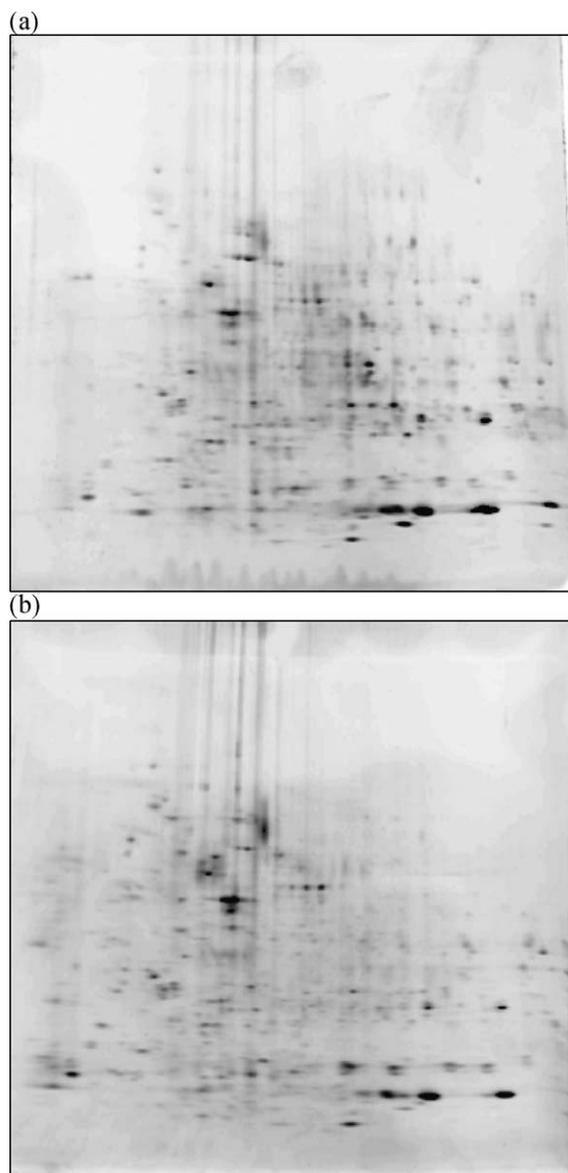


Fig. 1. Examples of the obtained 2D PAGE maps: control (a) and diseased (b) samples.

- six samples belonging to adrenal nude mouse gland mass, plus metastasis at other key organs (the pool of which was used to obtain 10 2D maps, Fig. 1b).

The procedure was repeated three times, thus producing three genuine replicates for each map (30 2D maps belonging to the healthy individuals and 30 2D maps belonging to the diseased ones).

3.3. Sample treatment

Healthy and tumoral adrenal glands from nude mice (nu/nu) were stored at -80°C . All samples were homogenized (5% homogenate) with a lysis solution containing 7 M urea, 2 M thiourea, 3% CHAPS, 40 mM Tris, 5 mM TBP, 0.5% carrier ampholytes pH 3–10, 5 mM EDTA, 1 mM PMSE, 50 U/ml DNase. After 1 h lysis, samples were centrifuged at 4°C at 6000 rpm for eliminating all residual particles. The collected supernatant was alkylated for 1 h with 10 mM acrylamide (1 M stock solution) and the reaction was stopped with 10 mM DTT (from 1 M stock solution). Interfering substances (lipids, salts) were removed by a precipitation step. A cold mix of acetone and methanol (8:1) was added; after 2 h at -20°C , the solution was centrifuged at 13,000 rpm for 30 min. Pellets were finally resuspended in sample solution containing urea, thiourea, CHAPS and Tris.

3.4. Electrophoretic separation: first dimension

The first dimension run was performed on strips (17 cm length, 0.5 mm thickness) with a linear pH gradient from pH 3 to 10 (Bio-Rad). The IPG strips were rehydrated with 400 μg of pooled samples (10 with pooled healthy samples and 10 with pooled tumor samples) and containing traces of bromophenol blue for monitoring the electrophoretic run. The passive gel rehydration was allowed to continue for 8 h before the focusing step. Isoelectric focusing (IEF) was carried out with a Protean IEF Cell (Bio-Rad), with a low initial voltage and then by applying a voltage gradient up to 10,000 V with a limiting current of 50 $\mu\text{A}/\text{strip}$. The total product time \times voltage applied was 75,000 V h for each strip and the temperature was set at 20°C .

3.5. Interfacing the IPG strips with the denaturing SDS solution

Each strip was equilibrated with an SDS denaturing solution containing 6 M urea, 2% SDS, 20% glycerol, 0.375 M Tris-HCl (pH 8.8). The contact lasted for 30 min in tubes containing 20 ml each of the equilibration solution. Each strip was then interfaced with a gel slab using 0.5% agarose, solubilized in the

Table 1
Results of PCA (for the first four PCs) performed on the overall data set

	% Explained variance	% Cumulative explained variance
PC ₁	47.3	47.3
PC ₂	14.5	61.8
PC ₃	10.6	72.4
PC ₄	8.3	80.7

cathodic buffer (192 mM glycine, 0.1% SDS, Tris to pH 8.3). All the gel slabs were cast with a two-vessel gradient mixer, with total sufficient volumes for polymerising 20 gel slabs of 1.5 mm in thickness and with a pore gradient from 7% to 20% T. Polymerization took place overnight.

3.6. Electrophoretic separation: second dimension

The second dimension run was performed by using a PROTEAN II xl Multi-Cell (Bio-Rad). The cathodic buffer was the same as in the previous step; the anodic buffer was a solution of 0.375 M Tris–HCl (pH 8.8). The electrophoretic run was performed by setting a current of 2 mA for each gel for 1 h, then 5 mA/gel for 2 h and 10 mA/gel until the end of the run. During the whole run, the temperature was set at 11 °C.

3.7. Fluorescent staining

All gels were fixed with a solution containing 40% ethanol, 10% acetic acid for 30 min. After this step, the gels were stained with Sypro Ruby (70 ml for each gel) overnight and then destained with 10% methanol, 7% acetic acid for 1 h. The gels were finally washed with milliQ water. All the samples were then digitized with the Versa Doc Scanner (Bio-Rad).

3.8. Protein identification by mass spectrometry

3.8.1. In situ digestion and extraction of peptides and MALDI-TOF analysis

The spots of interest were carefully excised from the gel with a razor blade and placed in Eppendorf tubes. The gel pieces were washed twice with a solution of acetonitrile/Tris 5 mM pH 8.5 (50:50) followed by a single wash with only Tris 5 mM pH 8.5. These pieces were dehydrated in a Speedvac device at room temperature and covered with 15 µl of Trypsin (0.02 mg/ml) in NH₄HCO₃ buffer (40 mM, pH 8.5) and left at 37 °C overnight. The peptides were extracted two times in 50 µl of acetonitrile/H₂O 1% v/v formic acid (50:50). The extraction was conducted in an ultrasonic bath for 15 min each time. The extract was brought to dryness in Speedvac and then resuspended with 10 µl of a H₂O 0.1% TFA solution.

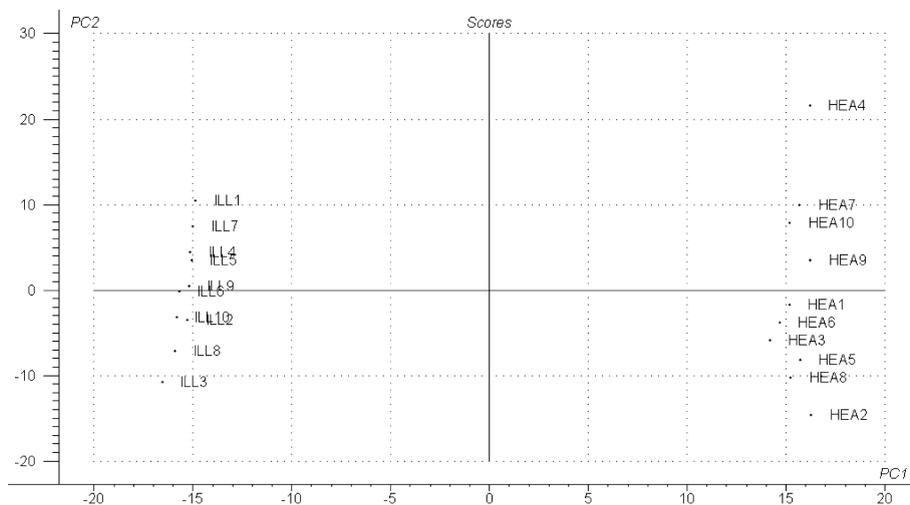


Fig. 2. Score plot of PC₂ vs. PC₁.

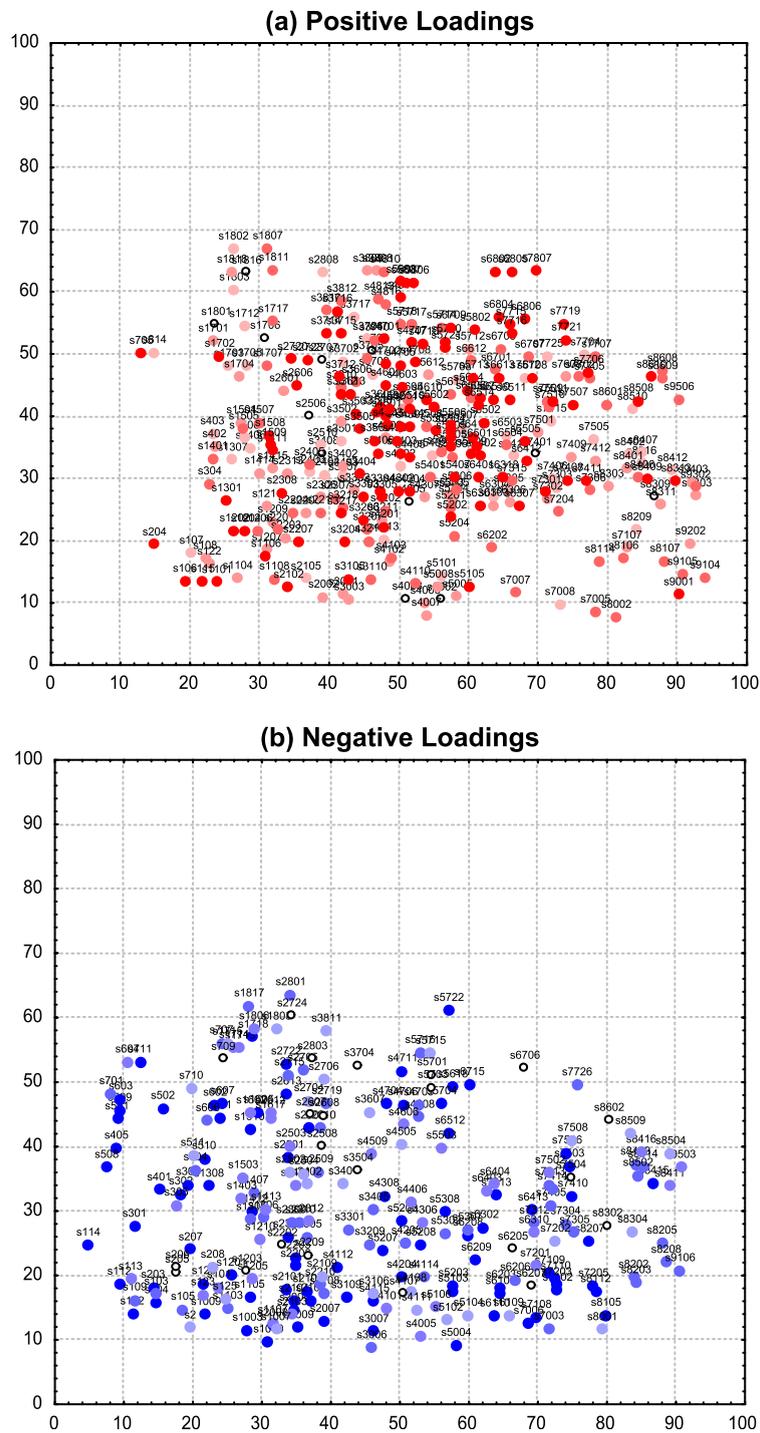


Fig. 3. Loading plots for the first principal component: positive (a) and negative loadings (b). Spots are identified by their SSP value.

The extracted peptides were loaded onto the MALDI target plate by mixing 1 μ l of each solution with the same volume of a matrix solution, prepared fresh every day by dissolving 10 mg/ml cyano-4-hydroxycinnamic acid in acetonitrile/ethanol (1:1, v/v) and allowed to dry. Measurements were performed using a ToFSpec 2E MALDI-TOF instrument (Micromass, Manchester, UK), operated in the reflectron mode, with an accelerating voltage of 20 kV. The laser wavelength was 337 nm and the laser repetition rate was 4 Hz. The final mass spectra were produced by averaging 50–200 laser shots. Peptide masses were searched against SWISS-PROT, TrEMBLE and NCBI nr databases by utilizing the ProteinLynx program from Micromass, Profound from Prowl and Mascot from Matrix Science.

3.9. Software

PCA was performed with UNSCRAMBLER (Camo, version 7.6), together with classification (SIMCA method). Graphical representations were performed with both UNSCRAMBLER and STATISTICA (Statsoft, ver. 5.1). 2D-PAGE maps were scanned with a Versa Doc (Bio-Rad), and analysed with the software PDQuest Version 6.2 (Bio-Rad).

4. Results and discussion

4.1. Principal component analysis

The differential analysis performed by PDQuest on 60 samples permitted the identification of 532 spots. The matching procedure produced a data set constituted by 532 variables (the optic density of each matched spot) and 60 objects (60 samples), thus giving a data matrix 60×532 . As the three genuine replicates of each map proved to have a good reproducibility, their results were averaged, thus reducing the data matrix to 20×532 , where the 20 samples are constituted by 10 averaged healthy samples and 10 averaged diseased ones. All variables were autoscaled before performing PCA, which gave the results reported in Table 1. The first three PCs explain more than 72% of the total variance contained in the original data set, but only the first PC was considered for the successive analysis be-

cause it showed to be able to clearly separate the two classes of samples.

Fig. 2 represents the score plot of the first two principal components. The score plot shows that the first PC allows an effective separation of the samples in the two classes: the diseased samples (ILL1–ILL10) lay at large negative scores on PC₁, while the control ones (HEA1–HEA10) are grouped at large positive values on the same PC.

The loading plots of PC₁ (Fig. 3) represents the spots as circles centred in the x - y position identified by means of PDQuest analysis. The red-coloured spots identify those spots showing a large positive loading on the first PC, while the blue-coloured ones identify the spots showing large negative values on the same PC. The spots are represented in a colour scale ranging from light to dark red or blue, where the increasing tone is proportional to the corresponding loading. The spots showing a small influence are light red (small positive loading) or light blue (small negative loading) coloured, while those showing a large influence are dark red (large positive loading) or dark blue (large negative loadings) coloured. The spots marked as a black circle do not have a relevant loading on the first PC. The loading plots (Fig. 3) allow the identification of the spots which show an increase in the optical density with the onset of the disease (blue-coloured spots, large negative loadings on PC₁) or which are absent in the control samples. The red spots show instead the proteins, which decrease or are absent after the appearance of the disease.

4.2. Classification

As PCA is able to identify two groups of samples in the data set, a classification method (SIMCA) is applied in order to verify the presence of the two classes and identify the modelling variables (the spots

Table 2
Results of PCA (for the first two PCs) performed on the control and diseased classes separately

	Control class		Diseased class	
	% Explained variance	% Cumulative explained variance	% Explained variance	% Cumulative explained variance
PC ₁	43.3	43.3	46.0	46.0
PC ₂	29.5	72.8	31.7	77.7

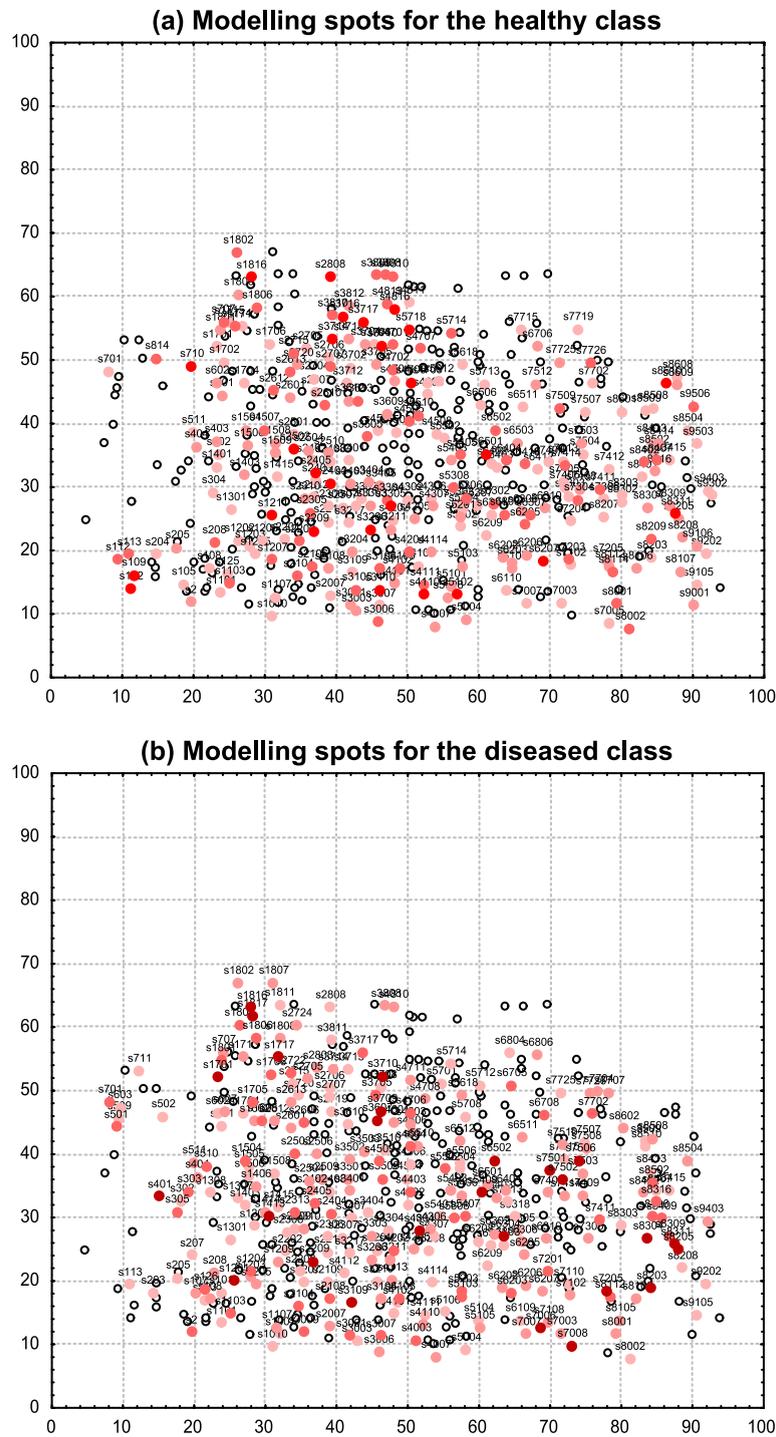


Fig. 4. Modelling variables obtained by the calculation of the SIMCA model for the two classes of samples: modelling variables for the healthy (a) and the diseased (b) classes. Spots are identified by their SSP value.

which describe each class) and the discriminating ones (the spots which are responsible for the differences between the two classes). SIMCA was selected as it is able to perform well even when small data sets are available.

PCA was then performed on the two classes (healthy and pathological samples) separately and a model based on principal components was calculated for each class. For both classes, the samples are exhaustively described by a model which includes only the first principal component; the results of PCA performed on the two separate data sets is reported in Table 2. The SIMCA method, performed on the overall data set, by considering the two models just calculated, allowed the correct classification of the samples in control and pathological ones, with a subsequent non-error rate (NER%) of 100%.

The SIMCA model, as just pointed out, allows the identification of both the modelling and the discriminating variables. The first ones are represented in Fig. 4 for both healthy and pathological

samples; the spots are represented in a colour scale, according to their modelling power (ranging from 0 to 1): spots from light to dark red show an increasing modelling power, while spots marked as a black circle do not play a relevant role in the definition of the SIMCA model built for each class. The spots showing the largest modelling power for each class are those, which describe the class more effectively; the comparison between the modelling spots of Fig. 4 and the loadings plots of Fig. 3 confirms the conclusions just driven by the performance of PCA. The modelling spots for the healthy class show positive loadings on PC₁, while the modelling ones for the diseased class show negative loadings on the same PC.

The discriminating spots are instead represented in Fig. 5. As in the previous case, the spots are presented in a colour scale from light to dark red, with respect to their increasing discriminating power; the spots marked as a black circle are not effective in the discrimination of the two classes. This representation sums up the information about the up- and down-

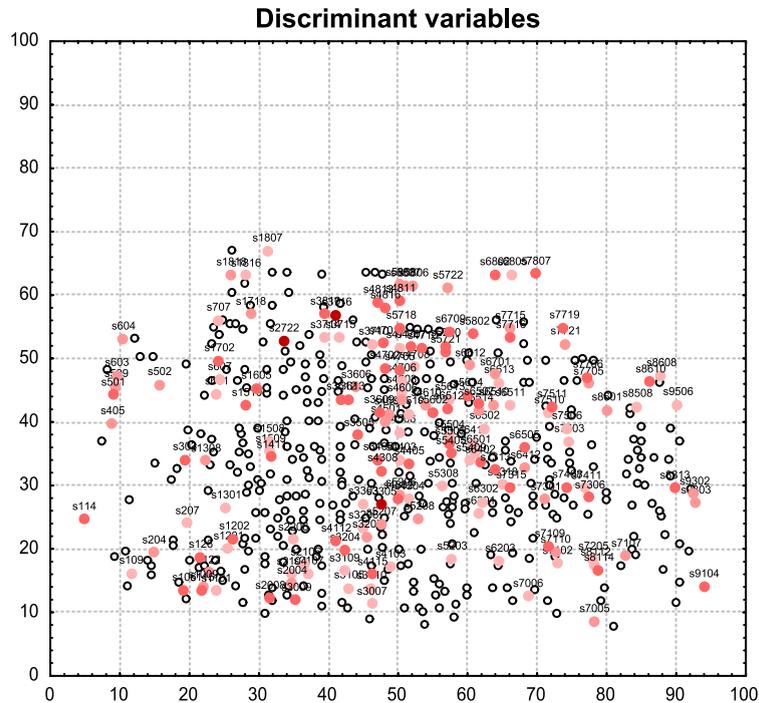


Fig. 5. Plot of the discriminant variables obtained by the calculation of the SIMCA model for the two classes of samples. Spots are identified by their SSP value.

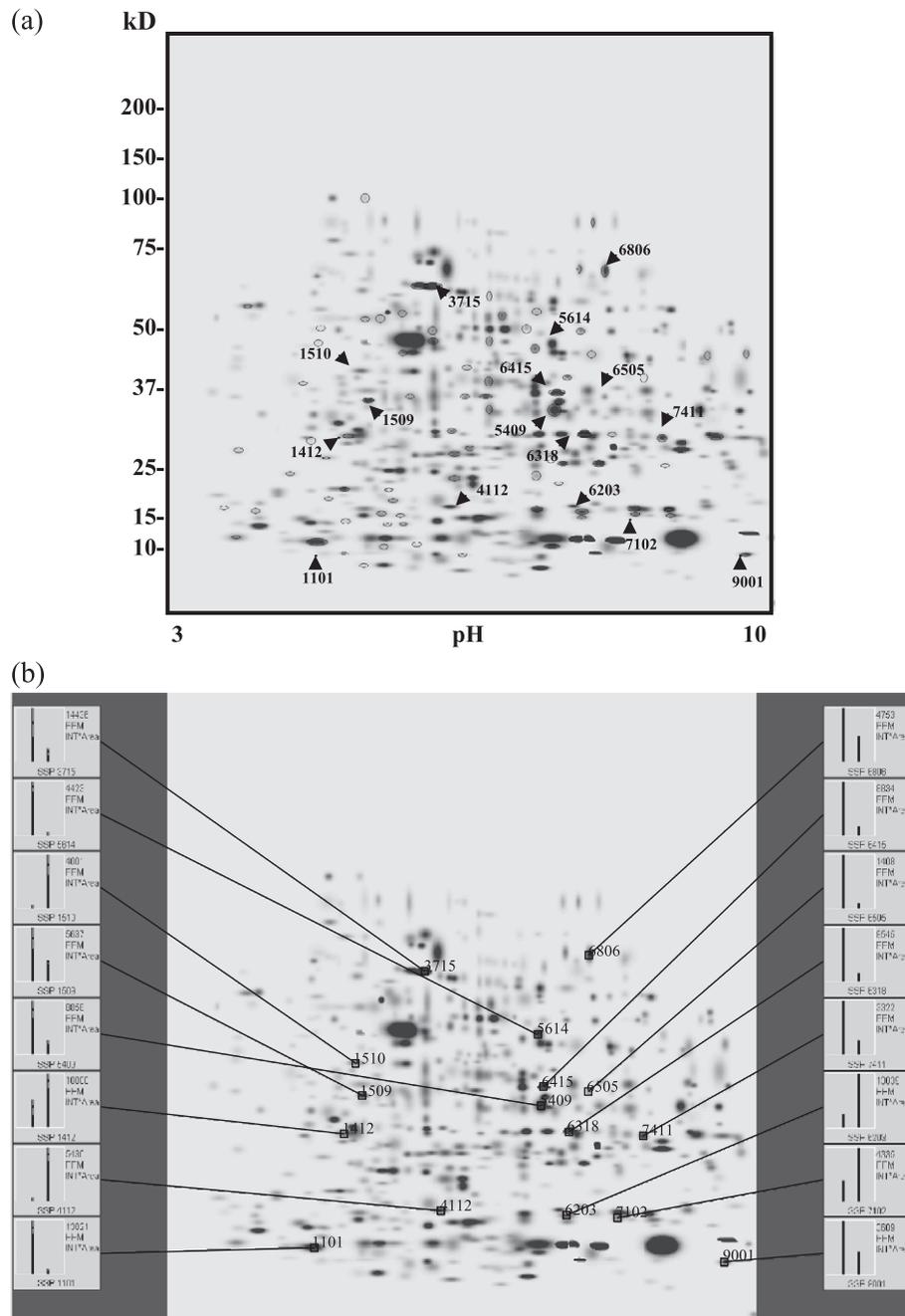


Fig. 6. Master maps of neuroblastoma tissue excised from the adrenal glands of nude mice. (a) The 84 spots differently expressed are highlighted and the identified spots are numbered. The numbers refer to the 14 spots, which could be identified by mass spectrometry upon excision. (b) Master map of neuroblastoma tissue with relevant histogram graphs of differentially expressed proteins identified by mass spectrometry. The standard spot (SSP) number is displayed beneath the histogram. The bars on the histograms display the maximum spot quantitation for each replicate group, the average spot quantitation and the standard deviation. Each bar represents one replicate group and the order of the bars corresponds to the order in which the replicate groups were made (the left bar corresponds to healthy samples and the right bar corresponds to tumoral ones).

Table 3
Identified proteins from the neuroblastoma tissue

Spot number	Experimental M_r (Da)	Experimental pI	Theoretical M_r (Da)	Theoretical pI	Z-score	MOWSE score	Protein name	Accession number	Coverage (%)	No. of peptides	Variation
6203	~ 15,000	7.8	17,840	8.3	2.39	1.93×10^8	Peptidyl-prolyl <i>cis-trans</i> isomeraseA	P17742	59.5	13	Increased 5
1101	~ 27,000	7.0	20,123	5.5	1.84	1.60×10^4	Adrenodoxin	P46656	25	8	Decreased 21
1510	~ 40,000	4.5	32,560	4.6	2.38	2.51×10^5	Nucleophosmin (NPM)	Q61937	31	8	Only tumoral samples
4112	~ 17,000	6.5	17,143	6.0	2.33	3.78×10^7	Stathmin (OP18)	P54227	64.9	17	Only tumoral samples
5614	~ 50,000	7.5	46,661	6.9	2.31	4.46×10^8	Isocitrate dehydrogenase	O88844	33.3	15	Only control samples
6318	~ 30,000	7.8	29,366	7.4	1.67	1.78×10^5	Carbonic anhydrase III	P14141	28	8	Decreased 8
6415	~ 37,000	7.5	35,857	7.2	2.29	1.77×10^{11}	Aldose reductase related protein 1	P21300	59.3	18	Decreased 8
6505	~ 35,000	8.0	33,335	8.2	2.1	2.68×10^7	Thiosulfate sulfurtransferase	P52196	39	13	Only control samples
3715	~ 65,000	6.5	60,956	6.1	2.37	2.61×10^{21}	60 kDa Heat shock protein	P19226	57	37	Decreased 4
5409	~ 34,000	7.5	34,973	8.6	2.38		Electron transfer flavoprotein	NCBI gi21704230	41	12	Decreased 4
6806	~ 72,000	8.1	67,631	7.7	2.41	2.09×10^9	Transketolase	P40142	28	18	Decreased 2
1412	~ 30,000	5.1	27,771	4.8	2.06	2.57×10^6	14-3-3 Protein zeta/delta (Protein kinase C inhibitor protein-1)	P35215	42	9	Increased 2
7411	~ 27,000	8.9	31,517	8.6	2.28	6.78×10^6	Enoyl-CoA hydratase	P14604	34	10	Decreased 4
1509	~ 36,000	5.3	35,753	4.8	2.38	2.70×10^{13}	Annexin V	P48036	64	22	Decreased 2
7102	~ 15,000	8.5	?	?	?	?	?	?	?	?	Increased 3
9001	~ 8500	9.8	?	?	?	?	?	?	?	?	Decreased 2

The comparison is performed between the matched spots in healthy samples and the matched spots in tumoral ones. A threshold value of 2.0 (which corresponds to a variation of 100%) was chosen as a meaningful variation in the comparison of tumour vs. healthy proteins.

regulated spots responsible for the differences occurring between the two classes investigated.

In order to check the validity of the results obtained by PCA data treatment, the same set of gels was analysed via commercially available software, namely the PDQuest, a software which has been continuously refined and improved over the years since its inception in 1979 [43]. The results of such data handling are shown in Fig. 6a, which displays, highlighted, a total of 84 spots differentially expressed in normal vs. pathological tissues. A number of these spots were eluted from the gels, digested and analysed by mass spectrometry in the MALDI-TOF mode. Fourteen of these polypeptide chains could be identified and are labelled in Fig. 6a by their respective SSP numbers. Two of these spots (No. 7102 and 9001) gave good spectra, but could not be found in any of the databases available. All the other unidentified spots were too faint and gave poor quality spectra. The master map of the neuroblastoma tissue with the relevant histogram graphs of the differentially expressed proteins, as identified by mass spectrometry, is shown in Fig. 6b. The standard spot (SSP) number is displayed beneath each histogram. The bars on the histograms display the maximum spot quantitation for each replicate group, the average spot quantitation and the standard deviation. Each bar represents one replicate group and the order of the bars corresponds to the order in which the replicate groups were made (the left bar corresponds to healthy samples and the right bar corresponds to tumoral ones). Table 3 gives the identification of the 14 spots via mass spectrometry and database searching. Among these proteins, of particular interest in tumour genesis appear to be the down-regulated proteins adrenodoxin (21-folds), carbonic anhydrase III (eightfolds) and aldose reductase related protein I (eightfolds), as well as the up-regulated protein peptidyl-propyl *cis-trans* isomerase A (five-folds). Moreover, new proteins which were absent in control samples (or probably too faint to be detected) were expressed in tumour samples, such as nucleophosmin and stathmin (oncoprotein 18).

Work is now in progress to obtain additional biological samples for identifying more of the very faint spots. In addition, we are extending this work to other tissues (liver, spleen) and we are also comparing the primary tumours with their metastases.

5. Conclusions

PCA is applied here to a data set composed by 20 samples belonging to control and diseased nude mouse adrenal cells. The performance of PCA on the overall data set turned out to be an effective tool for the identification of both classes of samples by means of only the first PC. The analysis of the loadings of PC₁ allowed us to evaluate the spots characterising each class of samples; the separation of the samples in the two classes (healthy and pathological) was also confirmed via the calculation of the SIMCA classification model. Moreover, this classification method, by means of one PC for each class, allowed the identification of the modelling spots and the discriminating ones. This analysis was in good agreement with the conclusions just driven by PCA. The results of PCA and SIMCA analysis were additionally compared to those obtained via dedicated commercial softwares, such as PDQuest, and found to be in very good agreement. Fourteen up- and down-regulated proteins, of relevance in tumour development, could be excised from the gels and identified via mass spectrometry and database searching. The use of 2D gel-electrophoresis coupled to multivariate statistical tools, e.g., PCA and classification methods, is certainly an interesting tool in the field of clinical chemical research, allowing the identification of the proteins responsible for the differences occurring between healthy and diseased samples.

Acknowledgements

P.G.R. is supported by grants from MIUR (Cofin 2003 and FIRB 2001, No. RBNF01KJHT) AIRC and ASI (No. I/R/294/02).

References

- [1] Righetti PG, Stoyanov A, Zhukov M. The proteome revisited: theory and practice of the relevant electrophoretic steps. Amsterdam: Elsevier; 2001.
- [2] Wilkins MR, Williams KL, Appel RD, Hochstrasser DF. Proteome research: new frontiers in functional genomics. Berlin, Germany: Springer-Verlag; 1997.
- [3] Fountoulakis M, Schlaeger EJ. *Electrophoresis* 2003;24: 260–75.
- [4] Kim J, Kim SH, Lee SU, et al. *Electrophoresis* 2002;23: 4142–56.

- [5] Gromov PS, Ostergaard M, Gromova I, et al. *Prog Biophys Mol Biol* 2002;80:3–22.
- [6] Dwek MV, Rawlings SL. *Mol Biotechnol* 2002;22:139–52.
- [7] Kostanjevecki V, Lannoy L, Daniels A, et al. *Neurobiol Aging* 2002;23:1383.
- [8] Castegna A, Aksenov M, Thongboonkerd V, et al. *J Neurochem* 2002;82:1524–32.
- [9] Sickmann A, Marcus K, Schafer H, et al. *Electrophoresis* 2001;22:1669–76.
- [10] Sinha P, Kohl S, Fischer J, et al. *Electrophoresis* 2000;21:3048–57.
- [11] Tsuji T, Shiozaki A, Kohno R, et al. *Neurochem Res* 2002;27:1245–53.
- [12] Westergren-Thorsson G, Malmstrom J, Marko-Varga G. *J Pharm Biomed Anal* 2001;24:815–24.
- [13] Bhatia K, Lord R, Stanton P. *Eur J Cancer* 2002;38:S156.
- [14] Hoffmann F, Kriegl K, Wenk C. *Discrete Appl Math* 1999;93:75–88.
- [15] Lee KH. *Trends Biotechnol* 2001;19:217–22.
- [16] Zhan X, Desiderio DM. *Electrophoresis* 2003;24(11):1834–46.
- [17] Marengo E, Robotti E, Gianotti V, Righetti PG. *Ann Chim (Rome)* 2003;93:105–16.
- [18] Marengo E, Robotti E, Righetti PG, Antonucci F. *J Chromatogr A* 2003;1004:13–28.
- [19] Marengo E, Robotti E, Gianotti V, Righetti PG, Domenici E, Cecconi D. *Electrophoresis* 2003;24:225–36.
- [20] Marengo E, Leardi R, Robotti E, Righetti PG, Antonucci F, Cecconi D. *J Proteome Res* 2003;2(4):351–60.
- [21] Davis JC. *Statistics and data analysis in geology*. Wiley; 1986.
- [22] Brereton RG. *Chemometrics: application of mathematics and statistics to laboratory systems*. Ellis Norwood; 1990.
- [23] Box GEP, Hunter WG, Hunter JS. *Statistics for experimenters*. New York: Wiley; 1978.
- [24] Massart DL, Vandeginste BGM, Deming SN, Michotte Y, Kaufman Y. *Chemometrics: a textbook*. Amsterdam: Elsevier; 1988.
- [25] Leonard RB, Mayer J, Sasser M, Woods ML, Mooney BR, Brinton BG, et al. *J Clin Microbiol* 1995;33:2723–7.
- [26] Johansson ML, Quednau M, Ahrne S, Molin G. *Int J Syst Bacteriol* 1995;45:670–5.
- [27] Couto MMB, Vogels JTWE, Hofstra H, Husiintveld JHJ, Vandervossen JMBM. *J Appl Bacteriol* 1995;79:525–35.
- [28] Boon N, De Windt W, Verstraete W, Top EM. *FEMS Microbiol Ecol* 2002;39:101–12.
- [29] Alike JE, Akenova ME, Fatokun CA. *Genet Resour Crop Evol* 1995;42:393–9.
- [30] Anderson NL, EsquerBlasco R, Richardson F, Foxworthy P, Eacho P. *Toxicol Appl Pharmacol* 1996;137:75–89.
- [31] Vohradsky J. *Electrophoresis* 1997;18:2749–54.
- [32] Kovarova H, Radzioch D, Hajduch M, et al. *Electrophoresis* 1998;19:1325–31.
- [33] Lacroix-Desmazes S, Bayry J, Misra N, Horn MP, Villard S, Pashov A, et al. *N Engl J Med* 2002;34:662–7.
- [34] Dewettinck K, Dierckx S, Eichwalder P, Huyghebaert A. *Lait* 1997;77:77–89.
- [35] Kovarova H, Hajduch M, Korinkova G, Halada P, Krupickova S, Gouldsworthy A, et al. *Electrophoresis* 2000;21:3757–64.
- [36] Ryan TE, Patterson SD. *Trends Biotechnol* 2002;20:S45–51.
- [37] Cacabelos R. *Ann Med* 2002;34:357–79.
- [38] Poirier F, Pontet M, Labas V, et al. *Electrophoresis* 2001;22:1867–77.
- [39] Steiner S, Witzmann FA. *Electrophoresis* 2000;21:2099–104.
- [40] Wold S. *Kem Kemi* 1982;9:401.
- [41] Frank IE, Lanteri S. *Chemometr Intell Lab* 1989;5:247.
- [42] Greene LA, Shain W, Chalazonitis A, Breakfield X, Minna J, Coon HG, et al. *Proc Natl Acad Sci U S A* 1975;72:4923–7.
- [43] Garrells JI. *J Biol Chem* 1979;254:7961–77.

Proteomic analysis of an orthotopic neuroblastoma xenograft animal model[☆]

Natascia Campostrini^a, Jennifer Pascali^a, Mahmoud Hamdan^b, Hubert Astner^a, Danilo Marimpetri^c, Fabio Pastorino^c, Mirco Ponzoni^c, Pier Giorgio Righetti^{a,*}

^a Department of Agricultural and Industrial Biotechnologies, University of Verona, Strada le Grazie 15, 37134 Verona, Italy

^b Computational, Analytical and Structural Sciences, GlaxoSmithKline, Via Fleming 4, 37135 Verona, Italy

^c Laboratory of Oncology, G. Gaslini Children's Hospital, Genoa, Italy

Received 16 March 2004; received in revised form 4 May 2004; accepted 17 May 2004

Available online 17 June 2004

Abstract

Neuroblastoma is the most common extracranial solid tumour of childhood and comprises up to 50% of malignancies among infants. There is a great need of designing novel therapeutic strategies and proteome analysis is one approach for defining markers useful for tumour diagnosis, as well as molecular targets for novel experimental therapies. We started by comparing healthy adrenal glands (which are the election organs developing primary neuroblastoma, NB, tumours) and adrenal glands carrying primary NB tumours, taken from nude mice. Standard maps of healthy and tumour samples were generated by analysis with the PDQuest software. The comparison between such maps showed up- and down-regulation of 84 polypeptide chains, out of a total of 700 spots detected by a fluorescent stain, Sypro Ruby. Spots that were differentially expressed between the two groups, were analysed by MALDI-TOF mass spectrometry and 14 of these spots were identified so far. Among these proteins, of particular interest are the down-regulated proteins adrenodoxin (21-folds), carbonic anhydrase III (eight-folds) and aldose reductase related protein I (eight-folds), as well as the up-regulated protein peptidyl-propyl *cis*–*trans* isomerase A (five-folds). Moreover new proteins, which were absent in control samples, were expressed in tumour samples, such as nucleophosmin (NPM) and stathmin (oncoprotein 18).

© 2004 Elsevier B.V. All rights reserved.

Keywords: Proteomics; Neuroblastoma; Two-dimensional maps

1. Introduction

Neuroblastoma (NB), together with lymphoma, osteosarcoma, Ewing's tumours, rhabdomyosarcoma and lymphoblastic leukemia, belongs to a group of undifferentiated paediatric malignancies known as the small round-cell tumours of childhood. NB is the most common extracranial solid tumour of infancy and childhood. It arises from primitive neuroepithelial cells of the neural crest and occurs in

the adrenal medulla or paraspinal sympathetic ganglia of the abdomen, chest or neck [1]. NB accounts for approximately 9% of all childhood cancers, occurring once out of 8000 live births. This results in an annual incidence of approximately one in 100,000 children less than 15 years of age world-wide. The median age at diagnosis is approximately 22 months with over one-third diagnosed at less than 1 year and over 88% diagnosed by the age of 5.

Despite various therapeutical treatments, including radio-, immune- and chemo-therapy, at elevated doses, with or without bone marrow transplant, childhood cancers caused by neural crest cells, are still characterised by a high percentage of relapse and a high rate of mortality. Starting from these considerations, there is a great need for designing novel therapeutic strategies. The effective treatment of NB, either at advanced stages or at minimal residual disease, remains one of the major challenges in paediatric oncology.

Abbreviations: NB, neuroblastoma; NPM, nucleophosmin; MALDI-TOF, matrix assisted laser desorption/ionisation time of flight

[☆] Presented at the 3rd Meeting of the Spanish Association of Chromatography and Related Techniques and the European Workshop, 3rd Waste Water Cluster, Aguadulce (Almeria), 19–21 November 2003.

* Corresponding author. Tel.: +39 045 8027901; fax: +39 045 8027929.

E-mail address: righetti@sci.univr.it (P.G. Righetti).

Prognosis for patients with this disease has improved with advances in medical care but the overall 5 year survival is still less than 60% [2]. Indeed, the incidence of fatal relapses is still high and long-term survival remains very low [3]. Innovative therapies are thus required and should be focused on the genes and biological pathways that might contribute to malignant transformation or progression of this neoplasia.

To be highly effective, cancer therapy has to be individualised. This requires a detailed understanding of the properties of each tumour with regard to metastasising propensity and drug sensitivity, information that is difficult or even impossible to gain from microscopic analysis of hematoxylin-stained tissue sections. Detailed information regarding mutations in growth-controlling genes, expression of genes controlling growth, metastasis and drug resistance must be obtained. Proteome analysis today, is perhaps one of the most valuable tools for defining markers useful for tumour diagnosis [4]. In addition, proteomics has the potential to unravel basic tumour biological questions regarding mechanisms involved in the pathogenesis of cancer. In this study, we present a first proteomic approach to NB analysis, based on an orthotopic neuroblastoma animal model.

2. Materials and methods

2.1. Chemicals and materials

Tris, mineral oil, DL-dithiothreitol (DTT), Tween 20 and EDTA were purchased from Sigma–Aldrich Chemie GmbH (Steinheim, Germany). Glycerol, methanol, ethanol, acetic acid and acetone were from Merck (Darmstadt, Germany). Forty percent acrylamide/Bis solution, *N,N,N',N'*-tetramethylethylene diamine (TEMED), acrylamide, ammonium persulfate (APS), the Protean IEF

Cell, the GS-710 Densitometer, the Versa Doc Scanner, the software PDQuest Version 6.2 as well as the linear Immobiline dry strips pH gradient 3–10 (17 and 7 cm long), pH gradient 4–7 (7 cm long) were obtained from Bio-Rad Labs (Hercules, CA, USA); glycine, sodium dodecyl sulphate (SDS), iodoacetamide, urea, thiourea, tributylphosphine (TBP) and CHAPS were from Fluka. Bromophenol blue, carrier ampholytes and agarose were purchased from Pharmacia Biotech (Uppsala, Sweden).

2.2. Orthotopic neuroblastoma animal model

Five-week-old female nude (nu/nu) mice were purchased from Harlan Laboratories (Harlan Italy-S. Pietro al Natisone, UD). Mice were anaesthetised and injected with 5×10^4 cultured murine NXS2 neuroblastoma cells [5] (kindly provided by Dr. Reisfield, Scripps Clinic, La Jolla, CA, USA) in 20 μ L of HEPES buffer, after laparotomy, in the capsule of the left adrenal gland. The lethality of the method was 0%. Mice were monitored at least two times weekly for evidence of tumour development, quantification of tumour size, and evidence of tumour-associated morbidity. Mice were sacrificed and organs, after washing in PBS, were frozen and stored in liquid nitrogen. All experiments involving animals have been reviewed and approved by the licensing and ethical committee of the National Cancer Research Institute and by the Italian Ministry of Health.

Four samples, divided in two classes, constituted the dataset:

- Four samples from healthy adrenal nude mouse glands (the pool of which was used to perform four 2D-maps; Fig. 1a);
- Four samples from adrenal nude mouse glands carrying primary Neuroblastoma tumours (the pool of which was used to perform four 2D-maps; Fig. 1b).

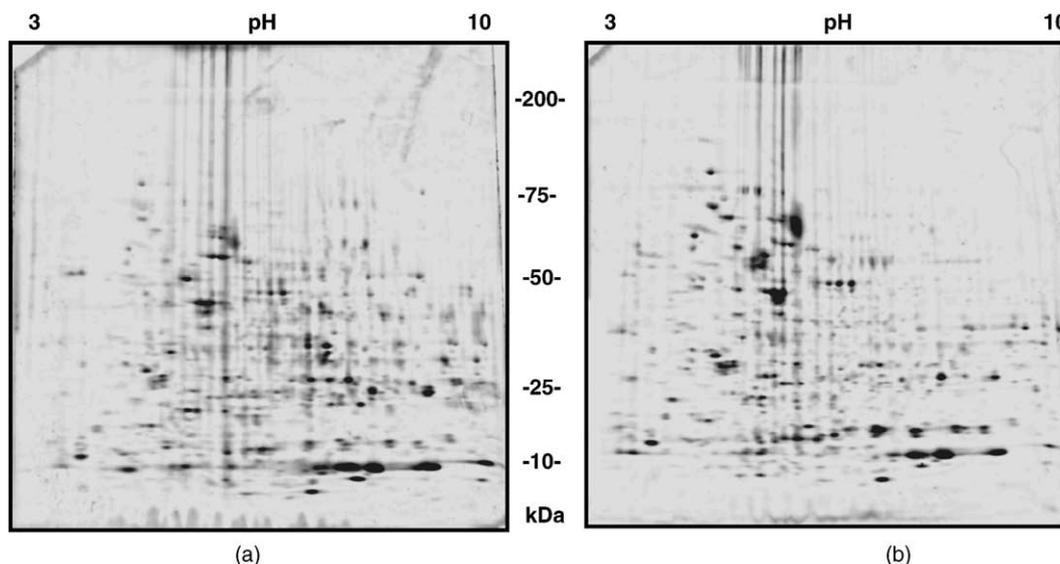


Fig. 1. Examples of the obtained 2D-PAGE maps: control (a) and tumour (b) samples.

The procedure was repeated three times, thus producing three genuine replicates for each map.

2.3. Two-dimensional gel electrophoresis

2.3.1. Sample treatment

Healthy and tumour adrenal glands from nude mice were stored at -80°C . All samples were homogenised (5% homogenate) with a lysis solution containing 7 M urea, 2 M thiourea, 3% CHAPS, 40 mM Tris, 5 mM TBP, 0.5% carrier ampholytes pH 3–10, 5 mM EDTA, 1 mM PMSF and 50 U/mL DNase. After 1 h lysis, samples were centrifuged at 4°C at 6000 rpm for eliminating all residual particles. The collected supernatant was alkylated for 1 h with 10 mM acrylamide (1 M stock solution) and the reaction was stopped with 10 mM DTT (from 1 M stock solution) [6]. Interfering substances (lipids, salts) were removed by a precipitation step. A cold mix of acetone and methanol (8:1) was added; after 2 h at -20°C , the solution was centrifuged at 13,000 rpm for 30 min. Pellets were finally resuspended in sample solution containing urea, thiourea, CHAPS and Tris.

2.3.2. IEF in IPG strips

The first dimension run was performed on strips (17 cm length, 0.5 mm thickness) with a linear pH gradient from pH 3 to 10 [7]. The IPG strips were rehydrated with 400 μg of pooled samples (four strips with pooled healthy samples and four strips with pooled tumour samples) and containing traces of bromophenol blue for monitoring the electrophoretic run. The passive gel rehydration was allowed to continue for 8 h before the focusing step. Isoelectric focusing (IEF) was carried out with a Protean IEF Cell (Bio-Rad), with a low initial voltage and then by applying a voltage gradient up to 10,000 V with a limiting current of 50 μA /strip. The total product time \times voltage applied was 75,000 Vh for each strip and the temperature was set at 20°C .

2.3.3. Interfacing the IPG strips with the denaturing SDS solution and SDS-PAGE

Each strip was equilibrated with a SDS denaturing solution containing 6 M urea, 2% SDS, 20% glycerol, 0.375 M Tris-HCl (pH 8.8). The contact lasted for 30 min in tubes containing 20 mL each of the equilibration solution. Each strip was then interfaced with a gel slab using 0.5% agarose solubilised in the cathodic buffer (192 mM glycine, 0.1% SDS, Tris to pH 8.3). All the gel slabs were cast with a two-vessel gradient mixer, with total sufficient volumes for polymerising eight gel slabs of 1.5 mm in thickness and with a porosity gradient from 7 to 20%T. Polymerization took place overnight. The second dimension run was performed by using a PROTEAN II x1 Multi-Cell (Bio-Rad). The cathodic buffer was the same as in previous step; the anodic buffer was a solution of 0.375 M Tris-HCl, pH 8.8. The electrophoretic run was performed by setting a current of 2 mA for each gel for 1 h, then 5 mA/gel for 2 h and 10 mA/gel un-

til the end of the run. During the whole run the temperature was set at 11°C .

2.4. Fluorescent staining

All gels were fixed with a solution containing 40% ethanol and 10% acetic acid for 30 min. After this step the gels were stained with Sypro Ruby (70 mL for each gel) overnight and then destained with 10% methanol, 7% acetic acid for 1 h. The gels were finally washed with milliQ water. All the samples were then digitised with the Versa Doc Scanner.

2.5. Protein pattern and statistical analysis

The digitised images were acquired with the software PDQuest (version 6.2), which was used for cropping and orienting the images, for detecting and identifying spots, for comparing and matching spots, for normalising and analysing the data and for preparing a report. A match set was created from the protein patterns of the two independent cellular extracts (healthy mouse adrenal gland, mouse adrenal gland carrying NB primary tumours). A standard gel was generated out of the image with the highest spot number. Spot quantities of all gels were normalised by removing non expression-related variations in spot intensity; for that, the raw quantity of each spot in a gel was divided by the total quantity of all the spots in that gel and included in the standard. The final synthetic image was a Gaussian scan image that contained all the Gaussian spots with a defined volume and quality. All subsequent spot matching and analysis steps in the PDQuest software were performed on Gaussian spots. The results were evaluated in terms of spot optical density (OD). Statistical analysis (Student's *t*-test) via PDQuest allowed the study of proteins that were significantly increased or decreased in pathological samples.

2.6. Protein identification by mass spectrometry

2.6.1. In situ digestion and extraction of peptides and MALDI-TOF analysis

The spots of interest were carefully excised from the gel with a razor blade and placed in Eppendorf tubes. The gel pieces were washed twice with a solution of acetonitrile/Tris 5 mM pH 8.5 (50/50) followed by a single wash with only Tris 5 mM pH 8.5. These pieces were dehydrated in Speed-vac device at room temperature and covered with 15 μL of Trypsin (0.02 mg/mL) in NH_4HCO_3 buffer (40 mM, pH 8.5) and left at 37°C overnight. The peptides were extracted two times in 50 μL of acetonitrile/ H_2O 1% (v/v) formic acid (50/50). The extraction was conducted in an ultrasonic bath for 15 min each time. After sonication, the excess of acetonitrile was evaporated and the peptides were resuspended in 10 μL of 0.1% TFA, concentrated and cleaned using ZipTip microcolumns (C18).

The extracted peptides were loaded onto the MALDI target plate by mixing 1 μL of each solution with the same

volume of a matrix solution, prepared fresh every day by dissolving 10 mg/mL cyano-4-hydroxycinnamic acid in acetonitrile/ethanol (1:1, v/v), and allowed to dry. Measurements were performed using a TofSpec 2E MALDI-TOF instrument (Micromass, Manchester, UK), operated in the reflectron mode, with an accelerating voltage of 20 kV [8]. The laser wavelength was 337 nm and the laser repetition rate was 4 Hz. The final mass spectra were produced by averaging 50–200 laser shots. Peptide masses were searched against SWISS-PROT, TrEMBLE and NCBI nr databases by utilizing the ProteinLynx program from Micromass, ProFound from Prowl and Mascot from Matrix Science.

2.7. Immunoblot analysis

Proteins derived from healthy and tumour adrenal glands, previously separated by 2D-PAGE, were transferred onto PVDF membranes (Immuno-Blot™ PVDF membrane, Bio-Rad) for 2 h at 60 V, using a Mini Trans-Blot system (Bio-Rad). Membranes were blocked with 3% BSA in TBST (10 mM Tris, 150 mM NaCl, 0.1% Tween 20, pH 7.5) overnight at 4 °C and then incubated for 1 h at room temperature with the anti-OP18/stathmin antibody (1:5000,

Sigma), recognising the C-terminus of this protein. Blots were developed with an enhanced chemiluminescence system (ECL plus, Amersham) and stathmin visualised on an autoradiography film (Hyperfilm, Amersham). Films were scanned using a densitometer (GS-710, Bio-Rad).

3. Results and discussion

3.1. Biological relevance of the orthotopic neuroblastoma xenograft model

Previous *in vivo* experiments, by using directly an *i.v.* administration of NB cells in nude mice, have proved the importance of choosing a specific xenograft animal model that mimics the metastatic spread observed in advanced-stage human NB patients [9,10]. However, a further and more realistic view of a comparative study could be obtained if a tumour model were available that better reflected the growth of advanced NB in children (*i.e.* large adrenal gland tumours and multiple small metastatic lesions). All current data support this concept and recommend that orthotopic implantation of tumour cells in recipient animals is mandatory for

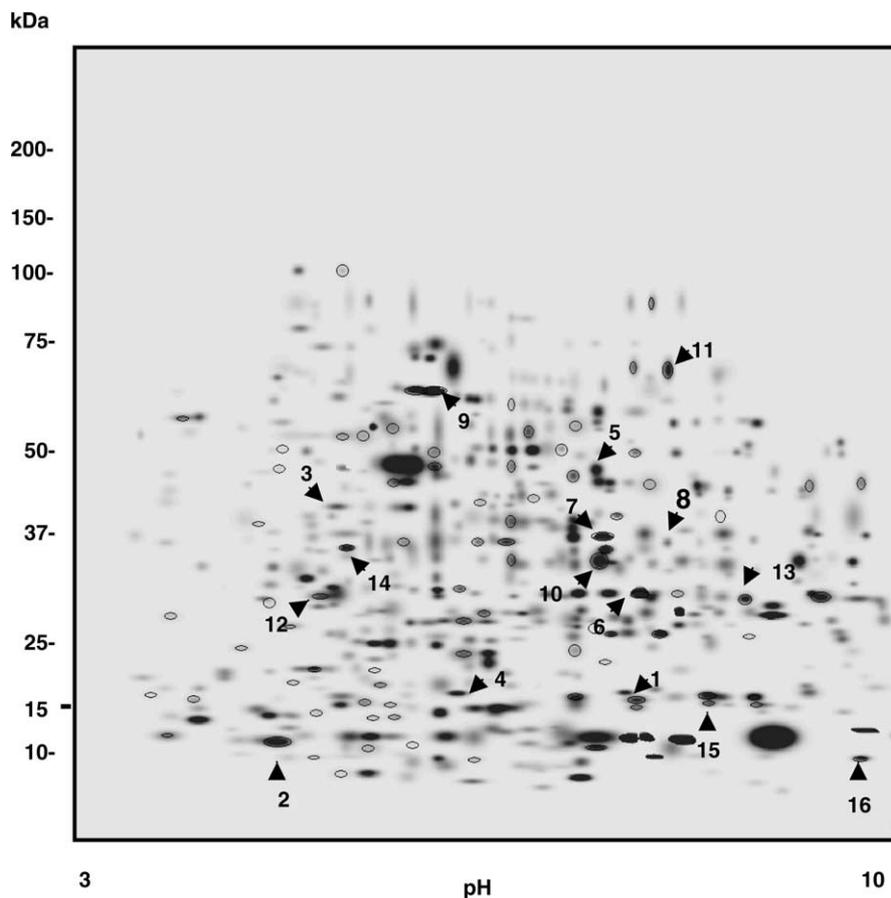


Fig. 2. Master map of a neuroblastoma tissue excised from the adrenal glands of nude mice. The 84 spots differently expressed are circled and the identified spots are numbered. The circled numbered spots refer to the 14 spots which could be identified by mass spectrometry upon excision (two of them, spots 15 and 16, gave good spectra but could not be found in any of the data bases available).

studies of tumour progression, angiogenesis, invasion, and metastasis [11]. In order to provide a well characterised, relevant, highly reproducible, angiogenic, and metastatic orthotopic model of NB, we initiated studies to define an adrenal NB xenograft model. We decided to use the intra-adrenal injection of NXS2 murine NB cells in mice because, 2 weeks after injection, adrenal gland tumours were always found in all animals. This model best reflected the typical growth pattern of human NB, since orthotopic injection of NXS2 cells resulted in solid adrenal tumours that were highly vascular, locally invasive into surrounding tissues, and metastatic to distant sites. Indeed, macroscopic metastases always occurred after 3–4 weeks of injection in the ovary and spleen, while micrometastases were apparent frequently in the contralateral adrenal gland, kidneys, liver, bone marrow and lung.

3.2. 2D map analysis

Fig. 2 shows the master map of the neuroblastoma developed in nude mice: about 750 spots could be counted upon Sypro Ruby staining. They seem to be quite evenly distributed in the pH 3–10 interval, with a Mr distribution from ca. 9000 to 140,000 Da. The circles mark 84 spots found to be differently expressed in pathological samples, of which 39 were up-regulated and 45 down-regulated. Moreover, a number of protein spots that were newly expressed and newly silenced in the tumour samples as compared to healthy ones could also be detected. Additionally, the numbers refer to the 14 spots which could be identified by MS analysis (an additional two, numbers 15 and 16, although producing good quality spectra, could not be matched to any of the databases available). Fig. 3 shows a small gel area with a few spots up- and down-regulated and newly expressed in pathological versus control specimens. Table 1 lists the protein spots that could be identified so far. In general, there was a good agreement between the theoretically-predicted and experimentally found pI and Mr values, except for a large deviation in the case of adrenodoxin, suggesting that this protein might be extensively modified at the post-translational level in this kind of tissue, or be a pre-protein, as discussed below. Adrenodoxin, in fact, (ADX) is a small iron–sulfur protein present in the mitochondrial matrix, where it transfers electrons from adrenodoxin reductase to mitochondrial forms of cytochrome P-450. The adrenal cortex is one of the more abundant sources of adrenodoxin, since it interacts with the three mitochondrial, steroidogenic P-450s. Adrenodoxin is also expressed in other steroidogenic tissues such as testis, ovary and placenta as well as in nonsteroidogenic tissues where mitochondrial P-450 forms are active including liver, kidney and brain. Adrenodoxin is encoded as larger precursors of 188 amino acids, and the signal peptide is cleaved during transfer into mitochondria (64 amino acids). The Mr of the mature adrenodoxin chain (124 amino acids) is 13,617 Da: this is in agreement with our experimental data. The mouse adrenodoxin pre-protein is 93, 74, 70 and

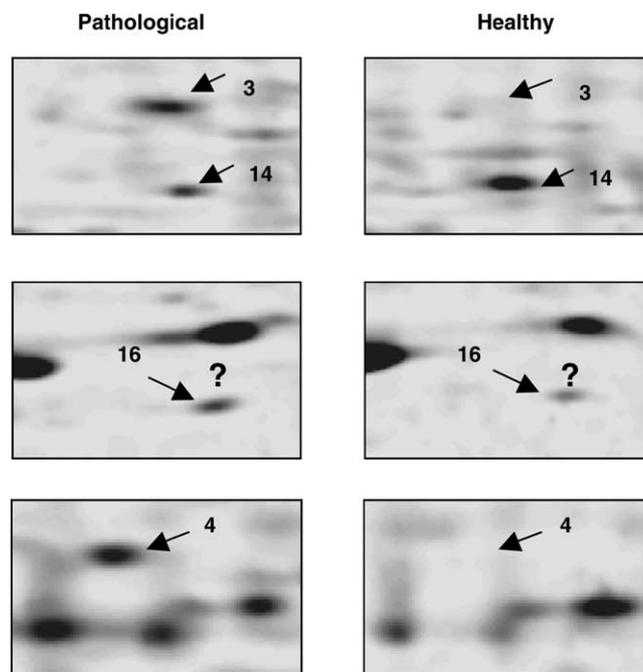


Fig. 3. Comparison of two-dimensional gel patterns of some proteins in neuroblastoma (left panel) and healthy (right panel) tissues. The corresponding spot number are shown in Table 1. The spot with a question mark corresponds to a two-fold up-regulated protein which could not be identified in any of the databases available, although it gave good MS spectra.

69% identical to the rat, human, bovine and porcine adrenodoxin pre-proteins, respectively. The mature adrenodoxin is 98% identical to rat adrenodoxin and 90% identical to the human, bovine and porcine adrenodoxin [12].

In order to further confirm the data obtained with the master map, we performed an immunoblot analysis on stathmin, a protein which, together with nucleophosmin (NPM), appears to be expressed only on tumour samples (Fig. 3). As shown in Fig. 4, with this method low levels of this stathmin can also be detected in normal murine tissue (right panel), whereas in the tumour sample (left panel) the same spot appears to be over-expressed by a factor of at least >15 times. Additionally, when the same experiment was repeated in a shallower pH gradient (pH 4–7) this very intense spot (which appeared oblong in the wide pH range) could be resolved into a well-separated string of minor spots, more acidic than the parental protein. This experiment suggests two important points, which might have escaped detection up to the present. First of all, it might not be true that a gene is fully silenced or newly turned-on in pathological tissues (or during ontogenesis or any other biologically-relevant phenomenon, such as differentiation, cell-cycle arrest, apoptosis, cell modification induced by drug treatments) as compared to the control, healthy one. Indeed, extremely low levels might be present in the control samples, and vice versa. What our experiments suggest is that, in the control samples, the levels of stathmin were so minute as to be below the sensitivity limit of Sypro Ruby staining. With immuno-detection (the most

Table 1
Identified proteins from the neuroblastoma tissue

	Experimental Mr (Da)	Experimental pI	Theoretical Mr	Theoretical pI	Z-score	MOWSE-score	Protein name	Accession number	Coverage (%)	Number of peptides	Variation
1	~15000	7.8	17,840	8.3	2.39	1.93E8	Peptidyl-prolyl <i>cis</i> - <i>trans</i> isomeraseA	P17742	59.5	13	Increased 5
2	~13000	5.0	20,123	5.5	1.84	1.60E4	Adrenodoxin	P46656	25	8	Decreased 21
3	~40000	4.5	32,560	4.6	2.38	2.51E5	Nucleophosmin (NPM)	Q61937	31	8	Only tumoral samples
4	~17000	6.5	17,143	6.0	2.33	3.78E7	Stathmin (OP18)	P54227	64.9	17	Only tumoral samples
5	~50000	7.5	46,661	6.9	2.31	4.46E8	Isocitrate dehydrogenase	O88844	33.3	15	Only control samples
6	~30000	7.8	29,366	7.4	1.67	1.78E5	Carbonic anhydrase III	P14141	28	8	Decreased 8
7	~37000	7.5	35,857	7.2	2.29	1.77E11	Aldose reductase related protein1	P21300	59.3	18	Decreased 8
8	~35000	8.0	33,335	8.2	2.1	2.68E7	Thiosulfate sulfurtransferase	P52196	39	13	Only control samples
9	~65000	6.5	60,956	6.1	2.37	2.61E21	60 kDa heat shock protein	P19226	57	37	Decreased 4
10	~34000	7.5	34,973	8.6	2.38		Electron tranfer flavoprotein	NCBI gi21704230	41	12	Decreased 4
11	~72000	8.1	67,631	7.7	2.41	2.09E9	Transketolase	P40142	28	18	Decreased 2
12	~30000	5.1	27,771	4.8	2.06	2.57E6	14-3-3 protein zeta/delta (protein kinase C inhibitor protein-1)	P35215	42	9	Increased 2
13	~27000	8.9	31,517	8.6	2.28	6.78E6	Enoyl-CoA hydrase	P14604	34	10	Decreased 4
14	~36000	5.3	35,753	4.8	2.38	2.70E13	Annexin V	P48036	64	22	Decreased 2
15	~15000	8.5	?	?	?	?	?	?	?	?	Increased 3
16	~8500	9.8	?	?	?	?	?	?	?	?	Increased 2

The comparison is performed between the matched spots in healthy samples and the matched spots in tumoural ones. A threshold value of 2.0 (which corresponds to a variation of 100%) was chosen as a meaningful variation in the comparison of tumour proteins vs. healthy proteins. The number, written in the last column, means folds of change.

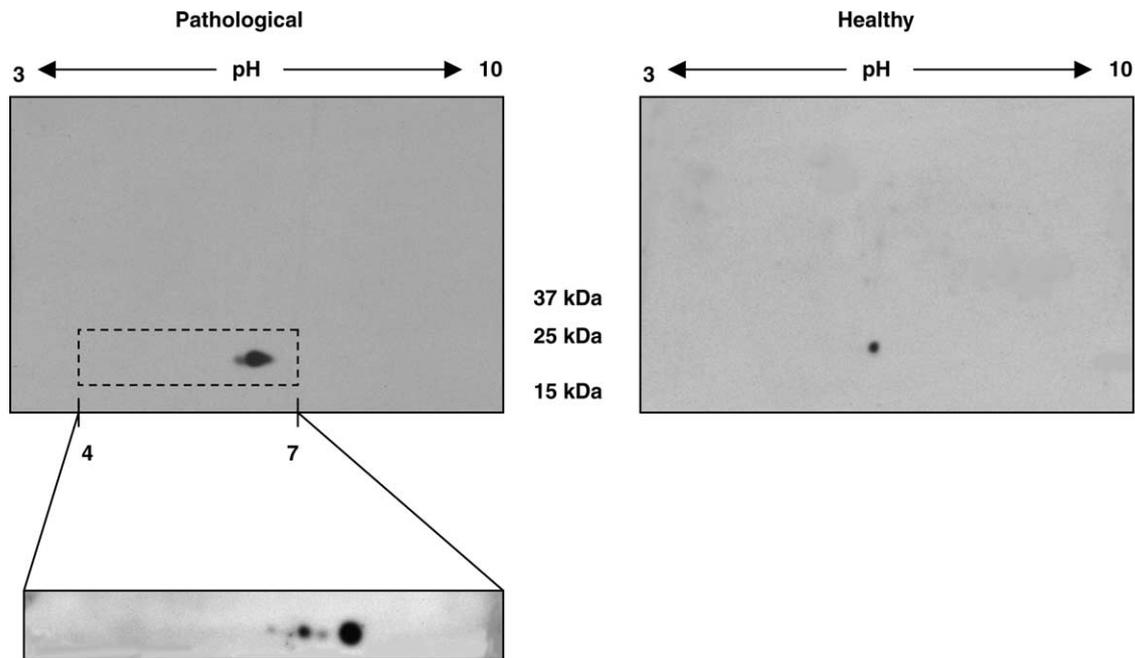


Fig. 4. Immunoblots of 2D-PAGE slabs from normal murine samples (right panel) and tumour tissues (left panel), developed by using an anti-OP18/stathmin antibody, recognising the C-terminus of this protein. Lower strip on the left: same boxed area of the upper gel, but run in a shallower (pH 4–7) IPG gradient. Note the string of spots to the left of the parental spot, suggesting a phosphorylated train.

sensitive detection method today available) detectable levels of stathmin could indeed be found also in control tissue (Sypro Ruby has a sensitivity at least three orders of magnitude lower than our chemiluminescence detection method on the immunoblot) [13]. The other very important point, is that only with such a sensitive method we could observe that, in tumour tissues, stathmin seems to be modified, possibly at the post-synthetic level, the *pI* variations in the string of minor spots suggesting that there could be phosphorylation events.

3.3. On the significance of nucleophosmin and stathmin up-regulation

Our group is now extensively involved in the proteomic analysis of different human tumours, such as mantle cell lymphomas [14,15] and pancreatic ductal carcinomas [16]. In both cases we found strong variations of these two proteins, whose significance will be discussed below. Nucleophosmin, a three-fold down-regulated protein, appears to be particularly important in oncogenesis. NPM is a ubiquitously expressed nuclear phosphoprotein that continuously shuttles between the nucleus and cytoplasm. One of its suggested roles is in ribosomal protein assembly and transport and also as a molecular chaperone that prevents proteins from aggregating. Evidence is accumulating that the NPM gene is involved in several tumour-associated chromosome translocations and in the oncogenic conversion of various associated proteins [17–19]. NPM appears to be present in most human tissues, with especially robust expression in pancreas and testis and lowest expression in lung [20]. In-

terestingly, a fusion protein, containing the amino-terminal 117 amino acid portion of NPM, joined to the entire cytoplasmic portion of the receptor tyrosine kinase anaplastic lymphoma kinase (ALK) has been found to be involved in oncogenesis in the case of non-Hodgkin's lymphoma [21]. Stathmin (oncoprotein 18, OC18) is a p53-regulated member of a novel class of microtubule-destabilising proteins known to promote microtubule depolymerisation during interphase and late mitosis [22–25]. Noteworthy, over-expression of stathmin, by inhibiting polymerisation of microtubules, permits increased binding to these structures of vinblastin, a well known chemotherapeutic agent, during treatment of human breast cancer [22].

4. Concluding remarks

To our knowledge, this is the first attempt, at a proteome level, of analysis of the paediatric solid tumour neuroblastoma. At present, we were forced to adopt a clinically relevant animal model, by injecting murine NB tumour cells in the capsule of the left adrenal gland of nude mice, due to the fact that human studies are hampered by the difficulty to procure proper, healthy controls. Although preliminary in nature, these studies will be quite valuable in view of extending them to human cases. At the Gaslini hospital, in fact, we have a rather large collection of NB tumour biopsies taken after surgery; as soon as control tissues will be available, the present data could be easily extended to human tumours, in view of finding the possible mechanisms, at the proteome, transcriptome and genome levels, responsible for

their development. Future studies will thus be addressed to human cases as well as to the correlation between regulation of protein expression and mRNA levels present in the cells.

Acknowledgements

Supported by grants from Associazione Italiana per la Ricerca sul Cancro (AIRC) and from Fondazione Italiana per la Lotta al Neuroblastoma. PGR is additionally supported by grants from MIUR (Cofin 2003 and FIRB 2001, N° RBNF01KJHT) and by the European Community (grant no. QLG2-CT-2001-01903).

References

- [1] J.M. Maris, K.K. Matthay, *J. Clin. Oncol.* 17 (1999) 2264.
- [2] D. Niethammer, R. Handgretinger, *Eur. J. Cancer* 31A (1995) 568.
- [3] R. Ladenstein, T. Philip, C. Lasset, O. Hartmann, A. Garaventa, R. Pinkerton, J. Michon, J. Pritchard, T. Klingebiel, B. Kremens, A. Pearson, C. Coze, P. Paolucci, D. Frappaz, H. Gadner, F. Chauvin, *J. Clin. Oncol.* 16 (1998) 953.
- [4] P.G. Righetti, A. Stoyanov, M. Zhukov, *The Proteome Revisited: Theory and Practice of All Relevant Electrophoretic Steps*, Elsevier Press, Amsterdam, 2001.
- [5] L.A. Greene, W. Shain, A. Chalazonitis, X. Breakfield, J. Minna, H.G. Coon, M. Nirenberg, *Proc. Natl. Acad. Sci. U.S.A.* 72 (1975) 4923.
- [6] B. Herbert, M. Galvani, M. Hamdan, E. Olivieri, J. McCarthy, S. Pedersen, P.G. Righetti, *Electrophoresis* 22 (2001) 2046.
- [7] P.G. Righetti, *Immobilized pH Gradients: Theory and Methodology*, Elsevier Press, Amsterdam, 1990.
- [8] H. Hamdan, P.G. Righetti, *Mass Spectrom. Rev.* 21 (2002) 287.
- [9] F. Pastorino, C. Brignole, D. Marimpietri, P. Saprà, E.H. Moase, T.M. Allen, M. Ponzoni, *Cancer Res.* 63 (2003) 86.
- [10] L. Raffaghello, G. Pagnan, F. Pastorino, E. Cosimo, C. Brignole, D. Marimpietri, P.G. Montaldo, C. Gambini, T.M. Allen, E. Bogenmann, M. Ponzoni, *Int. J. Cancer* 104 (2003) 559.
- [11] A.F. Chambers, A.C. Groom, I.C. MacDonald, *Nat. Rev. Cancer* 2 (2002) 563.
- [12] M. Stromstedt, M.R. Waterman, *Biochim. Biophys. Acta* 1261 (1995) 126.
- [13] W.F. Patton, *BioTechniques* 28 (2000) 944.
- [14] F. Antonucci, M. Chilosi, M. Santacatterina, B. Herbert, P.G. Righetti, *Electrophoresis* 23 (2002) 356.
- [15] F. Antonucci, M. Chilosi, C. Parolini, M. Hamdan, H. Astner, P.G. Righetti, *Electrophoresis* 24 (2003) 2376.
- [16] D. Cecconi, A. Scarpa, M. Donadelli, M. Palmieri, M. Hamdan, H. Astner, P.G. Righetti, *Electrophoresis* 24 (2003) 1871.
- [17] E. Colombo, J.C. Marine, D. Danovi, B. Falini, P.G. Pelicci, *Nat. Cell Biol.* 4 (2002) 529.
- [18] M. Okuda, *Oncogene* 21 (2002) 6170.
- [19] C. Yang, D.A. Maiguel, F. Carrier, *Nucleic Acid Res.* 30 (2002) 2251.
- [20] G.M. Shackleford, A. Ganguly, C.A. MacArthur, *BMC Genomics* 2 (2001) 8.
- [21] D. Bischof, K. Pulford, D.Y. Mason, S.W. Morris, *Mol. Cell. Biol.* 17 (1997) 2312.
- [22] E. Alli, J. Vash-Babula, J.M. Yang, W.N. Hait, *Cancer Res.* 62 (2002) 6864.
- [23] C.L. Chang, N. Hora, N. Huberman, R. Hinderer, M. Kukuruga, S.M. Hanash, *Proteomics* 1 (2001) 1415.
- [24] G. Chen, H. Wang, T.G. Gharib, C.C. Huang, D.G. Thomas, K.A. Shedden, R. Kuick, J.M. Taylor, S.L. Kardia, D.E. Misek, T.J. Giordano, M.D. Iannettoni, M.B. Orringer, S.M. Hanash, D.G. Beer, *Mol. Cell Proteomics* 2 (2003) 107.
- [25] N. Melhem, N. Hailat, R. Kuick, S.M. Hanash, *Leukemia* 11 (1997) 1690.

Review

Critical survey of quantitative proteomics in two-dimensional electrophoretic approaches

Pier Giorgio Righetti^{a,*}, Annalisa Castagna^a, Francesca Antonucci^a,
Chiara Piubelli^a, Daniela Cecconi^a, Natascia Campostrini^a, Paolo Antonioli^a,
Hubert Astner^b, Mahmoud Hamdan^b

^a Department of Agricultural and Industrial Biotechnologies, University of Verona, Strada Le Grazie 15, Verona 37134, Italy

^b Computational, Analytical and Structural Sciences, GlaxoSmithKline, Via Fleming 4, 37135 Verona, Italy

Available online 31 July 2004

Abstract

The present review attempts to cover a number of methods that appeared in the last few years for performing quantitative proteome analysis. However, due to the large number of methods described for both electrophoretic and chromatographic approaches, we have limited this excursus only to conventional two-dimensional (2D) map analysis, coupling orthogonally a charge-based step (isoelectric focusing) to a size-based separation (sodium dodecyl sulfate (SDS)-electrophoresis). The first and oldest method applied in 2D mapping is based on statistical analysis performed on sets of gels via powerful software packages, such as the Melanie, PDQuest, Z3 and Z4000, Phoretix and Progenesis. This method calls for separately-running a number of replicas for control and treated samples, the merging and comparing between these two sets of data being accomplished via the softwares just mentioned. Recent developments permit analyses on a single gel containing mixed samples differentially labelled and resolved by either fluorescence or isotopic means. In one approach, a set of fluorophors, called Cy3 and Cy5, are selected for differentially tagging Lys residues, via a “minimal labelling” protocol. A variant of this, adopts a newer set of fluorophors, also of the Cy3 and Cy5 type, reacting on Cys residues, via a strategy of “saturation labelling”. There are at present two methods for quantitative proteomics in a 2D gel format exploiting stable isotopes: one utilizes tagging Cys residues with [²H₀]/[²H₃]-acrylamide; the other one, also based on a Cys reactive compound, exploits [²H₀]/[²H₄] 2-vinylpyridine. The latter reagent achieves 100% efficiency coupled to 100% specificity. The advantages and limitations of the various protocols are discussed.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Reviews; Proteomics; Isotope labelling; Acrylamide; Vinylpyridine; Dyes

Contents

1. Introduction	4
2. Statistical analysis of separately-run two-dimensional maps	5
3. Differential, in-gel electrophoresis based on Lys tagging	7
4. Differential, in-gel electrophoresis based on Cys tagging	10
5. Isotope-coded two-dimensional maps: [² H ₀]/[² H ₃] acrylamide	12
6. Isotope-coded two-dimensional maps: [² H ₀]/[² H ₄] 2-vinylpyridine	16
7. Conclusions	16
Acknowledgements	16
References	16

* Corresponding author. Tel.: +39 045 8027901; fax: +39 045 8027929.

E-mail address: righetti@sci.univr.it (P.G. Righetti).

1. Introduction

A major goal of proteomics is the qualitative and quantitative analysis of all the proteins expressed in an organism, a tissue, a cell, an organelle, or even a body fluid, determined quantitatively at a certain moment and under a precise condition [1]. Changes in protein expression owing to stimulus or conditioning are measured in a systematic manner, and are used for elucidating mechanisms of cell function and signalling. The strength of proteomics is that a “shot gun” approach, requiring no prior knowledge of the system under investigation, is often used and does not assume a model prior to data collection. Therefore, proteomics provides the ability to deal with the complexity of biological systems with minimal experimental bias. Such a complexity arises from the numerous parallel signalling pathways that interact with each other. The ability to monitor many proteins simultaneously yields a global view of protein expression and post-translational modification, which is much more informative than monitoring a few proteins [2]. An important area of application for proteome analysis is the recognition of proteins that are correlated with a certain state: the desired assessment is a comparison between two samples. For this purpose, the protein patterns of, e.g., healthy and pathological, of drug-treated and untreated cells, tissues, or body fluids are compared. The presence or altered levels of specific proteins can be biomarkers of disease, either individually or as a signature of multiple proteins. Thus, the comparison of treated versus untreated samples, and the detection of differences in protein expression therefrom, can provide unique markers of biological activity. Additionally, such differences can point to mechanisms of action, or they can be used for predicting or understanding drug toxicity, or a number of other relevant biological/pharmacological phenomena. In protein analysis, consideration must be given to the fact that the number of proteins expressed at any one time in a given cellular system is in the thousands or tens of thousands. Thus, a proteomic technology would consist of a combination of the following features:

- (1) High-throughput.
- (2) The ability to recognize differentially expressed proteins.
- (3) The ability to quantitatively display and analyse all the proteins present in a sample.

Quantitative proteomics is becoming particularly interesting in the field of medicine, due in large part to the prospects that a proteomic approach to disease investigations will overcome some of the limitations of routes based largely, up to recent times, in screening of gene defects. As correctly pointed out by Storchman [3,4], “only 2% of our total disease load is related to monogenic causality, and even here the final phenotype is modulated by many factors”, a statement highlighting the primary role of expressed proteins in disease processes and evolution. Up to the present, in fact, the primary

technology platform, for screening for a variety of pathological states, has been the gene expression micro-array (GEM), a spotted grid of up to 30 000 oligonucleotides or cDNAs representing expressed genes [5]. Although GEMs have allowed researchers to generate a huge amount of mRNA expression data for many cancer types, there are a number of disadvantages in the interpretation of purely transcriptomics data that would preclude the identification of all-tumour associated changes. Firstly, there is a poor correlation between transcript and disease-associated protein levels, due to different kinetics of protein translation and turnover, in the cell environment, for different polypeptide chains [6]. Secondly, the disease state may be brought about by a translocation of a protein within the cell rather than simply differential levels of mRNA [7]. Thirdly, current transcriptomic analyses provide only limited information on alternative splicing and none on post-translational modifications. The protein content is more dynamic than the transcriptome, conferring reactive and compensatory functions that do not rely on the relatively slow process of transcriptional activation.

Given the above shortcomings, proteomic analysis appears thus to be a most useful tool in biomedicine [8], as well as the identification of therapeutic targets and development, e.g., of new anticancer strategies and remedies to a host of diseases [9]. The opportunities as well as the challenges facing disease proteomics are formidable. Particularly promising areas of research include:

- (1) Delineation of altered protein expression; not only at the whole cell or tissue levels, but also in subcellular structures, in protein complexes and in biological fluids.
- (2) The development of novel biomarkers for diagnosis and early detection of disease.
- (3) The identification of new targets for therapeutics and the potential of accelerating drug development through more effective strategies for evaluating therapeutic effects and toxicity.

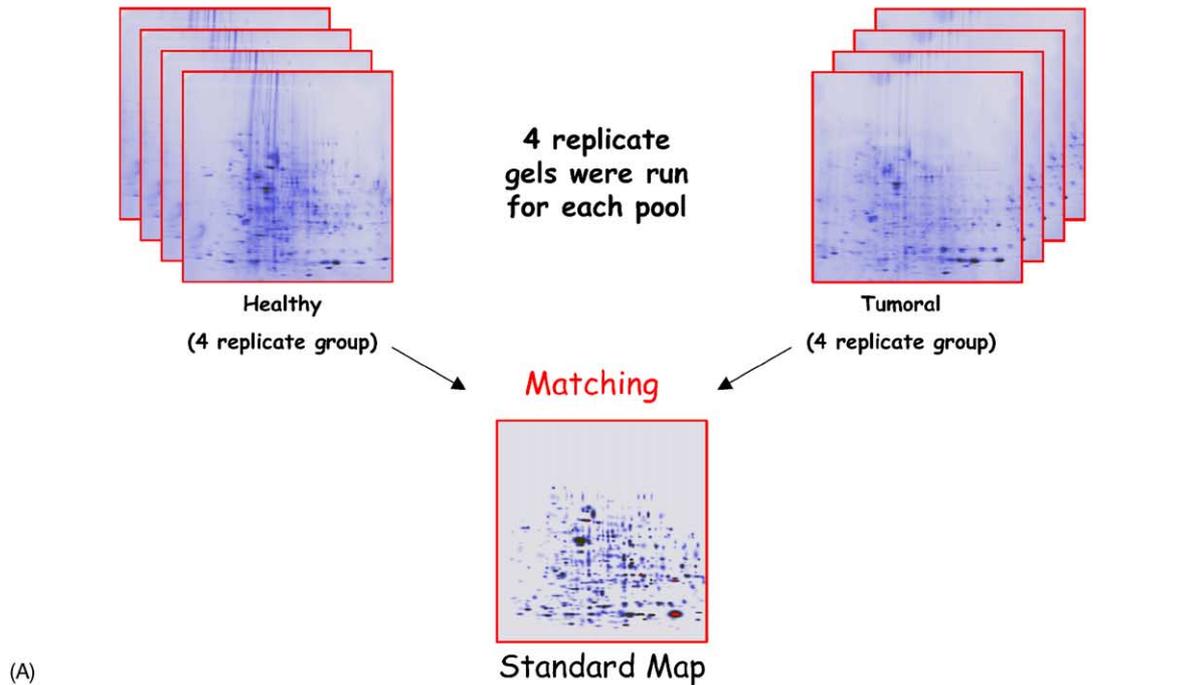
There has been a sudden burst, in the last few years, of methods describing novel approaches to quantitative proteomics. Quite a few of them have been reviewed in a number of papers dedicated to these topics [10–13]. Such methods comprise not only differential proteomics in the well-ingrained two-dimensional (2D) map analysis, but also a host of approaches developed in purpose for 2D chromatography processes. There are two fundamental distinctions between the two methodologies: whereas, in the 2D map protocol, the sample is analyzed as intact species, i.e., as synthesized by the organism under analysis, in 2D chromatography the sample is in general tagged after having been digested into a mixture of peptides. Given the two quite different protocols, we will restrict this review only to electrophoretic methodologies and give figures of merits to the various approaches.

2. Statistical analysis of separately-run two-dimensional maps

Comparison of 2D maps, separately-run, by powerful softwares (similar to those used by astronomers for mapping stars in a given portion of the night sky; in fact, one

of the first, embryonic programs developed was nicknamed Tycho, in honour of Tycho Brae, a famous Danish astronomer of the 17th century [14]) is one of the oldest and most popular methods in the electrophoretic approach to proteome analysis. The sequence of panels in Fig. 1 gives an example of such a procedure. It refers to neuroblastomas,

Differential Analysis on Neuroblastoma samples:
 comparison between healthy and tumoral samples
 (gels matching and analysis by PDQuest software)



Up-regulated proteins in Neuroblastoma

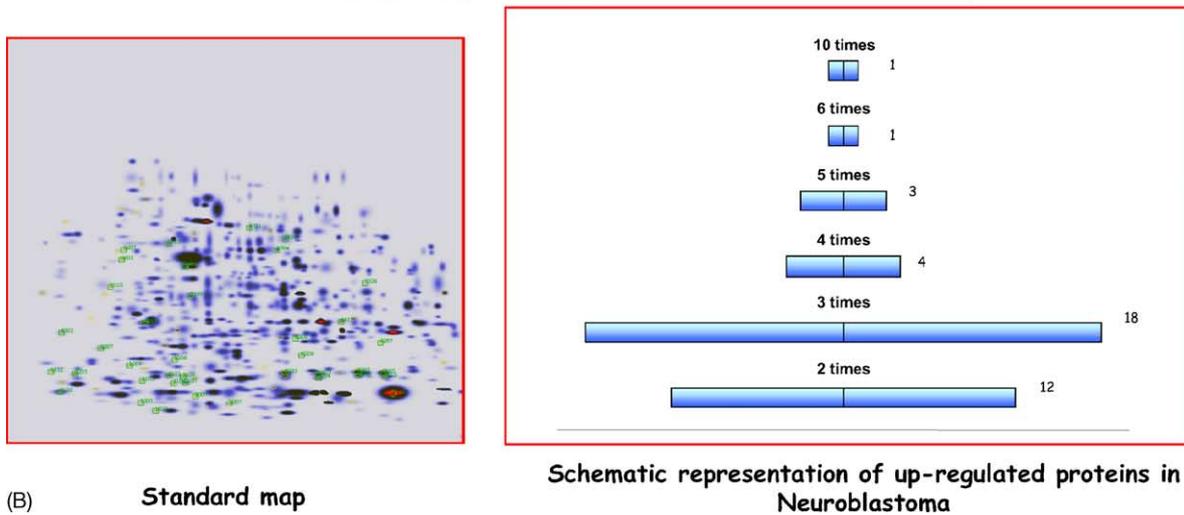
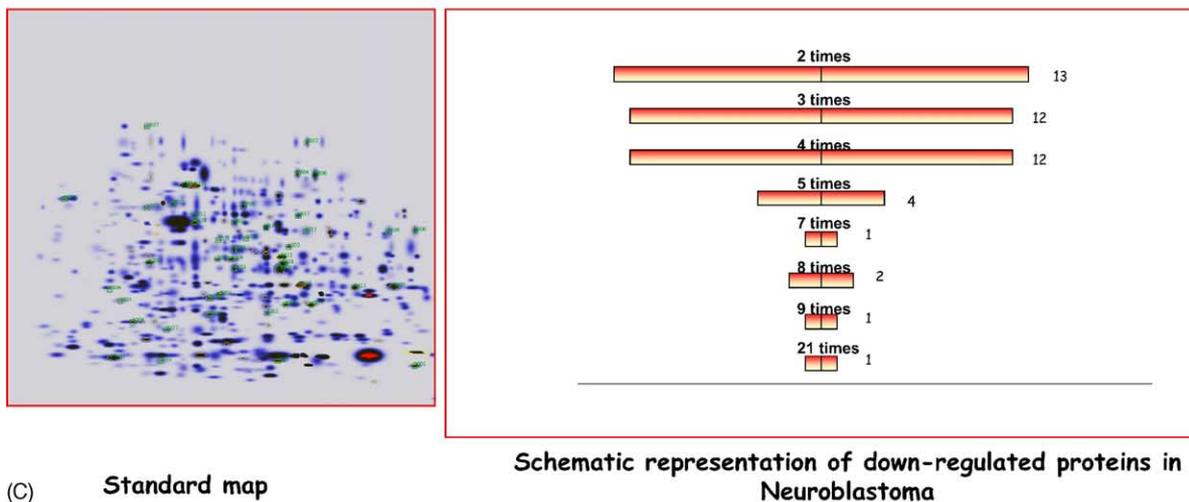


Fig. 1. Experimental design for generating sets of 2D maps from control and treated samples and for comparing master maps, stained with colloidal Coomassie Blue, for detection of up- and down-regulated proteins in the paired samples. All comparative steps performed with the PDQuest software. (A) Creating master maps from replica gels. (B) Display of up-regulated proteins. (C) Display of down-regulated proteins. (D) Display of newly expressed or suppressed proteins in neuroblastoma.

Down-regulated proteins in Neuroblastoma



Newly Expressed or suppressed proteins in Neuroblastoma

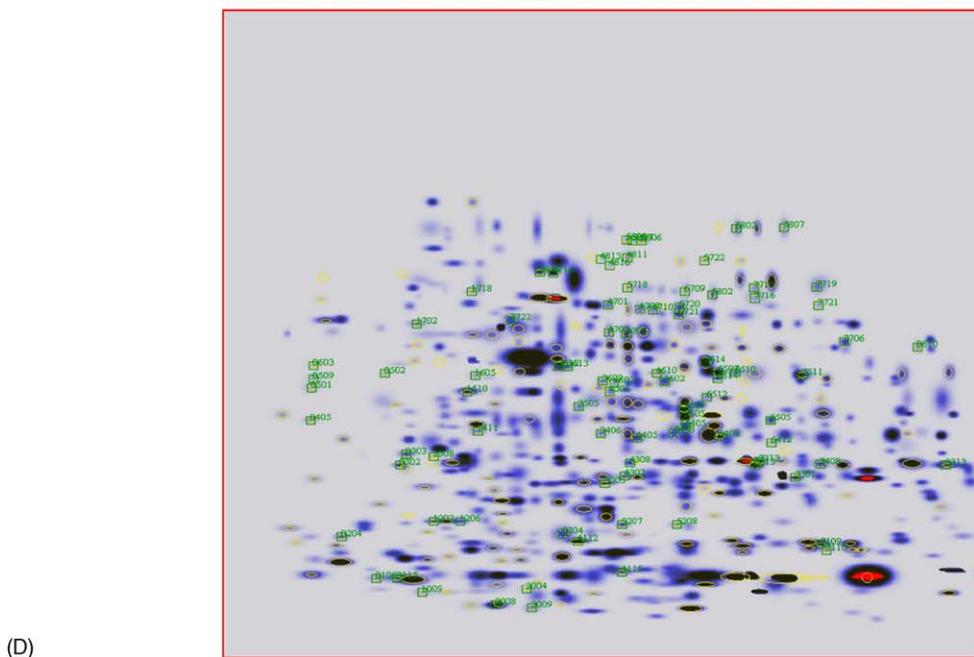


Fig. 1. (Continued).

a type of tumour that accounts for approximately 9% of all childhood cancers, occurring once out of 8000 live births, as analyzed in an experimental mice model. As illustrated in Fig. 1A, 4–5 replicas of such 2D maps should be run simultaneously, so as to maximize spot reproducibility (in general, we prefer fairly large-size 2D maps, 18 cm in the focusing dimension, 20 cm in the sodium dodecyl sulfate (SDS) dimension, although even larger sizes, e.g., 24 cm × 30 cm, have been reported). From the replicas of the control and pathological states, master maps are produced, which contain all spots found in the individual gels. Spot intensities were normalized in each gel and a statistical test was adopted to evaluate significant differences between the

healthy and tumoral groups, thus eliminating artefacts due to gel running. The comparison between the two master maps offers a clue about polypeptide chains whose expression is either up- or down-regulated. Fig. 1B gives an example of the up-regulated proteins in the tumour tissue, the bar graph to the right side listing the number of spots having experienced increments from two up to ten-folds (two-fold being the threshold for a statistically significant change in spot volume). Fig. 1C gives an analogous scheme for down-regulated proteins in neuroblastomas. Such analysis can offer additional information too, as shown in Fig. 1D. It can detect protein spots that are newly expressed in the tumoral samples as regard to control ones and protein spots

that are newly silenced in the tumoral samples as compared to healthy ones. These kinds of spots are highlighted in green. Once this differential analysis has been performed, all the spots of interest are excised, in-gel digested and subsequently characterized by mass spectrometry [e.g., using matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) or LC-electrospray ionization MS]. Once the precise fragmentation spectrum of each tryptic fragment is obtained, together with a lead sequence, interrogation of a number of databases (e.g., SwissProt, TrEMBL, NCBI, and the like) enables proper identification of the unknown protein, provided, of course, that it is listed in any of them. Although this procedure has been amply demonstrated in innumerable publications up to the present, it suffers from quite a few shortcomings: first of all, the extremely laborious and time-consuming set-up, requiring generation of at least five maps for each state (control versus disease); in a second instance, the fact that, for statistical reasons, the significant level of variation has to be set at quite high values, at least 100% (two-fold change in absorbance for each pair of spots under analysis). This means that any change below the threshold value of 100%, although of potential biological significance, has to be rejected. Thirdly, due to the large number of gels which have to be run, for minimizing experimental error, the method is highly demanding on the quantity of sample sacrificed for the assay, a serious problem in case of medical research, where, often, truly minute biopsies are available. Lastly, there is one thing that we found disturbing in some of these programs. E.g., when matched spots were missing in some gels, the PDQuest software assigned an arbitrary value to the missing spots, introducing a false normalised quantity in the calculation of the Student's *t*-test. In this case, we prefer to perform the statistical test with the GraphPad Prism for Windows (GraphPad Software, San Diego, CA, USA) and consider only the real values of the analyzed spots. Fortunately, a number of softwares are today available for image analysis and differential spot quantitation in 2D maps, as listed in Table 1. With some of them, the creation of master maps is greatly facilitated, and the acquisition time strongly shortened, due to the fact that the operator does not have to manually enter and verify each individual spot on all the maps; the software automatically takes care of that, thus shortening dramatically the elaboration time and minimizing operational errors. Some of these programs have been highly refined over the years, like the PDQuest, since they have been around at least from 1979 [15,16]. Some papers have also recently appeared evaluating and comparing the above-mentioned software packages [17–19]. The overall success of differential protein display in proteome research depends critically on the accuracy and the reliability of the analysis software. In addition, the software has a profound effect on the interpretation of the results obtained, and the amount of user intervention demanded during the analysis. The choice of analysis software that best meets specific needs is therefore of interest to the research laboratory. Different packages show dif-

ferent strengths and weaknesses. We will give here some general conclusions drawn by a pool of users: ImageMaster (Amersham Biosciences) is quoted among the most accurate packages, Z3 (Compugen) appears to be the most robust to poor S/N ratio and PDQuest (Bio-Rad Labs.) the most robust to spot overlap. Melanie III (GeneBio) performs well in all evaluations and Progenesis (Nonlinear Dynamics) has the advantage of a parameter-free spot detection, whilst also performing well in most evaluations. One should not forget, however, what is stated in footnote (a) of Table 1: all the packages listed under number 5–5.3 appear to be “essentially the same as “Phoretix 2D Evolution”, marketed under different trade names”! It is here additionally recalled that some companies might offer a range of packages to meet different experimental needs. A case in point is that of Nonlinear Dynamics, which proposes no less than four different types of softwares. “Phoretix 2D” is the standard work horse, robust and reliable, but competitively priced, for everyday use in any laboratory with a low throughput. Next on-line is “Phoretix 2D Evolution”, meant for laboratories with a low to medium throughput 2D gels. The “Progenesis Workstation” offers multiple analysis functionalities with limited user intervention and is targeted to those laboratories with medium to high daily gel production. Finally, “Progenesis Discovery”, the top of the line, is offered to those users who perform high-throughput proteomics and who require a fully automated analysis solution. The prices, of course, vary accordingly, from just 6750 US\$ for the simplest package up to as high as 120 000 US\$ for the top version.

3. Differential, in-gel electrophoresis based on Lys tagging

An alternative to the above protocol, could be the method known under the acronym of DIGE, differential in-gel electrophoresis, as first described in 1997 by Unlu et al. [20]. It is based on differential labelling with *N*-hydroxy-succinimide ester-modified cyanine fluors, the most popular couple being named Cy3 and Cy5 (see Fig. 2 for their formulas). Cy3 is excited at 540 nm and has an emission maximum at 590 nm, while Cy5 is excited at 620 nm and emits at 680 nm. The two samples to be compared are separately labelled with either Cy3 or Cy5, which covalently modify Lys residues in proteins. These dyes have positive charges to replace the loss of charge on the ϵ -amino group of Lys, and the molecular masses of the dyes are similar to each other (434 and 464 Da, respectively). The reaction is carried out so as to label only a few Lys residues per macromolecule (ideally, in fact, just one). As long as the extent of the reaction is similar between the samples to be compared, the mass shift will be uniform and the isoelectric point (pI) should be essentially unaltered. Given the distinguishable spectra of the two fluorophores, the two samples can then be combined and run in a single 2D gel. The differences between the quantities of the individual proteins from each sample can then be determined

Table 1
Commercial software packages currently available for 2D gel-image-analysis^{a,b}

No.	Software	Company	Year of arrival	Comments	Platforms	Images supported
1	Delta 2D	DECODON GmbH Http://www.decodon.com	2000	Save-disabled evaluation version available	PC [Windows 98, ME, 2000, NT], Linux, Sun Solaris, Mac OS X	TIFF (8, 12 and 16 bit), JPEG, BMP, GIF, PNG.
2	GELLAB II+	Scanalytics http://www.scanalytics.com/	1999	Trial version available	PC [Windows 95, NT]	TIFF (8 bit)
3	Melanie	Geneva Bioinformatics S.A. http://www.genebio.com	N/A ^c	30 day fully functional trial version available	PC [Windows 95, 98, 2000, NT]	TIFF (8, 16 bit), GIF, Bio-Rad Scan
4	PDQuest	Bio-Rad Laboratories Inc. http://www.bio-rad.com	1998	30 day fully functional trial version available	PC [Windows 95, 98, 2000, XP, NT], Macintosh Power PC	TIFF (8, 16 bit), 1 SC
5 ^d	Phoretix 2D Evolution	Nonlinear Dynamics Ltd. http://www.nonlinear.com http://www.phoretix.com	1991	Trial version available through sales agent	PC [Windows 95, 98, 2000, NT]	TIFF (8, 12 and 16 bit)
5.1	AlphaMatch 2D	Alpha Innotech Corporation http://alphainnotech.com	1999	Trial version available through sales agent	PC [Windows 95, 98, 2000, Me, NT]	TIFF (8,12 and 16 bit)
5.2	Image Master 2D Elite	Amersham Pharmacia Biotech http://www.apbiotech.com	2001	Trial version available through sales agent	PC [Windows 95, 98, 2000, Me, NT]	TIFF (8,12 and 16 bit)
5.3	Investigator HT Analyzer	Genomic Solutions Inc. http://www.genomicsolutions.com	2000	Trial version available through sales agent	PC [Windows 98, 2000, NT]	TIFF (8, 12 and 16 bit)
6	Progenesis	Nonlinear Dynamics Ltd. http://www.nonlinear.com http://www.phoretix.com	2001	Special hardware and software requirements	PC [Windows 2000]	TIFF (8, 12 and 16 bit), GEL, MEL, IMG
7	Z3	Compugen http://www.2dgels.com	2000	21 day fully functional trial version available	PC [Windows 98, 2000, NT]	TIFF (8, 12 and 16 bit), JPEG, BMP, GIF, PNG, GEL, FLT
8	ProteomeWeaver	Definiens (Munich, Germany) http://www.definiens-imaging.com	2002	21 day fully functional trial version available	PC [Windows 2000, XP]	TIFF (8, 12 and 16 bit), JPEG, BMP, GIF, PNG, GEL, FLT

^a The software packages listed in the table are only comprehensive off the shelf commercial software packages available for 2D gel-image-analysis. The information listed in the table has been obtained from various sources, including internet, literature and sales agents. Misinformation, if any, is purely unintentional.

^b Modified from Raman et al. [37].

^c Not available.

^d The software packages listed under #5 are essentially the same as “Phoretix 2D Evolution”, marketed under different brand names. Please contact individual companies to know about any differences that there may be.

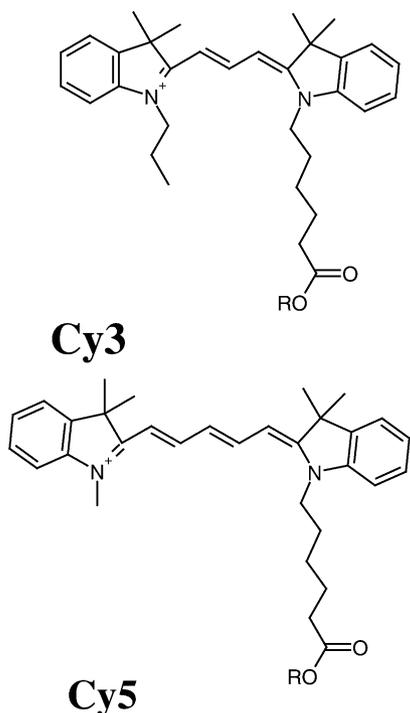


Fig. 2. Chemical formulae of the Cy3 and Cy5 dyes.

using specialized 2D image analysis software. Since both samples to be compared are separated in a single gel, this eliminates gel-to-gel variation, resulting in improved spot matching. As a corollary, the number of parallel and replicate gels required for obtaining reliable results is greatly reduced. Furthermore, fluorescence imparts the ability of detecting proteins over a much broader linear dynamic range of concentrations than visible gel stains [21]. Fig. 3 gives an example of the DIGE technique, as applied to the analysis of breast cancer cells ErbB-2-transformed [22]. Proteins that are present at equal levels in the two cell populations give a uniform violet hue. Proteins present in only one of the two tissues under comparison are either purely red or blue in colour, according to the Cy3/Cy5 label which they carry. Proteins up- and down-regulated give intermediate hues which are properly quantified by specialized software [23,24]. Because the labelling in the DIGE involves only a few Lys residues in each protein, the great part remains unlabelled. It is thus possible to stain the gel with another method in order to be able to perform further analysis such as peptide mapping. In Fig. 3 it is of interest to note that in the SYPRO RUBY image more proteins are visualized.

Just as an example of the power of this technique, Fig. 4 shows the differential analysis of one protein, L-plastin, detected only in tumour cells and not in the control. With this kind of analysis, it is also possible to perform a kinetic study

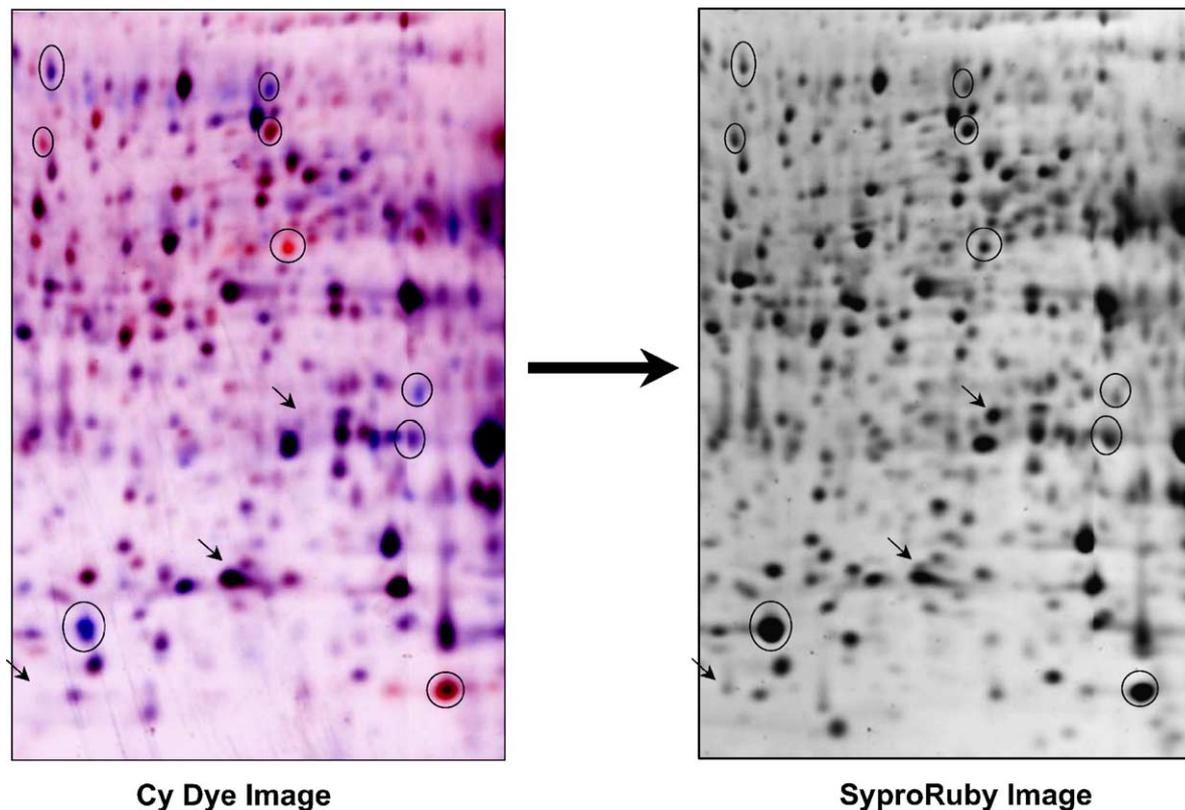


Fig. 3. Comparison of 2D-DIGE imaging and Sypro Ruby post-staining. Left panel: merged Cy dye image of HB4a lysate labelled with Cy3 (red) and HBc3.6 lysate labelled with Cy5 (blue). The same gel was post-stained with Sypro Ruby (right panel). Circles represent differentially expressed proteins detectable by both methods. Arrows represent spots detected by Sypro Ruby but not Cy-dye labelling. From [22] with permission.

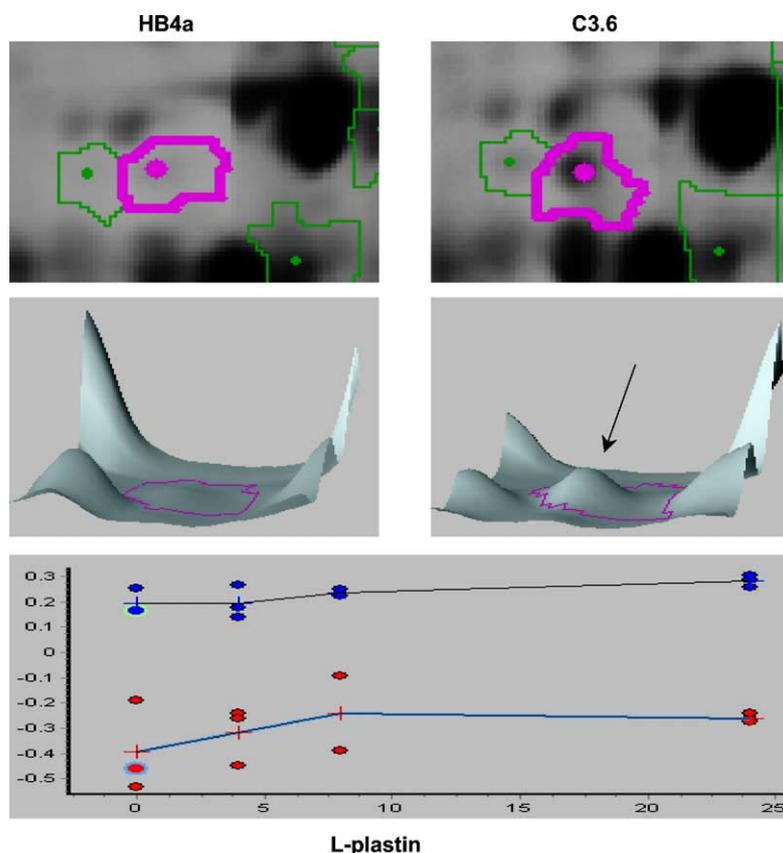


Fig. 4. Detailed gel area on a Cy3/Cy5 labelled sample pool, showing the induction of L-plastin in tumour cells. Bottom panel: kinetics of L-plastin induction. From [22] with permission.

on the induction of this protein, as shown in the bottom panel.

Differential proteome analysis is also applied in clinical medicine: for example in the study of body fluids from patients suffering from rheumatoid arthritis, reactive arthritis or osteoarthritis. This method has proven effective for identification of multiple molecular markers and determination of associated protein structure modifications that are thought to play a role for specifically determining defined pathological states of diseased joints [25]. One potential limitation of this method is that excision of spots of interest could be unreliable because, with minimal labelling conditions, only a few percent of a specific protein is labelled and this minor fluorescent population is generally shifted to slightly higher mass position due to the mass of the covalently bound dye. Therefore, the position of the bulk amount of unlabelled protein could be shifted about one spot diameter down (lower M_r values), but this could lead to excision of contaminants, different from the protein of interest. Should one carry the labelling of Lys to higher extents, the situation would be even more disastrous: not only this would generate more elongated spot areas along the second dimension (and possibly also along the first one), but it would surely impede trypsin action on the blocked Lys residues, thus generating a large number of missed cutting sites, much larger peptides

and inability to enter databases with correct values for protein identification.

4. Differential, in-gel electrophoresis based on Cys tagging

Among the drawbacks reported by users of the Cy3/Cy5 tagging, another one has been lamented: due to the “minimal labelling” approach, the stain sensitivity is not even comparable to that of silvering protocols. E.g., in the report by Zhou et al. [23], the total amount of spots remained less than 1000, whereas it is well-known that in any silvering procedure a minimum of 1500 spots in a total cell lysate are routinely detected. Perhaps to overcome this, and other limitations, Shaw et al. [26] have now reported another protocol for differential Cy3/Cy5 labelling, based on the reaction of a similar set of dyes not any longer on Lys, but on Cys residues. This technique is based on the opposite principle as compared to the original DIGE idea: not any longer “minimal”, but “maximal” labelling, i.e., saturation of all possible Cys reacting sites. This would fulfil two goals at once: on the one hand, it would automatically enhance the stain sensitivity; on the other hand, it would block further reactivity of reduced Cys residues. It will be briefly recalled

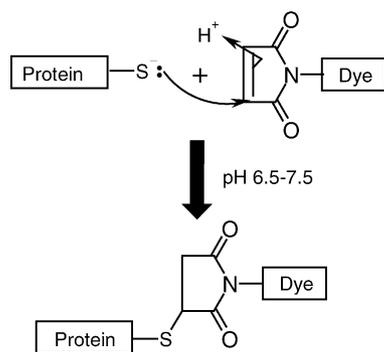


Fig. 5. Reaction scheme of maleimido cyanine dyes with the $-SH$ group of proteins. From [26] with permission.

here that, up to the year 2001, it was customary in 2D electrophoretic mapping to adopt a curious protocol, based on reducing the $-S-S-$ bridges of proteins, prior to the isoelectric focusing (IEF) step, but on performing the alkylation reaction only in between the first and second step, i.e., just prior to the SDS-polyacrylamide gel electrophoresis (PAGE) dimension. This was a disastrous protocol, of course, since alkylation at this point would not repair the artefactual spot pattern generated in the first dimension, due to spontaneous re-oxidation of $-SH$ to $-S-S-$ bridges in the alkaline pH region, with the formation of homo- and hetero-oligomers [27]. A labelling protocol aimed at Cys residues would automatically extinguish any further reactivity. As shown in Fig. 5, which gives the type of reaction of such compounds, it must be stated that the reacting tail of these two fluorophores is also quite appropriate in 2D analysis, since it is not an iodinated tail, which would automatically be destroyed by the thiourea in the solubilizing medium [28,29]. The reacting end is indeed a maleimide residue, permitting an addition of the $-SH$ group to the double bond of the maleimide moiety, thus forming a thioether link (although the structure of the dyes as not been disclosed as yet, their mass has been reported to be 673 and 685, respectively). We had in fact demonstrated that species with a reactive double bond, such as acrylamide, would not be scavenged by thiourea in the sample [29] and had suggested that α - β unsaturated compounds should be preferred as alkylating agents for $-SH$ groups in lieu of iodoacetamide.

Kondo et al. [30] have indeed adopted this couple of fluorescent dyes in a cancer proteomic study, aimed at the analysis of normal intestinal epithelium with that of adenoma in Min mice. They have claimed at least three major advantages with these newer tags:

- (1) First of all, the much higher sensitivity as compared with Lys-tagging, permitting detection of >1500 spots.
- (2) As a result of this "saturation labelling", the need for much decreased amounts of tissue biopsies, of the order of barely 6 μ g per gel.
- (3) A lowering of the statistical threshold of significant variation in spot intensity from 100% (as customary

in differential proteomics in separate gels, followed by PDQuest, or other software, analysis) to only 20% (due to the fact that samples are run admixed in a single gel, as typical of the DIGE strategy).

Although they [30] and Shaw et al. [26] have not claimed any disadvantages of this Cys-differential labelling to a saturation level, the situation might not be as rosy as depicted by this authors. Let us first examine more closely the data of Kondo et al. [30]. It is surprising that, although they claim that, by this procedure, they have detected significant expression level changes in 37 protein spots (of which 27 up- and 10 down-regulated), only a handful were indeed identified (eight, but indeed only four, since five of them were variant of the same family of 14–3–3 proteins). What is even more striking, these very few proteins identified were recognized by Western blots with specific antibodies, although they claim (without giving the relevant spectra) that the same set was also confirmed by mass spectrometry (MS, as customarily done in proteome analysis). There is a strong suspicion, here, that the extensive labelling with this bulky reagent might interfere in more than one way, not only by suppressing the MS signal (by quenching the ionisation of peptides, as candidly admitted by the authors) but also, perhaps, by interfering with the trypsin digestion too, thus producing fewer cuts than expected. This might be corroborated by their own statements: "the number of ion peaks from labelled protein spots was less than those from unlabelled proteins and the ability to identify the individual proteins with MS appeared to be affected" [30]. There are other matters of concern, of course. Among them, the extent of reaction: does the Cys blocking procedure achieve 100%, or is it considerably less? In a study, we performed on the reaction kinetics of iodoacetamide (or also acrylamide, for that matter), we could prove that the extent of reaction hardly reaches 80% final yield [28]. Insisting with an overnight incubation would only worsen the matter: the reaction would be even greater than 100%, simply because it will continue not on Cys, but on Lys residues, thus aggravating the matters when attempting to identify the relevant peptides by MS [28]. If the situation is so poor with such simple reactants, would it be any better with these fluorescent maleimide cyanine dyes, considering their bulky structure?

Another matter of concern comes also from a close inspection of Fig. 6, which compares identical maps obtained by staining with silver, with Cy3/Cy5 Lys and Cy3/Cy5 Cys. First of all, it would appear that, even with the saturation dye approach, the high sensitivity of silvering is not quite reached (compare panels A and C). Moreover, other serious changes are apparent: first of all, the massive shift of all proteins spots towards higher apparent M_r values, due to the bulky size of the cyanine dye. Secondly, the fact that quite a few of the spots appear blurred and out of focus, as though they have a tendency to precipitate along the migration path. In addition, the fact that the total number of spots is considerably less than in the silver gel image, notwith-

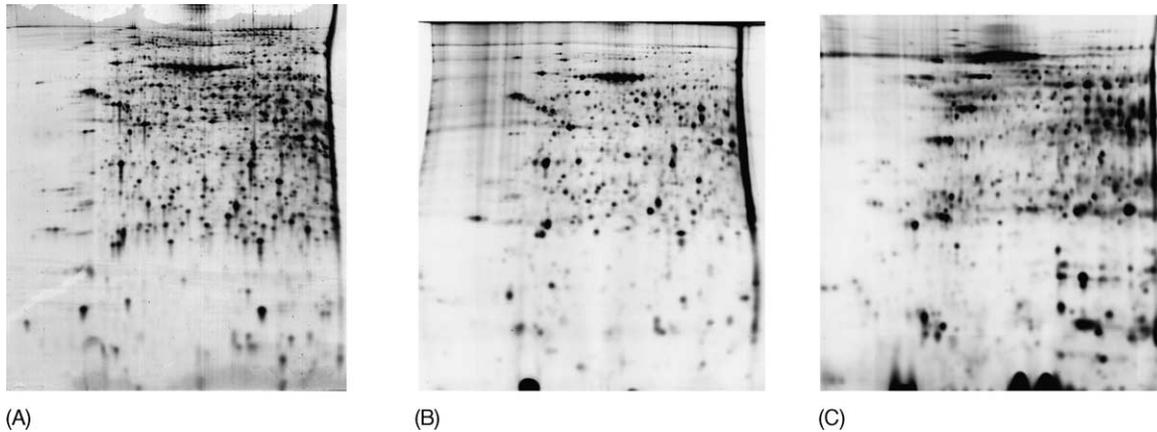


Fig. 6. 2D gel images of (A) silver stained, (B) Cy3 minimal dye-labelled and (C) Cy3 saturation dye-labelled liver homogenate (50 μ g, pH 4–7 18 cm IPG strip). Whilst the silver stained image and the Cy3 minimal dye image are very similar, the Cy3 saturation dye image shows an altered spot pattern. From [26] with permission.

standing the much higher fluorescent signal of the saturation label, makes one wonder if, by any chance, during the labelling protocol, a number of barely soluble proteins might precipitate out of solution due to increased hydrophobicity brought about by the cyanine dyes, thus disappearing from the map just at the onset of the 2D mapping procedure. This is in fact candidly admitted by Shaw et al. [26]: “on average, 25% of protein material was lost to precipitation during the labelling reaction, (*passim*) the losses are more significant with higher-molecular-mass proteins”. But what if additional protein losses were to occur during the focusing step? A number of labelled proteins, barely soluble at the pH of tagging, might precipitate during the IEF run at or close to the *pI* value, due to the well-known fact that the *pI* of a protein is a point on the pH scale of minimum of solubility.

5. Isotope-coded two-dimensional maps: [$^2\text{H}_0$]/[$^2\text{H}_3$] acrylamide

Isotope coding, for quantitative proteomics, was the brilliant brain child of Aebersold’s group, who proposed this protocol, called ICAT (isotope-coded affinity tags) already in 1999 [31,32]. In this novel procedure, stable isotopes are incorporated, in the two different samples to be compared, by the selective alkylation of Cys residues with either a “heavy” or “light” reagent; after that, the two protein pools to be compared are mixed. The ICAT reagent is composed of three parts: a biotin portion, used as an affinity tag; a linker, which can incorporate either the heavy or light isotopes and a third terminal group, which contains a reactive iodine atom able to alkylate specifically thiol groups (Cys residues). The

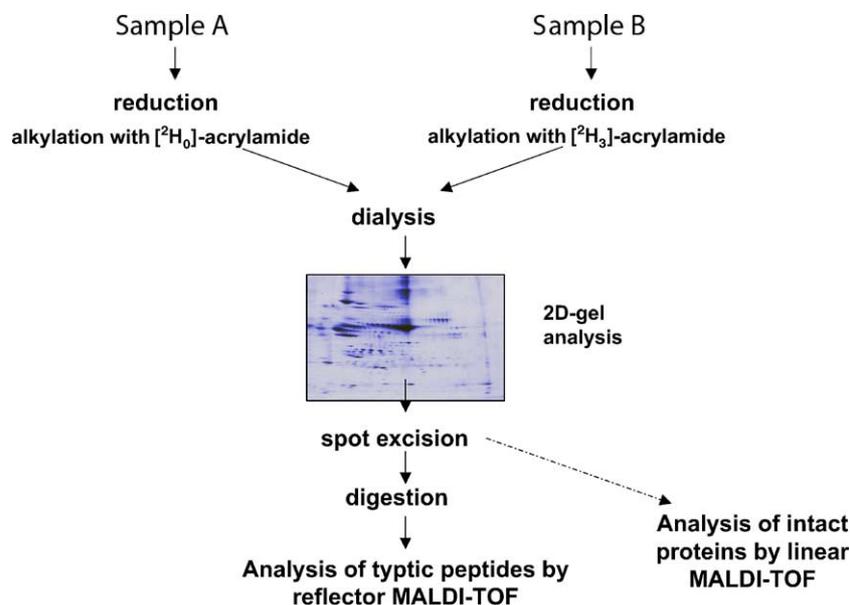


Fig. 7. Scheme for differential labelling of two samples with [$^2\text{H}_0$]/[$^2\text{H}_3$]-acrylamide (alkylation of Cys residues). The central map refers to rat sera, labelled separately with either [$^2\text{H}_0$]- or [$^2\text{H}_3$]-acrylamide and mixed in a 30:70% ratio. From [34] with permission.

“heavy” ICAT contains eight deuterium atoms, which in the “light” one are replaced by standard hydrogen atoms. Proteins from two different cell states are harvested, denatured, reduced and labelled at Cys residues with either light or heavy ICAT reagent. The samples are then combined and digested with trypsin. ICAT-labelled peptides can be further isolated by biotin-affinity chromatography and then analyzed by on-line HPLC coupled to tandem MS. The ratio of the ion intensities for any ICAT-labelled pair quantifies the relative abundance of its parent protein in the original cell state. In addition, the tandem MS approach produces the sequence of the peptide, and thus can unambiguously identify the protein of interest. This strategy, ultimately, results in the quantification and identification of all protein components in a mixture and, in principle, could be applied to

protein mixtures as complex as the entire genome. Needless to say, this protocol cannot be applied to 2D map analysis. Not just because this procedure calls for trypsin digestion prior to sample analysis (one could omit this step), but because this reagent contains a reactive iodine tail. As stated above, since essentially all 2D map procedures adopted today contain 2 M thiourea in the sample solubilization buffer, the ICAT would quickly be destroyed as soon as added to such a sample buffer.

Aware of this limitations, yet fascinated by the brilliant idea of ICAT, we explored the possibility of exploiting this very ICAT technique in electrophoretic 2D maps. This approach would utilize the same ICAT concept, but by labelling intact macromolecules and disposing of the affinity tail, certainly not needed in conventional mapping strate-

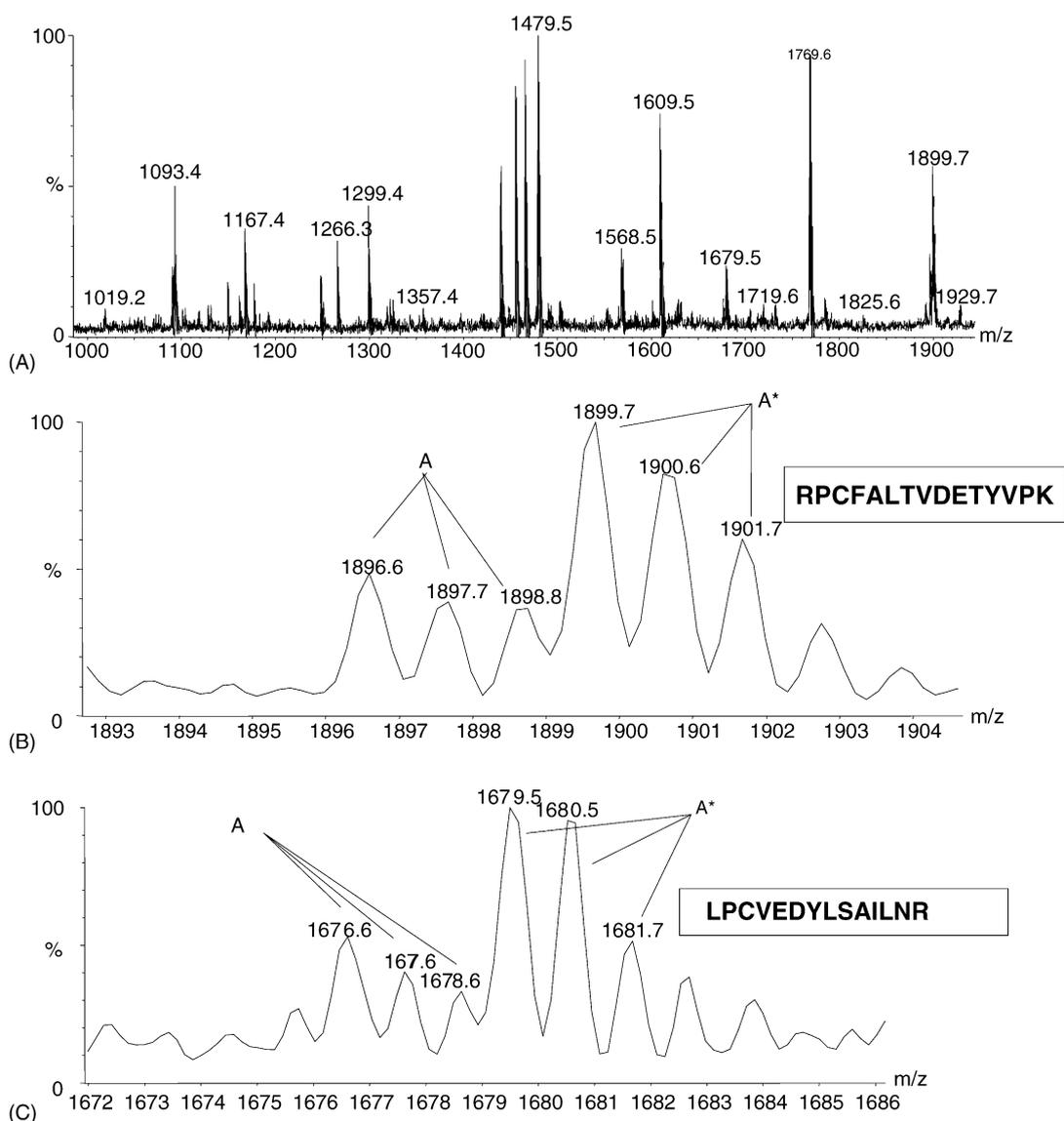


Fig. 8. (A) Reflector MALDI mass spectrum of an in-situ digest of apo-transferrin taken from the 2D map of rat sera displayed in Fig. 7, that were alkylated with $[^2\text{H}_0]$ - and $[^2\text{H}_3]$ -acrylamide and mixed in a 30%/70% ratio. (B) and (C) are two short intervals taken from (A), and are associated with the two indicated peptide sequences. From [34] with permission.

gies. An example of such an approach could be the use of [$^2\text{H}_0$]/[$^2\text{H}_3$]-acrylamide for blocking Cys residues in intact protein molecules. The use of light/heavy acrylamide to alkylate proteins prior to their 2D electrophoretic separation was in fact simultaneously and independently described by Sechi [33] and by Gehanne et al. [34]. Both reports have demonstrated that this procedure, when combined with MALDI-TOF-MS, could be a valid tool for protein identi-

fication and relative quantification. The basic steps in such approach are depicted in Fig. 7. Basically, relative quantification of individual proteins in two different samples is achieved by alkylating one sample with [$^2\text{H}_0$]-acrylamide, and the second with its [$^2\text{H}_3$] counterpart; the two samples are then combined with predetermined ratios, dialyzed, and subjected to 2D gel electrophoresis. Following visualisation of the separated proteins, each spot can be excised,

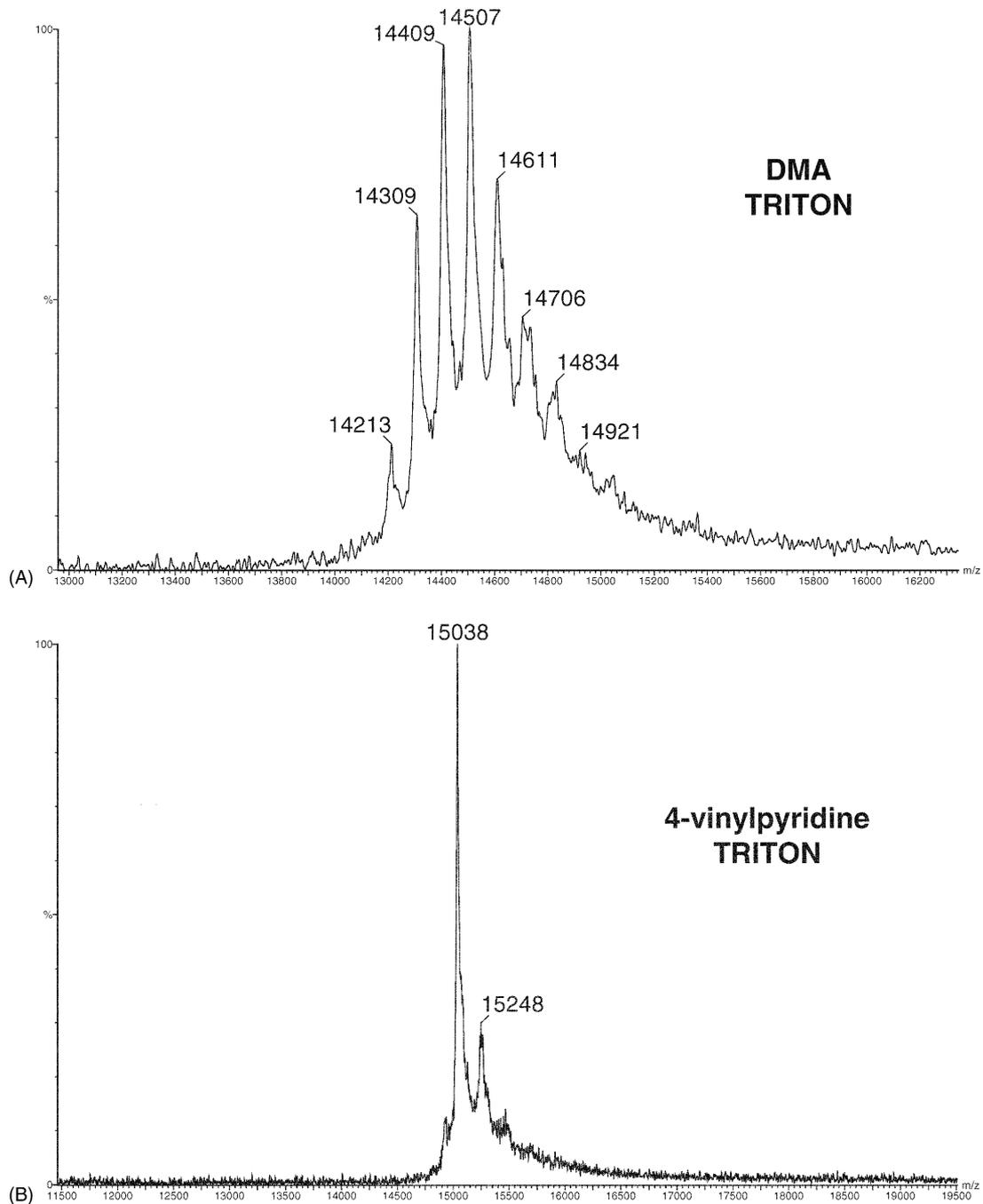


Fig. 9. MALDI-TOF mass spectra of bovine α -lactalbumin after 1 h incubation with DMA (A) or 4-vinylpyridine (B), both in presence of the surfactant 2% Triton X-100. Note that, in panel B, the peak at m/z 15248 represents an adduct of LCA with the MALDI matrix, sinapinic acid. From [36] with permission.

digested with trypsin, and examined by MALDI-TOF. The relative quantification of a number of proteins would then be obtained by comparing the relative peak heights within a reflector MALDI spectrum of two adjacent isotopic envelopes that happen to differ by m/z 3. The application of this approach to quantitation of various proteins within the 2D map of rat serum shown in Fig. 7 is illustrated below. The map in Fig. 7, covering the pH 3–10 IPG interval, was obtained by mixing in different proportions two fractions of rat sera, the first (30%) being alkylated with [$^2\text{H}_0$]-acrylamide, and the second (70%) reacted with [$^2\text{H}_3$]-acrylamide. A representative example of a reflector MALDI spectrum that pertains to apo-transferrin is given in Fig. 8A–C. The spectrum of the entire digest is given in (A), whereas (B) and (C) display two short intervals of the same spectrum and show two isotopic distributions marked A and A* in which a difference of 3 Da in the m/z values of the corresponding peaks is clearly evident. A database search yielded the two indicated peptides, each of which contains a single cysteine. Considering the relative peak heights in

both isotopic distributions, a ratio of 34:66 was obtained, which is in good agreement with the labelling ratio 30:70 prior to 2D separation. Interestingly, this method has been recently validated by Cahill et al. [35] and given a good performance score. For instance, these authors have found that spots labelled with either [$^2\text{H}_0$] or [$^2\text{H}_3$]-acrylamide effectively co-migrate in the IEF dimension (i.e., there is no isotope effect shifting the pK , thus the pI values, of proteins). In addition, whereas we have always reported incomplete alkylation of proteins by both iodoacetamide or acrylamide (typically, 80–85% extent of reaction), they have claimed 100% alkylation ability of this system, provided, though, that such alkylation is conducted in boiling 2% SDS (for 1 h in the case of acrylamide and for 15 min for iodoacetamide), conditions perhaps not fully compatible with the first, IEF dimension of 2D maps. Being more conservative, we prefer to stick to our figures of 80–85% conversion of –SH groups in Cys residues, which might be one of the limitations of this protocol (its obvious advantages being the ease of reaction and the very low cost of the deuterated chemicals).

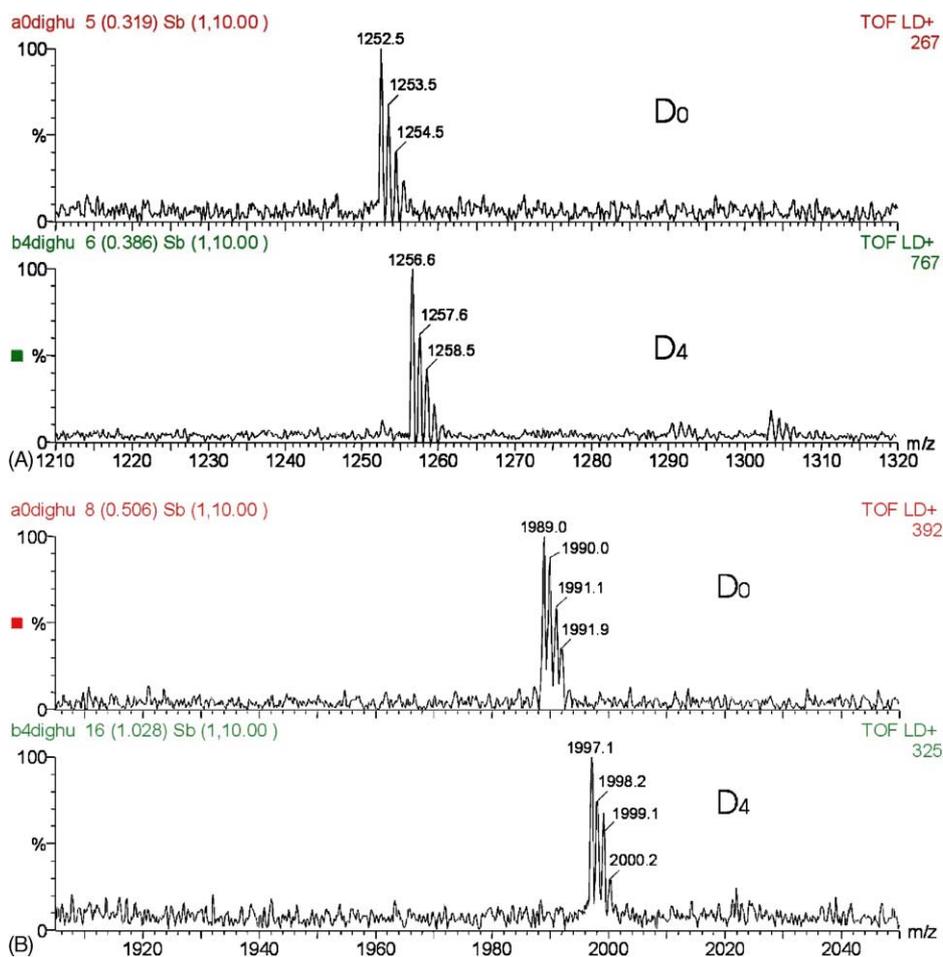


Fig. 10. (A) Zoom-in to monoisotopic distributions of the m/z 1252.5 peptide, obtained by tryptic digestion of α -lactalbumin, labelled with [$^2\text{H}_0$] 2-VP (D₀) and the corresponding m/z 1256.6 peptide labelled with [$^2\text{H}_4$] 2-VP (D₄). The peptide contains a single Cys residue. (B) Zoom-in to monoisotopic distributions of the m/z 1989.0 peptide labelled with [$^2\text{H}_0$] 2-VP and the corresponding m/z 1997.1 peptide labelled with [$^2\text{H}_4$] 2-VP. Note that this peptide contains 2 Cys residues. From [36] with permission.

6. Isotope-coded two-dimensional maps: [$^2\text{H}_0$]/[$^2\text{H}_4$] 2-vinylpyridine

We have seen that there are inherent shortcomings of the above methods exploiting stable isotope labelling. To start with, both the ICAT and acrylamide rarely achieve better than 80% conversion of all –SH groups in Cys, a major drawback when attempting protein quantitation of all phenotypes in a biological specimen. In addition, ICAT would be rapidly destroyed by thiourea, a common protein solubilizer in modern electrophoretic 2D map analyses. We have thus wondered if there could be a special chemical coupling 100% reactivity with 100% specificity, reaction features rarely met when attempting any kind of protein derivatization. Preliminary experiments had indeed demonstrated that weakly basic molecules containing a double bond, such as 2- and 4-vinylpyridines (VPs), were able to react and selectively alkylate –SH groups in proteins, thus preventing their re-oxidation to disulphur bridges. Contrary to conventional alkylating agents, such as iodoacetamide and non-charged acrylamide derivatives, such molecules seemed to offer 100% alkylation of all –SH residues, even in complex proteins, without reacting with other functional groups [12]. This can be easily appreciated in Fig. 9, which shows the alkylation power of dimethylacrylamide (DMA), as compared with 4-VP, in presence of the surfactant Triton X-100, known to quench such reactions. Whereas the control panel (A), shows a large number of reaction channels, starting with the mono-up to the barely traces of the octa-alkylated (the target) species, panel (B) shows just a single reaction product, corresponding to the target, octa-alkylated species (the second peak to the right being the adduct with sinapinic acid). We thus set out to synthesize a tetra-deuterated 2-VP, and measured its reactivity with α -lactalbumin, a protein containing eight –SH groups. MALDI-TOF analysis showed that all (and only) the peptides containing a Cys residue were fully alkylated [36]. Zooms of two of these peptides, as shown in Fig. 10A and B, indeed show that, when exploring their mono-isotopic distribution, the [$^2\text{H}_0$]/[$^2\text{H}_4$]-tagged peptides were spaced apart by 4 Da (in case of single Cys peptides, Fig. 10A) and by 8 Da, in the case of double-Cys peptides (Fig. 10B).

7. Conclusions

We have reviewed here a number of approaches to quantitative proteomics in 2D map analysis (for more on informative tools for proteome profiling, see also Chakravarti et al. [38]). The good old method of separate replicas of 2D maps, stained with colloidal Coomassie and then analyzed and matched by softwares able to detect up- and down-regulation (and appearance or disappearance of spots) of proteins via differential dye uptake, although terribly time-consuming and labour intensive, is still a good and reliable work horse. We have nicknamed it the “peones” approach, since one

of its characteristic is its relatively inexpensive set-up and low cost, making it suitable in all labs surviving on a tight budget. The DIGE technology, either via “minimal” (on Lys residues) or “saturation” (on Cys residues) labelling, might certainly be a powerful approach, but it surely requires equipment of very high cost, coupled to a quite expensive reagent kit. Even this “elitist” system, though, might be prone to problems, especially in the Cys-tagging procedure, where protein spots identification via MS appears to be problematic. Perhaps a good compromise might be alkylation with the stable isotope approach, namely with [$^2\text{H}_0$]/[$^2\text{H}_4$] 2- or 4-vinylpyridines. It would appear that either 2- or 4-VPs are ideal alkylators for Cys groups in proteins: they guarantee 100% reactivity coupled to 100% specificity, properties which lack in all other alkylating agents investigated, both with a reactive iodine tail or with a reacting double bond. In addition to these unique properties, 2- or 4-VP appear to be insensitive to reaction inhibition typically exhibited by all other reagents in presence of neutral or zwitterionic surfactants, common additives in solubilization cocktails for 2D map analysis. An extra bonus, shared by all reagents containing a double bond, is their unreactivity towards thiourea, a fundamental solubilizer in total cell lysates and membrane analysis, contrary to reagents containing a terminal reactive iodine, which are rapidly destroyed in presence of thiourea. As a final comment, it must be stated that most procedures of Cys alkylation, here reported, have been recently challenged by Luche et al. [39], on the grounds that essentially all Cys alkylators either under- or over-alkylate proteins. However, these authors did not seem to be aware of the work of Sebastiano et al. [36] on VPs, neither of the work of Mineki et al. [40], who reported 97% alkylation of SH groups in BSA with high levels of acrylamide. With the latter compound, Luche et al. [39] report spurious alkylation on Lys residues, but they make the fundamental mistake of not removing the excess alkylant during the IPG run (where alkylation will continue undisturbed!).

Acknowledgements

Supported in part by the European Union, Grant No. QLG2-CT-2001-01903, by FIRB-2001 (No. RBNF01KJHT), by PRIN-2003 from MIUR (Rome, Italy), by AIRC (Associazione Italiana per la Ricerca sul Cancro) and by Fondazione Cassa di Risparmio di Verona.

References

- [1] P.G. Righetti, A. Stoyanov, M.Y. Zhukov, *The Proteome Revisited: Theory and Practice of All Relevant Electrophoretic Steps*, Elsevier, Amsterdam, 2001.
- [2] M.J. Powell, A.T. Timperman, *Methods Mol. Biol.* 251 (2004) 387.
- [3] R.C. Storhman, *Perspect. Biol. Med.* 37 (1993) 112.
- [4] R.C. Storhman, *BioTechnology* 12 (1994) 156.
- [5] J.H. Leamon, W.L. Lee, K.R. Tartaro, J.R. Lanza, G.J. Sarkis, A.D. Dewinter, J. Berka, K.L. Lohman, *Electrophoresis* 24 (2003) 3769.

- [6] S.P. Gygi, B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb, R. Aebersold, *Nature Biotech.* 17 (1999) 994.
- [7] J.P.A. Baak, F.R.C. Path, M.A.J.A. Hermsen, G. Meijer, J. Schmidt, E.A.M. Janssen, *Eur. J. Cancer* 39 (2003) 1199.
- [8] R.E. Banks, M.J. Dunn, D.F. Hochstrasser, J.C. Sanchez, W. Blackstock, D.J. Pappin, *Lancet* 356 (2000) 1749.
- [9] R.J. Simpson, D.S. Dorow, *Trends Biotech.* 19 (2001) S40.
- [10] M. Hamdan, M. Galvani, P.G. Righetti, *Mass Spectrom. Rev.* 20 (2001) 121.
- [11] H. Hamdan, P.G. Righetti, *Mass Spectrom. Rev.* 21 (2002) 287.
- [12] P.G. Righetti, A. Castagna, M. Hamdan, *Adv. Chromatogr.* 42 (2003) 270.
- [13] H. Hamdan, P.G. Righetti, *Mass Spectrom. Rev.* 22 (2003) 272.
- [14] N.L. Anderson, J. Taylor, A.E. Scandora, B.P. Coulter, N.G. Anderson, *Clin. Chem.* 27 (1981) 1807.
- [15] J.A. Garrells, *J. Biol. Chem.* 254 (1979) 7961.
- [16] J.A. Garrells, J.T. Farrar, B.C. Burwell IV, in: J.E. Celis, R. Bravo (Eds.), *Two-Dimensional Gel Electrophoresis of Proteins*, Academic Press, Orlando, 1984, p. 37.
- [17] B. Raman, A. Cheung, M.R. Marten, *Electrophoresis* 23 (2002) 2194.
- [18] A. Rubinfeld, T. Keren-Lehrer, G. Hadas, Z. Smilansky, *Proteomics* 3 (2003) 1930.
- [19] A.T. Rosengren, J.M. Salmi, T. Aittokallio, J. Westerholm, R. Lahesmaa, T.A. Nyman, O.S. Nevalainen, *Proteomics* 3 (2003) 1936.
- [20] M. Unlu, M.E. Morgan, J.S. Minden, *Electrophoresis* 18 (1997) 2071.
- [21] W.F. Patton, *Electrophoresis* 21 (2000) 1123.
- [22] S. Gharbi, P. Gaffney, A. Yang, M.J. Zvelebil, R. Cramer, M.D. Waterfield, J.F. Timms, *Mol. Cell Proteomics* 2 (2002) 91.
- [23] G. Zhou, H. Li, D. DeCamps, S. Chen, H. Shu, Y. Gong, M. Flaig, J.W. Gillespie, N. Hu, P.R. Taylor, M.R. Emmert-Buck, L.A. Liotta, E.F. Petricon III, Y. Zhao, *Proteomics* 1 (2002) 117.
- [24] R. Tonge, J. Shaw, B. Middleton, R. Rawlinson, S. Rayner, J. Young, F. Pognan, E. Hawkins, I. Currie, M. Davison, *Proteomics* 1 (2001) 377.
- [25] A. Sinz, M. Bantscheff, S. Mikkat, B. Ringel, S. Drynda, J. Kekow, H.J. Thiesen, M.O. Glocker, *Electrophoresis* 23 (2002) 3445.
- [26] J. Shaw, R. Rowlinson, J. Nickson, T. Stone, A. Sweet, K. Willimas, R. Tonge, *Proteomics* 3 (2003) 1181.
- [27] B. Herbert, M. Galvani, M. Hamdan, E. Olivieri, J. McCarthy, S. Pedersen, P.G. Righetti, *Electrophoresis* 22 (2001) 2046.
- [28] M. Galvani, M. Hamdan, B. Herbert, P.G. Righetti, *Electrophoresis* 22 (2001) 2058.
- [29] M. Galvani, L. Rovatti, M. Hamdan, B. Herbert, P.G. Righetti, *Electrophoresis* 22 (2001) 2066.
- [30] T. Kondo, M. Seike, Y. Mori, K. Fujii, T. Yamada, S. Hirohashi, *Proteomics* 3 (2003) 1758.
- [31] S.P. Gygi, B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb, R. Aebersold, *Nat. Biotech.* 17 (1999) 994.
- [32] S.P. Gygi, R. Aebersold, in: W. Blackstock, M. Mann (Eds.), *Proteomics: a trend guide*, Elsevier, London, 2000, p. 31.
- [33] S. Sechi, *Rapid Commun. Mass Spectrom.* 16 (2002) 1416.
- [34] S. Gehanne, D. Cecconi, L. Carboni, P. G. Righetti, E. Domenici, M. Hamdan, *Rapid Commun. Mass Spectrom.* 16 (2002) 1692.
- [35] M.A. Cahill, W. Wozny, G. Schwall, K. Schroer, K. Hoelzer, S. Poznanovic, C. Hunzinger, J.A. Vogt, W. Stegmann, H. Matthies, A. Schratzenholz, *Rapid Commun. Mass Spectrom.* 17 (2003) 1283.
- [36] R. Sebastiano, A. Citterio, M. Lapadula, P.G. Righetti, *Rapid Commun. Mass Spectrom.* 17 (2003) 2380.
- [37] B. Raman, A. Cheung, M.R. Marten, *Electrophoresis* 23 (2002) 2194.
- [38] D.N. Chakravarti, B. Chakravarti, I. Moutsatsos, *Comput. Proteomics Suppl.* 32 (2002) S4.
- [39] S. Luche, H. Diemer, C. Tastet, M. Chevallet, A. van Dorsselaer, E. Leize-Wagner, T. Rabilloud, *Proteomics* 4 (2004) 551.
- [40] R. Mineki, H. Taka, T. Fujimura, M. Kikkawa, N. Shindo, K. Murayama, *Proteomics* 2 (2002) 1672.

REVIEW

Numerical approaches for quantitative analysis of two-dimensional maps: A review of commercial software and home-made systems

Emilio Marengo¹, Elisa Robotti¹, Francesca Antonucci², Daniela Cecconi²,
Natascia Campostrini² and Pier Giorgio Righetti²

¹ Department of Environmental and Life Sciences, University of Eastern Piedmont, Alessandria, Italy

² Department of Agricultural and Industrial Biotechnologies, University of Verona, Verona, Italy

The present review attempts to cover a number of methods that have appeared in the last few years for performing quantitative proteome analysis. However, due to the large number of methods described for both electrophoretic and chromatographic approaches, we have limited this review to conventional two-dimensional (2-D) map analysis which couples orthogonally a charge-based step (isoelectric focusing) to a size-based separation step (sodium dodecyl sulfate-electrophoresis). The first and oldest method applied to 2-D map data reduction is based on statistical analysis performed on sets of gels *via* powerful software packages, such as Melanie, PDQuest, Z3 and Z4000, Phoretix and Progenesis. This method calls for separately running a number of replicas for control and treated samples. The two sets of data are then merged and compared *via* a number of software packages which we describe. In addition to commercially-available systems, a number of home made approaches for 2-D map comparison have been recently described and are also reviewed. They are based on fuzzyfication of the digitized 2-D gel image coupled to linear discriminant analysis, three-way principal component analysis or a combination of principal component analysis and soft-independent modeling of class analogy. These statistical tools appear to perform well in differential proteomic studies.

Received: June 22, 2004
Revised: September 6, 2004
Accepted: September 10, 2004

Keywords:

Fuzzy logic / Linear discriminant analysis / Principal component analysis / Review / Two-dimensional maps

Contents

1	Introduction.....	655
2	A practical example: The PDQuest system.....	656
2.1	Scanning.....	656
2.2	Image filtering.....	656
2.3	Automated spot detection.....	656
2.4	Matching of protein profiles.....	656
2.5	Normalization.....	657
2.6	Differential analysis.....	657
2.7	Statistical analysis.....	657
3	Home-made approaches: Fuzzy logics.....	658
3.1	Image digitalization.....	658
3.2	Image defuzzyfication.....	658
3.3	Refuzzyfication.....	658
3.4	Fuzzy logic and PCA.....	659
3.5	Fuzzy logic and multidimensional scaling.....	660
4	Home-made approaches: Three-way PCA.....	661

Correspondence: Professor Pier Giorgio Righetti, Scientific and Technological Department, University of Verona, Strada Le Grazie No. 15, Verona 37134, Italy
E-mail: righetti@sci.univr.it
Fax: +39-045-8027929

Abbreviations: MDS, multidimensional scaling; PC, principal component; PCA, principal component analysis; SIMCA, soft-independent model of class analogy; TSA, Trichostatin-A

5	Home-made approaches: PCA coupled to classification methods and cluster analysis	662
5.1	PCA and soft-independent model of class analogy	663
5.2	PCA and cluster analysis	663
6	Concluding remarks	664
7	References	664

1 Introduction

The power of the 2-D gel technique lies in its capacity to separate simultaneously thousands of proteins for their subsequent identification and quantitative comparison.

Such a proteomics approach typically requires the quantitative analysis of numerous sets of gels for revealing differential protein expression across multiple experiments. As the experiments result in large amounts of data, efficient use of the 2-D techniques relies on powerful and user-friendly data analysis by means of computer algorithms. A number of software packages have appeared in the last decade and are listed in Table 1. A few papers have been published comparing some of these packages: PDQuest and Progenesis [1]; Z3 and Melanie 3.0 [2]. Other reports assess just a single product, e.g. Phoretix 2-D full [3]; Compugen's Z4000 [4]. Other articles deal with general aspects of such algorithms, such as point pattern matching, reproducibility, matching efficiency *etc.* [5–10].

Table 1. Commercial software packages currently available for 2-D gel image analysis^{a), b)}

#	Software	Company	Year of arrival	Comments	Platforms	Images supported
1	Delta 2-D	DECODON http://www.decodon.com	2000	Save-disabled evaluation version available	PC (Windows 98, ME, 2000, NT), Linux, Sun Solaris, Mac OS X	TIFF (8, 12 and 16 bit), JPEG, BMP, GIF, PNG.
2	GELLAB II+	Scanalytics http://www.scanalytics.com/	1999	Trial version available	PC (Windows 95, NT)	TIFF (8 bit)
3	Melanie	Geneva Bioinformatics http://www.genebio.com	N/A ^{d)}	30 day fully functional trial version available	PC (Windows 95, 98, 2000, NT)	TIFF (8, 16 bit), GIF, Biorad Scan
4	PD Quest	Bio-Rad Laboratories http://www.biorad.com	1998	30 day fully functional trial version available	PC (Windows 95, 98, NT), Macintosh Power PC	TIFF (8, 16 bit)
5 ^{c)}	Phoretix 2-D Advanced	Nonlinear Dynamics http://www.nonlinear.com http://www.phoretix.com	1991	Trial version available through sales agent	PC (Windows 95, 98, 2000, NT)	TIFF (8, 12 and 16 bit)
5.1	AlphaMatch 2-D	Alpha Innotech Corporation http://alphainnotech.com	1999	Trial version available through sales agent	PC (Windows 95, 98, 2000, Me, NT)	TIFF (8,12 and 16 bit)
5.2	Image Master 2-D Elite	Amersham Biosciences http://www.apbiotech.com	2001	Trial version available through sales agent	PC (Windows 95, 98, 2000, Me, NT)	TIFF (8,12 and 16 bit)
5.3	Investigator HT Analyzer	Genomic Solutions http://www.genomicsolutions.com	2000	Trial version available through sales agent	PC (Windows 98, 2000, NT)	TIFF (8, 12 and 16 bit)
6	Progenesis	Nonlinear Dynamics http://www.nonlinear.com http://www.phoretix.com	2001	Special hardware and software requirements	PC (Windows 2000)	TIFF (8, 12 and 16 bit)
7	Z3	Compugen http://www.2dgels.com	2000	21 day fully functional trial version available	PC (Windows 98, 2000, NT)	TIFF (8, 12 and 16 bit), JPEG, BMP, GIF, PNG, GEL, FLT
8	Proteome-Weaver	Definiens (Munich, Germany)	2002	21 day fully functional trial version available	PC (Windows 2000, XP)	TIFF (8, 12 and 16 bit), JPEG, BMP, GIF, PNG, GEL, FLT

a) The software packages listed in the table are comprehensive off the shelf-commercial software packages available for 2-D gel image-analysis. The information listed in the table has been obtained from various sources, including the internet, literature and sales agents. Misinformation, if any, is purely unintentional

b) Modified from Raman *et al.* [2]

c) The software packages listed under 5 are essentially the same as Phoretix 2-D advanced, marketed under different brand names. For comparisons of software packages please contact the individual companies

d) Not available

We will review in detail the steps required for analyzing and matching data obtained from control and treated samples run in separate gels and subsequently subjected to matching *via* different software packages. Given the aim of this review, we will not treat systems that deal with matching and comparison of spots from two different samples run in a single gel, such as the DIGE technique [11, 12] or techniques utilizing sample populations treated with stable isotopes, such as those described by Sechi [13] and Gehanne *et al.* [14], since in these cases the samples, after separate labelling with different reporters, are mixed, usually in a 1:1 ratio, loaded on a single gel and thus displayed, at the end of the two electrophoretic passages, in a single gel slab. Accurate quantitation relies upon computer analysis of a digitized representation of the stained gel. In the majority of cases, images are captured by laser densitometry, phosphor imagery or *via* a CCD camera. Once the image has been digitized, a standard computer-assisted analysis of 2-D gels includes at least the following three basic steps: (i) protein spot detection; (ii) spot quantitation; (iii) gel-to-gel matching of spot patterns. Since we have gathered extensive experience with the PDQuest system in our laboratory, we will deal in depth with it, while, of course, not endorsing it.

2 A practical example: The PDQuest system

PDQuest is a software system for imaging, analyzing, and data basing 2-DE gels. It is one of the oldest software packages with early reports on it dating back to 1979 [15]. It has been continually refined over the years [16, 17]. Extensive literature also exists on other programs developed in the early eighties, such as the Tycho system [18], Melanie [19–21], GELLAB [22], and the GESA processing system [23], to name just a few (for more on these early systems, see [24, 25]). Once a gel has been scanned, advanced algorithms are available for removing background noise, gel artefacts, and horizontal or vertical streaking from the image. PDQuest then uses a spot segmentation facility to detect and quantify protein spots. Protein spots that change over time can be traced, quantified, displayed on-screen, and exported to other applications for statistical analysis. The spot matching and data basing facilities make it possible to objectively compare hundreds of different gels. Specifically, the series of steps necessary for proper gel evaluation can be summarized as follows: (i) scanning; (ii) filtering images; (iii) automated spot detection; (iv) matching of protein profiles; (v) normalization; (vi) differential analysis; (vii) statistical analysis. We will briefly describe here the various steps.

2.1 Scanning

This process converts signals from biological samples into digital data. A data object, as displayed on the computer, is composed of tiny individual screen pixels. Each pixel has an

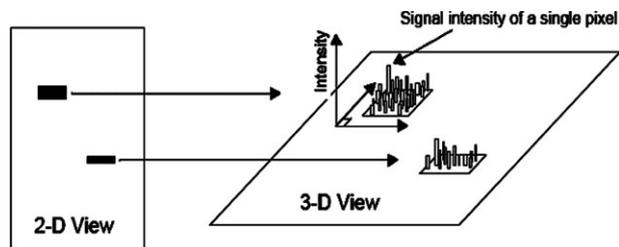


Figure 1. Representation of the pixels in two digitally imaged bands in a gel, as a 2-D (left) and three-dimensional (right) view.

X and Y coordinate, which are the pixel's horizontal and vertical positions on the image, and a Z value, which is the signal intensity of the pixel (see Fig. 1). The total intensity of a data object is the sum of the intensities of all the pixels that make up the object. The mean intensity of a data object is the total intensity divided by the number of pixels in the object.

2.2 Image filtering

This is a fundamental process, as it removes small noise features on an image while leaving larger features (like spots) unaffected. A filter wizard helps the user through this selection process, for removing specks and other imperfections in the imaged gel.

2.3 Automated spot detection

A spot detection wizard is designed to guide users through this process. One first selects the faintest spot in the scan (this will set the sensitivity and minimum peak value parameters) and then the smallest spot (this will set the size scale parameter); after that the largest spot on the image that one wants to detect is selected. Once the gel image has been created (*i.e.* each pixel of a 2-D scan is originally assigned an OD) the program will help the user through a number of steps, consisting of initial smoothing, background subtraction, final smoothing (necessary for removing extraneous spots at or near the background level), locating spots in the gel image (*i.e.* locating the center and position of each recognizable spot), fitting and quantifying spots (this step fits ideal Gaussian distributions to spot centers). As a final product, three separate images are obtained: the original raw 2-D scan, which remains unchanged; the filtered image, which is a copy of the original scan that has been filtered and processed; and the Gaussian image, which is a synthetic image containing the Gaussian spots. A representation of a filtered image, a Raw 2-D scan and a Gaussian image, is shown in Fig. 2, which refers to a 2-D map of rat serum.

2.4 Matching of protein profiles

Groups of gels can be edited and matched to one another in a match set. A match set consists of gel spots files and gel images. In a match set, protein spots are matched to each

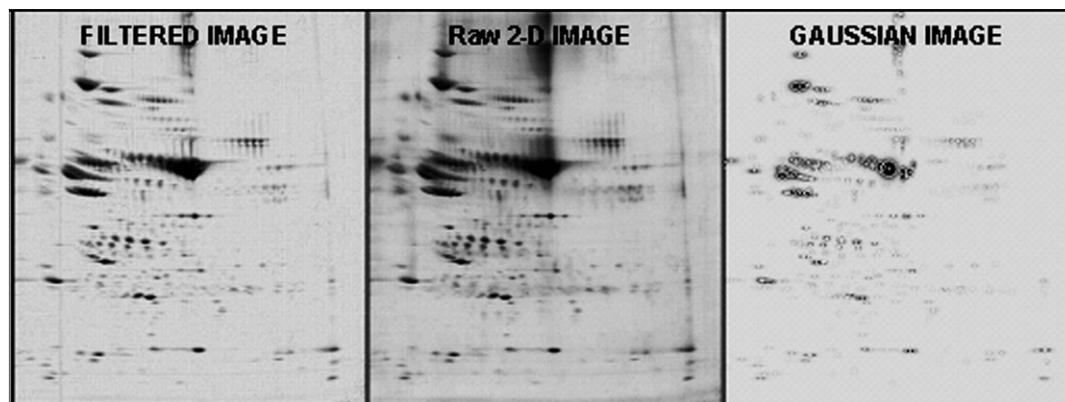


Figure 2. Representation of a filtered image, a raw 2-D scan and a Gaussian image. The data shown is from a 2-D map of rat serum.

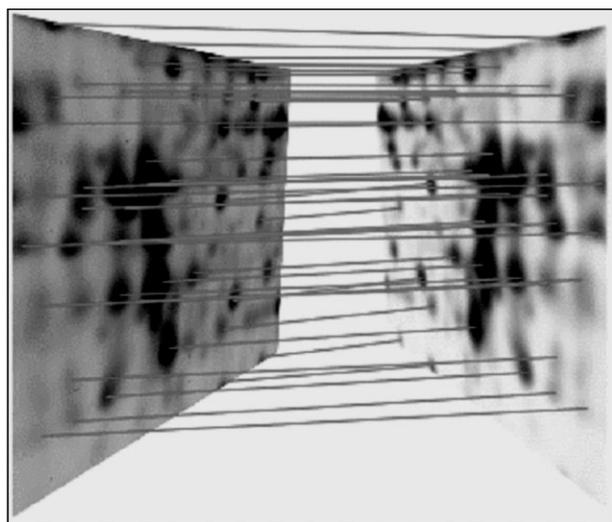


Figure 3. Schematic drawing on how protein spots from two gels are matched with PDQuest.

other, enabling the user to compare their quantities. For matching the same protein spots between different gels landmarks are needed. Landmarks are reference spots used by PDQuest to align and position match set members for matching. They are used to compensate for slight differences and distortions in the member gels. How protein spots of two gels are matched with PDQuest is shown schematically in Fig. 3.

2.5 Normalization

When comparing gels in a match set, there is often some variation in spot size and intensity between gels that is not due to differential protein expression. This variation can be caused by a number of factors, including pipetting errors during sample preparation and loading, variations in sample density, inconsistencies in staining, *etc.* To accurately com-

pare spot quantities between gels, one must compensate for these nonexpression-related variations in spot intensity. This is the process of normalization.

2.6 Differential analysis

For quantitative analysis, at least five replicate gels should be run *per* sample. A replicate group is the name given to a set of gels that are duplicates of each other. For example, five separate gels of a control cell lysate theoretically should produce the same quantitation for all spots on a 2-D gel. In practice, however, slight variations in the quantitation will probably be seen. Instead of choosing the one that the user thinks is best, one can take the average and use those values for spot quantitation. Once the sample groups have been created (for example control and drug-treated cell line), it is possible to perform the comparison between the protein profiles to find differentially expressed proteins (down-regulated, up-regulated, silenced, *etc.*). Due to the high variability from sample comparisons run in different gels, the threshold for accepting a meaningful variation is set at a factor of 2.0, *i.e.* only spots whose quantity in gel B is at least twice that of the corresponding spot in gel A are accepted as significantly changed (100% variation). Conversely, in the DIGE technique, due to the fact that both treated and control samples are run within the same gel, the error seems to be greatly reduced and spot variations of only 20% are accepted.

2.7 Statistical analysis

Once differentially expressed proteins have been detected, it is necessary to perform statistical analysis to find significant differences between the two compared samples. The statistical considerations are usually performed using the Student's *t*-test ($p < 0.05$).

Efficient analysis of protein expression by using 2-D data relies on the use of automated image processing techniques. The overall success of this research depends critically on the

accuracy and the reliability of the analysis software. In addition, the software has a profound effect on the interpretation of the results obtained, and the amount of user intervention demanded during the analysis. The choice of analysis software that best meets specific needs is therefore of interest to the research laboratory. Different packages have different strengths and weaknesses. ImageMaster (Amersham Biosciences, Uppsala, Sweden) is one of the most accurate packages, Z3 (Compugen San Jose, CA, USA) appears to have the most robust to poor S/N ratio and PDQuest (Bio-Rad, Hercules, CA, USA) the most robust to spot overlap. Melanie III (GeneBio, Geneva, Switzerland) performs well in all evaluations and Progenesis (Nonlinear Dynamics, Newcastle, UK) has the advantage of a parameter-free spot detection, whilst also performing well in most evaluations. One should not forget, however, that all the packages listed under number 5 to 5.3 in Table 1 appear to be essentially the same as Phoretix 2-D Advanced, marketed under different trade names. Furthermore, some companies might offer a range of packages to meet different experimental needs.

3 Home-made approaches: fuzzy logics

It is of interest to mention some other procedures for 2-D map comparisons that have recently become available. Although not yet elaborated into commercially available software packages, these methods appear to be just as powerful and user friendly. In one approach, as reported by Marengo *et al.* [26–29], statistical treatment of 2-D maps is performed *via* fuzzy logics [30].

It is well known that 2-DE is characterized by low reproducibility which is due to several features; complex specimen and sample pretreatment, large numbers of instrumental parameters (staining-destaining conditions, polymerization conditions of the gel for the second dimension *etc.*), large numbers of spots present on each map. The low reproducibility produces significant differences even among map replicates of the same electrophoretic run, which is reflected in changes in spots position, size and shape. It appears thus very difficult to locate precisely and univocally a spot on a map by the two coordinates x and y . In order to consider this effect, fuzzy logic principles are applied. The procedure is based on three steps: (i) image digitalization, (ii) image defuzzification, (iii) refuzzification.

3.1 Image digitalization

Each image obtained by scanning the correspondent map with a densitometer, is turned into a grid of a given step (generally 1×1 mm) containing in each cell the OD ranging from 0 to 1. A cut-off value is chosen (generally ranging from 0.3 to 0.4) in order to eliminate the contribution to the signal given by the background. All values below the cut-off value are turned into 0. The threshold value has to be chosen independently for each dataset.

3.2 Image defuzzification

The digitalized image is turned into a grid of binary values; 0 when no signal is detected; 1 when a value above the cut-off threshold is present. This step, which represents a defuzzification of each map, is focused on the elimination of the sensitivity to the destaining protocol.

3.3 Refuzzification

This step is focused on reintroduction of information due to the spatial uncertainty about the localization, size and shape of each spot on the map. Each cell containing a 1 in the digitalized image is substituted with a 2-D probability function.

The statistical distribution used in step 3, is a 2-D Gaussian function. The probability of the presence of a signal in cell x_i, y_j when a signal is present in the cell x_k, y_l is calculated by the following function:

$$f(x_i, y_j, x_k, y_l) = \frac{1}{2\pi\sigma_x\sigma_y} \cdot e^{-\frac{1}{2(1-\rho^2)} \left[\frac{(x_i-x_k)^2}{\sigma_x^2} + \frac{(y_j-y_l)^2}{\sigma_y^2} \right]} \quad (1)$$

where ρ is the correlation between the two dimensions x and y (fixed at 0, since a complete independence of the two electrophoretic runs is expected); σ_x and σ_y are two constants corresponding to the SD of the Gaussian function along each of the two dimensions. The two parameters σ_x and σ_y can be kept identical, so that the Gaussian function presents the same SD (the same uncertainty) on both dimensions (IEF/IPG followed by SDS-PAGE as the second dimension run). This behavior corresponds to an identical repeatability of the result with respect to the two electrophoretic runs. So, the parameter which is analyzed for its effect on the final result is $\sigma = \sigma_x = \sigma_y$. Alternatively, the two parameters can be set at different values, typically $\sigma_x = 1.5 \sigma_y$, corresponding to an uncertainty along the second dimension (indicated by x in our case) which is about 50% larger than that along the first dimension. The second dimension in fact often presents a larger uncertainty, due to the self-made polymerization of the gel for the second run, while the first dimension is run on commercial strips.

Changing the value of the parameter σ (or of parameters σ_x and σ_y) corresponds to modifying the distance at which an occupied cell exerts its effect: high values of σ correspond to a perturbation operating at larger distances. A low value of σ corresponds to a perturbation operating at a smaller distance, with spots that show a smaller effect on their neighbor and a more crisp image. Thus, an increase in the parameter σ means a higher fuzzyfication level of the maps; a decrease in σ corresponds instead to fuzzy maps more similar to the original 2-D PAGE. In general, best results are obtained with intermediate levels of the fuzzyfication parameters, corresponding to not too fuzzyfied maps.

The Gaussian distribution was chosen as the best function since the spots can be described as intensity/probability distributions, with the highest intensity/probability value in

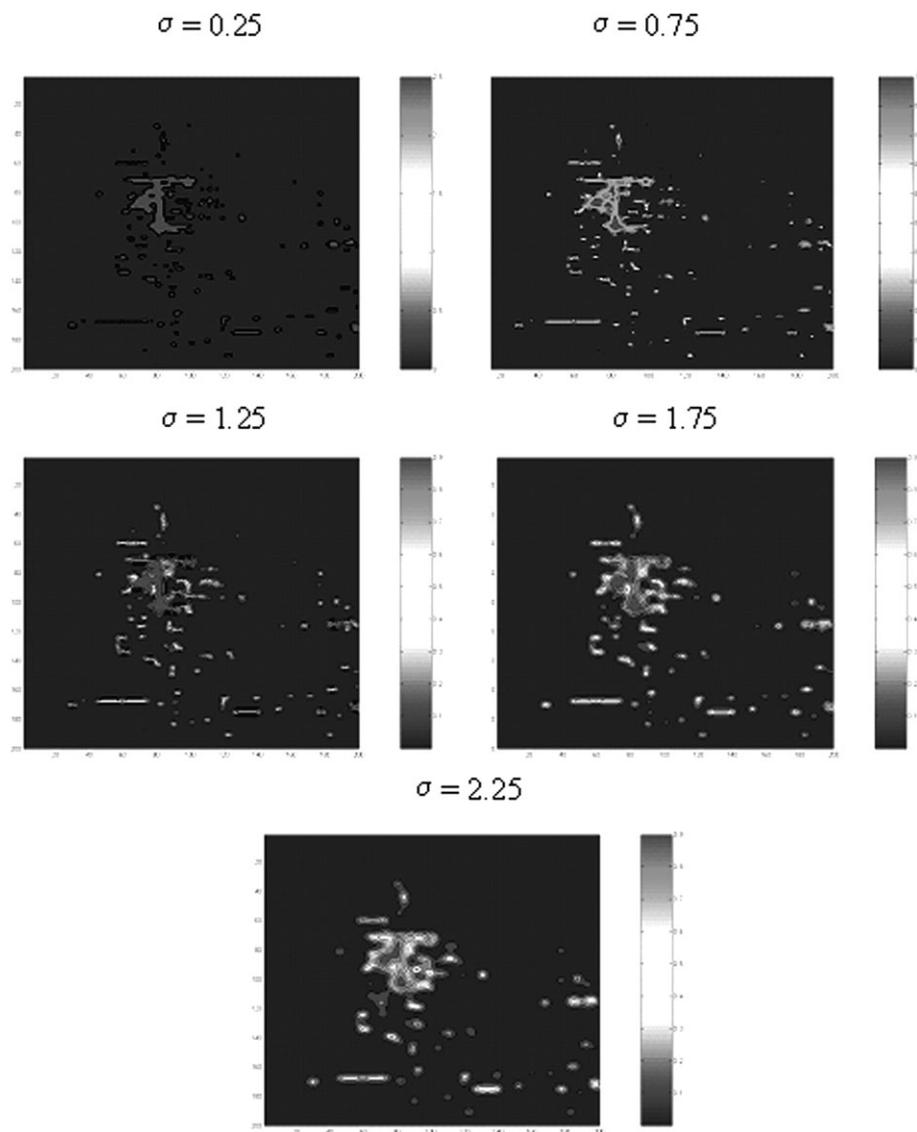


Figure 4. Example of fuzzy maps obtained for a control human lymph-node for different values of the σ parameter. Reprinted from [27] with permission.

the center of the spot itself and decreasing intensities/probabilities as the distance from the center of the spot increases. Moreover, the integral of the Gaussian function on the whole domain of the 2-D PAGE is 1, so that the total signal is blurred but is maintained quantitatively coherent. The value of the signal S_k in each cell x_i, y_j of the fuzzy map is calculated by the sum of the effect of all neighbor cells $x_{i'}, y_{j'}$, containing spots, as expressed in Eq. 2:

$$S_k = \sum_{i',j'=1,n} f(x_i, y_j, x_{i'}, y_{j'}) \quad (2)$$

The sum runs on all cells in the grid, but is dependent on the value of the parameter σ . Only the neighboring cells are affected significantly by the presence of a signal. Each digital-

ized image is thus turned into a virtual map containing in each cell the sum of the influence of all the spots of the original 2-D PAGE; these virtual maps can be called fuzzy matrices or fuzzy maps. The use of such a function corresponds to a sort of Gaussian filter applied to each map. The choice to associate the Gaussian probability function to each cell instead of to each spot is due to the presence in the maps of several complex spots, whose shape is irregular, so that the substitution of the spot, with a Gaussian probability function would not describe properly the situation. Our method allows, in principle, spots containing several proteins, a frequent hazard in 2-D maps of complex samples [31, 32], to be taken into consideration. Figure 4 shows an example of fuzzyfication of a map at different σ values. The first two steps, namely digitalization and fuzzyfication, are by all means not sufficient to exhaust a comparative 2-D map analysis. For that, different multivariate statistical tools can be applied; we use principal component analysis (PCA) coupled to classification methods and multidimensional scaling.

3.4 Fuzzy logic and PCA

The fuzzy maps obtained can be compared by means of multivariate statistical tools, *i.e.* PCA and classification methods. This approach was applied to a set of eight 2-D maps: four control samples and four samples belonging to mantle cell lymphomas. PCA is a multivariate statistical method which allows the representation of the original dataset in a new reference system characterized by new variables, called principal components (PCs). Each PC has the property of explaining the maximum possible amount of variance contained in the original dataset. The PCs, which are expressed as linear combinations of the original variables, are orthogonal to each other and can be used for an effective representation of the system under investigation, with a lower number of variables than in the original case. The

coordinates of the samples in the new reference system are called scores while the coefficient of the linear combination describing each PC, *i.e.* the weights of the original variables on each PC, are called loadings.

PCA is performed on the fuzzy maps in order to understand which parts of the 2-D PAGE maps contain the same type of information and which contain different information [33, 34]. Moreover, PCA can permit the identification of clusters and groups of samples. Finally it provides a size reduction of the dataset since from thousands of original variables only the relevant PCs can be maintained. In this case the variables are the probabilities contained in every cell of each unwrapped map (40 000 variables for a map of 200×200 cells).

Classification methods can then be applied in order to discriminate between the classes of samples involved in the study and to identify the spots responsible for the separation of the original dataset into classes. For this purpose we selected linear discriminant analysis (LDA) [35, 36] and applied it to the relevant PCs obtained from the previous step. Briefly, LDA is a Bayesian classification method that allows for the discrimination of the samples present in a dataset considering its multivariate structure. An object x is assigned to the class g for which the posterior probability $P(g|x)$ is maximum, assuming a Gaussian multivariate probability distribution:

$$P(g|x) = \frac{P_g}{(2\pi)^{p/2} |S_g|^{1/2}} e^{-0.5(x-c_g)S_g^{-1}(x-c_g)} \quad (3)$$

where P is the prior probability, S_g is the covariance matrix of class g that, in the case of LDA, is approximated with the pooled (between the classes) covariance matrix, c is the centroid of class g , and p is the number of descriptors.

The argument of the exponential function is the Mahalanobis distance between the object x and the centroid of class g , which takes into consideration the class covariance structure:

$$(x - c_g)S_g^{-1}(x - c_g) \quad (4)$$

From the logarithm of the posterior probability (by eliminating the constant terms) each object is classified in class g if it is minimum, the so-called discriminant score:

$$D(g|x) = (x - c_g)S_g^{-1}(x - c_g) + \ln |S_g| - 2 \ln P_g \quad (5)$$

The selection of the variables for the LDA model which discriminates between the classes present in the dataset was performed by a stepwise algorithm in forward search ($F_{to-enter} = 4.0$). The results of this novel method, as applied to the just described dataset of control and diseased lymphnodes, can be appreciated in the differential display shown in Fig. 5. The spots that characterize the pathological class are represented by negative values, whereas spots characterizing the control samples from healthy individuals have positive values [27].

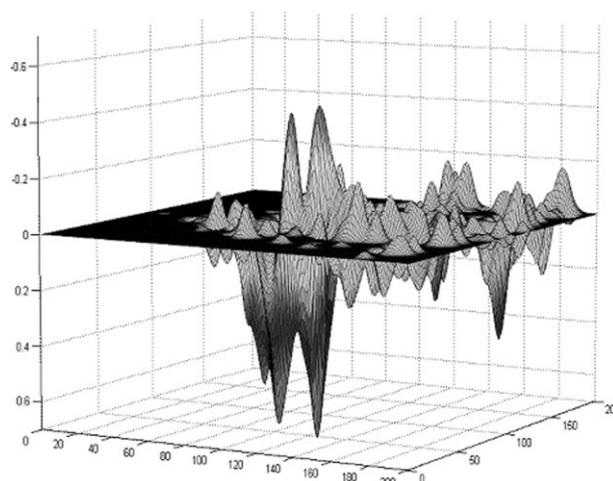


Figure 5. Display of the fuzzy map obtained as the difference of the mean healthy and pathological fuzzy maps. Reprinted from [27] with permission.

3.5 Fuzzy logic and multidimensional scaling

Fuzzy maps can also be compared by multidimensional scaling (MDS), a multivariate tool which allows a substantial dimensional reduction and effective graphical representation of the data. Given a set of objects and a measure of their similarity, MDS searches for a smaller number of dimensions in which objects can be represented as points, while matching as much as possible the distances between the objects in the new reference system with those calculated in the original one. In this case, the calculations were performed by the Kruskal iterative method [37, 38]. The search for the coordinates is based on the steepest descent minimisation algorithm, where the target function is the so-called stress (S), which is a measure of the ability of the configuration of points to simulate the original distance matrix.

For each value of the fuzzyfication parameter, a similarity matrix has to be built. Similarities between each couple of 2-D maps are calculated after matching the two fuzzy maps. From the match between the two fuzzy maps k and l , the common signal SC_{kl} (the sum of all signals present in both maps) and the total signal ST_{kl} , can be computed:

$$SC_{kl} = \sum_{i=1}^n \min(S_i^k, S_i^l) \quad (6)$$

$$ST_{kl} = \sum_{i=1}^n \max(S_i^k, S_i^l) \quad (7)$$

where n is the number of cells in the grid. The similarity index is then computed by:

$$S_{kl} = \frac{SC_{kl}}{ST_{kl}} \quad (8)$$

The similarity index ranges from 0, corresponding to two maps showing no common structure, to 1, corresponding to two identical maps. MDS was applied to a dataset of ten 2-D

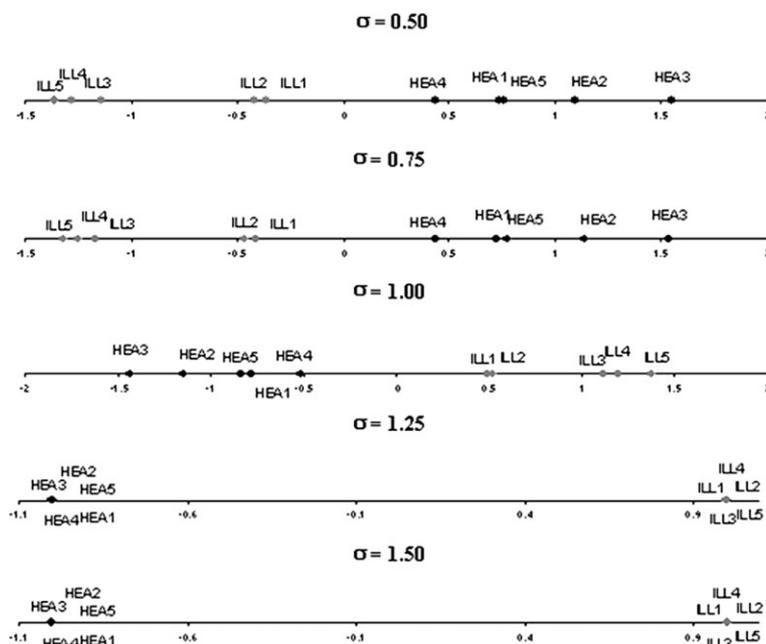


Figure 6. Rat sera dataset: MDS final configurations for different values of the σ parameter. Reprinted from [28] with permission.

maps: five belonging to control rat sera samples and five belonging to nicotine treated rat sera. MDS, performed for different values of the σ parameter, allowed the separation of the two classes of samples by means of only one dimension for all the σ values considered, as shown in Fig. 6.

4 Home-made approaches: Three-way PCA

This method, as applied to the comparison of 2-D maps of control and treated samples, consists of four steps: (i) digitalization of the image, as in all other protocols; (ii) data transformation, to scale all the samples and make them comparable; (iii) three-way PCA, to identify the classes of samples present in the dataset and to identify the zones of the maps responsible for the differences occurring between the classes; (iv) difference analysis, by which the maps, rebuilt by using the relevant factors, are compared to identify the differences occurring between the centroids of the two classes of control and treated samples.

For data transformation a normalization is essential before performing three-way PCA, in order to make all the samples comparable with each other. The chosen transformation is maximum scaling: the digitalized 2-D PAGE maps are scaled one at a time to the maximum value for each map, according to the following mathematical expression:

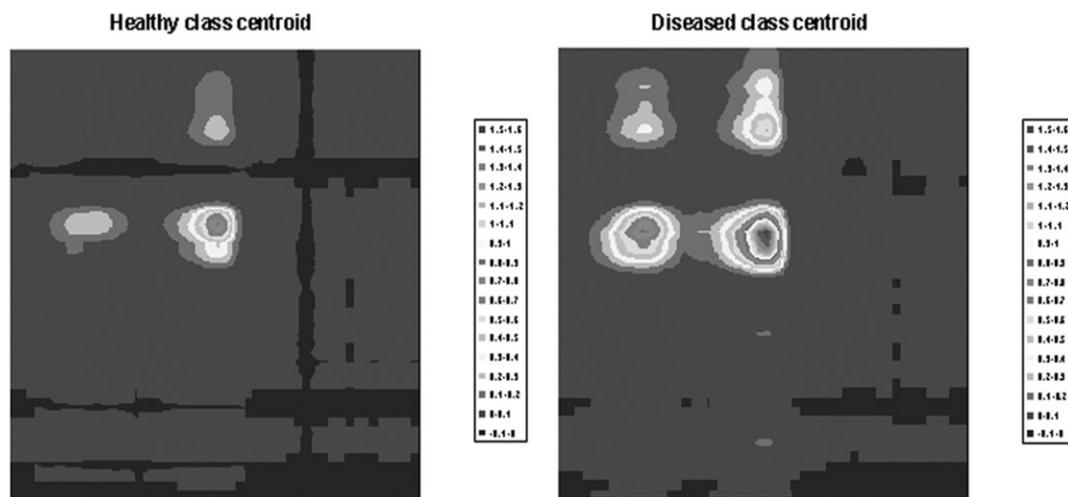
$$x_k(i,j) = \frac{X_k(i,j)}{\max(X_k)} \quad (9)$$

where $x_k(i,j)$ is the value corresponding to the cell in (i,j) position in the k -th 2-D PAGE map and $\max(x_k)$ is the maximum value in all the cells of the k -th 2-D PAGE map. By applying such a transformation to each 2-D map, the maximum signal intensity value of every 2-D PAGE map becomes a unit value; all the samples are thus ranged from 0 to 1 and the dataset becomes independent from the intensity differences due to the staining step. This scaling is suggested by the fact that the large variability of the staining procedure causes a systematic error (*i.e.* maps being consistently darker or lighter). If not removed, this error would account for the major amount of variation.

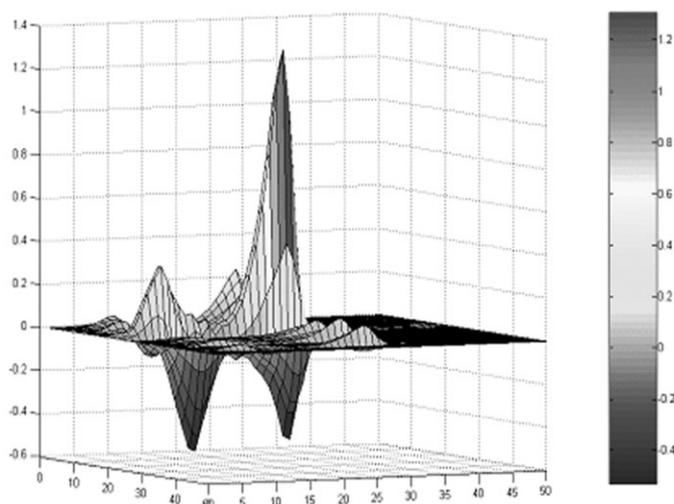
Three-way PCA, based on the Tucker-3 model [39–45], has been used for the identification of classes of samples present in the two datasets. Three-way PCA allows the three-way structure of the dataset which can be considered as a parallelepiped of size $I \times J \times K$ (conventionally defined as objects, variables and conditions), where, in our case: I is the number of rows of the grid (the x coordinates, *i.e.* M), J is the number of columns of the grids (the y coordinates, *i.e.* pH), and K is the number of samples to be taken into consideration. Three-way PCA is based on the fact that the observed modes I , J , and K can be synthesized in more fundamental modes, each element of a reduced mode expressing a particular structure existing between all or a part of the elements of the associated observation mode. The final result is given by three sets of loadings together with a core array describing the relationship among them. Each of the three sets of loadings can be displayed and interpreted in the same way as a score plot of standard PCA. Mathematically, this is expressed as follows:

$$x_{ijk} = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R a_{ip} b_{jq} c_{kr} g_{pqr} + e_{ijk} \quad (10)$$

where x_{ijk} denotes the elements of the initial matrix \underline{X} , a_{ip} , b_{jq} and c_{kr} denote reduced elements of the component matrices A , B and C of order $I \times P$, $J \times Q$ and $K \times R$ respectively, g_{pqr} denotes the elements (p, q, r) of the $P \times Q \times R$ core array \underline{G} , and e_{ijk} denotes the error term for element x_{ijk} and is an element of the $I \times J \times K$ array \underline{E} . In the case of a cubic core array (*i.e.* if $P = Q = R$), a series of orthogonal rotations can be performed on the three spaces of the three modes, looking for the common orientation for which the core array is body-diagonal as much as possible. If this condition is sufficiently achieved, *i.e.* if the elements g_{111} , $g_{222} \dots$ are the only elements of the core matrix significantly different from 0, then the rotated sets of loadings can also be interpreted jointly by overlapping them. The datasets were then analyzed with a program developed by Dr. R. Leard (Genova, Italy) in the MATLAB 6.1 (The mathworks; Natick, MA, USA) environment.



(a)



(b)

Figure 7. Rat sera dataset. Results of difference analysis: (a) control and nicotine-treated centroids; (b) map obtained as the difference of the mean control and nicotine-treated maps. Reprinted from [47] with permission.

Three-way PCA was applied successfully to the just described datasets of human lymph-nodes and rat sera [46]; in both cases, the difference analysis allowed the identification of the under- and over-expressed proteins discriminating the different classes. The result of difference analysis performed on the rat sera dataset is reported in Fig. 7. Under- and over-expressed proteins can be identified by a matching procedure between the map given by the difference analysis and the original 2-D maps. Certainly, this approach has the disadvantage of loss of information about the exact localization of the under- and over-expressed proteins, which are no longer described by pI and M_r coordinates. On the other hand, the low reproducibility affecting 2-DE can be taken into consideration.

5 Home-made approaches: PCA coupled to classification methods and cluster analysis

As a variant of the above methods, the comparison of classes of 2-D maps can be performed on the spot volume dataset obtained as a result of the PDQuest analysis. In this case the variables are the volumes of the spots revealed by PDQuest analysis. Classification methods and cluster analysis techniques can then be applied to the significant PCs in order to identify the differences occurring between the classes in terms of under-expressed and over-expressed spots.

5.1 PCA and soft-independent model of class analogy

Soft-independent models of class analogy (SIMCA) [47, 48] is a classification method based on PCA. This method builds boxes containing the samples of each class, on the basis of a class dependent PC model. Each object is then assigned to a class by calculating its distance from the different PC models obtained for the samples of each class. Since this method is based on PCs and not on the original variables, the main advantage of its application is that the objects are assigned to each class by considering only the systematic and useful information contained in the dataset, thus overlooking the unresponsive ones. Moreover the method is also suitable for investigating small datasets. This depends on the possibility of performing a substantial size reduction of the descriptors through the calculation of the PC. In the comparison of sets of gels from control and treated samples, SIMCA allows for the identification of modelling and discriminating spots, *i.e.* of those spots which are differentially expressed in the two sets of samples under comparison.

Combined PCA and SIMCA were applied to the analysis of control and neuroblastoma murine samples, in order to detect changes in protein expression between the two sets of samples [49]. Very good correlation was found between the data obtained by analysis of 2-D maps *via* the commercial software PDQuest and the present PCA-SIMCA data reduction. In both cases, the comparison between such maps showed up- and down-regulation of 84 polypeptide chains, out of a total of 700 spots detected by a fluorescent stain, Sypro Ruby (see Figs. 8A and B). In Fig. 8A, the spots that are presented in a color scale from light to dark red are those that exhibit an increasing discriminating power; the spots marked as black circles are not effective in the discrimination of the two classes. This representation sums up the information about the up- and down-regulated spots responsible for the differences occurring between the two classes investigated (control *vs.* neuroblastoma specimens). In Fig. 8B, the same set of 84 modulated spots, as detected by PDQuest analysis, is highlighted by red ovals. Spots that were differentially expressed between the two groups, were analyzed by MALDI-TOF MS and 14 of these spots were identified (indicated by arrows in Fig. 8B).

5.2 PCA and cluster analysis

Cluster analysis techniques allow the relationships between the objects or the variables of a dataset to be investigated, in order to recognize the existence of groups. The methods most frequently used belong to the class of the agglomerative hierarchical methods [50], where the objects are grouped (linked together) on the basis of a measure of their similarity. The most similar objects or groups of objects are linked first. The result of such analyses is a graph, called a dendrogram, where the objects (X axis) are connected at decreasing levels of similarity (Y axis). The results of hierarchical clustering methods depend on the specific measure of similarity and on

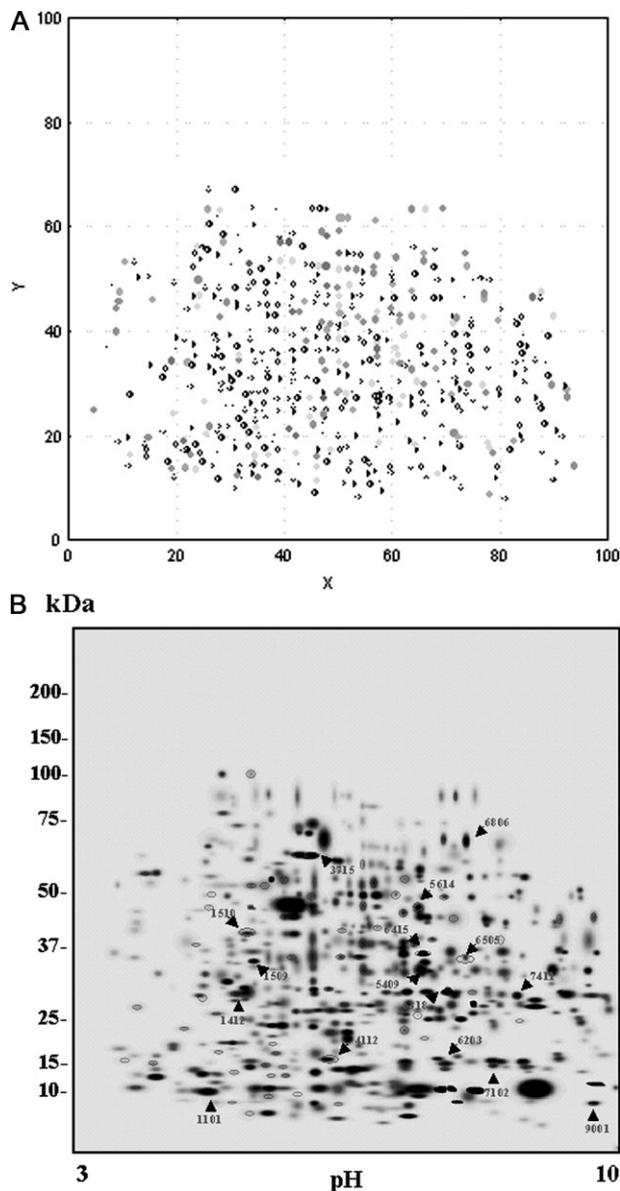


Figure 8. Data reduction of control *vs.* neuroblastoma murine samples *via* (A) combined PCA-SIMCA data handling and (B) PDQuest analysis. In (A) the red spots of different color intensity represent polypeptide chains which are differentially expressed in the two sets of samples; the same spots, in (B) are highlighted by red ovals. The black arrows indicate the 14 spots that have been identified, so far, *via* elution, digestion and MALDI-TOF MS analysis (from [50] with permission).

the linking method. In contrast to the classical cluster analysis performed on the original variables (in this case the volumes of the spots detected by PDQuest), it can be applied to the relevant PCs.

This method was applied to a dataset consisting of by 18 samples belonging to human pancreatic cancer cells from two cell lines (Paca44 and T3M4) with or without exposure to Trichostatin-A (TSA), a potential chemotherapeutic agent

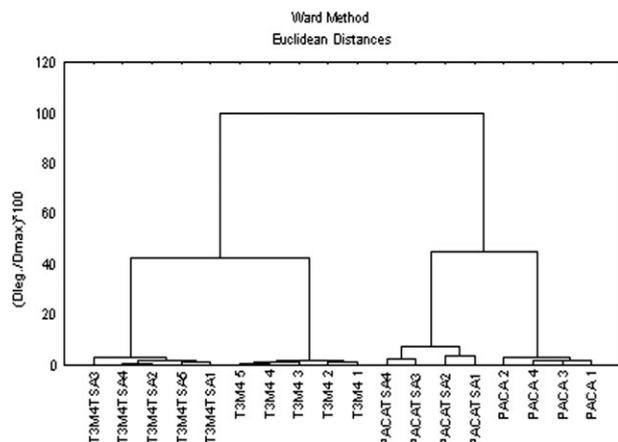


Figure 9. Dendrogram calculated on the basis of the first three PCs (Ward method, Euclidean distances) for a human pancreatic cancer cell dataset. Reprinted from [52] with permission.

[51]. The first three PCs accounted for the main differences existing between the four groups: PC₁ explains the information related to the two cell lines; PC₂ carries the information about the TSA effect, mainly for the Paca44 cell line; PC₃ carries the information about cellular sensitivity to TSA, mainly for the T3M4 cell line. Cluster analysis was then performed on the three relevant PCs by means of the Ward method (Euclidean distances were used in the calculations). Cluster analysis revealed the clear presence of four groups of samples: the samples are divided between the two cell lines and then into diseased and treated samples, as can be seen by the dendrogram in Fig. 9. As in the previous case, the up- and down-regulated spots could be identified by means of PCA.

6 Concluding remarks

We have reviewed here a number of approaches to quantitative proteomics in 2-D map analysis. Replicas of 2-D maps, stained with colloidal Coomassie and then analyzed and matched by software able to detect up- and down-regulation (and appearance or disappearance or spots) of proteins *via* differential dye uptake, although terribly time consuming and labor intensive, is still a good and reliable method. We have nicknamed it the peones approach, since one of its characteristic is its relatively inexpensive set-up and low cost, making it suitable for laboratories surviving on a tight budget. In addition to an evaluation of the various software packages in use today (as listed in Table 1), we have also reviewed a number of home-made approaches that, although not yet available as fully automated software programs, also appear to be quite powerful, user friendly and comparable to some of the presently available systems. These methods have some advantages with respect to the classical approach: first, they do not involve human intervention and no operator dependent alignment step is

required; moreover, all replicate maps are taken into consideration and the evaluation of the differences between classes of 2-D maps is carried out on all the replicates instead of on the master gels.

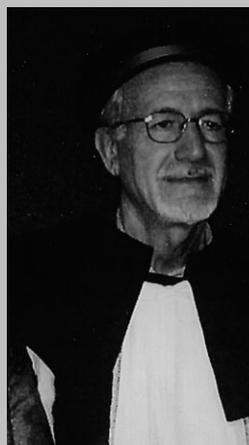
The approaches described in Section 5 show another important advantage, related to the possibility of identifying the differences between groups of samples smaller than the classical 2-fold: this is achieved by means of PCA, which also allows the identification of small differences due to its peculiar robustness. The main disadvantage, on the other hand, is presented by the techniques based on fuzzy logic, in which the spot identity is lost. We have left out the DIGE approach on purpose, since in this methodology the sample are mixed and run on a single gel and therefore require approaches for differential display completely different from those adopted for samples run in separate sets of gels. The DIGE technology, either *via* minimal (Lys residues) [11, 12] or saturation (Cys residues) [53] labelling, may be a powerful approach, but requires expensive equipment and software packages and quite an expensive reagent kit. Even this elitist system, though, might be prone to problems, especially in the Cys-tagging procedure [52], where protein spot identification *via* MS appear to be problematic and where a number of barely soluble proteins might precipitate out of solution due to the increased hydrophobicity from the cyanine dyes, thus disappearing from the map just at the onset of the 2-D mapping procedure.

PGR is supported by FIRB-2001 (grant no. RBNE01KJHT), by PRIN-2003 from MIUR (Rome, Italy) and by AIRC (Associazione Italiana per la Ricerca sul Cancro). EM is supported by Regione Piemonte, Ricerca Scientifica Applicata, DD. RR. No. 78-9416 of 19-May-2003 and No. 111-20277 of 1-August-2003.

7 References

- [1] Rosengren, A. T., Salmi, J. M., Aittokallio, T., Westerholm, J. *et al.*, *Proteomics* 2003, 3, 1936–1946.
- [2] Raman, B., Cheung, A., Marten, M. R., *Electrophoresis* 2002, 23, 2194–2202.
- [3] Mahon, P., Dupree, P., *Electrophoresis* 2001, 22, 2075–2085.
- [4] Rubinfeld, A., Keren-Lehrer, T., Hadas, G., Smilansky, Z., *Proteomics* 2003, 3, 1930–1935.
- [5] Panek, J., Vohradsky, J., *Electrophoresis* 1999, 20, 3483–3491.
- [6] Pleissner, K. P., Hoffman, F., Kriegel, K., Wenk, C. *et al.*, *Electrophoresis* 1999, 20, 755–765.
- [7] Voss, T., Haberl, P., *Electrophoresis* 2000, 21, 3345–3350.
- [8] Cutler, P., Heald, G., White, I. R., Ruan, J., *Proteomics* 2003, 3, 392–401.
- [9] Molloy, M. P., Brzezinski, E. E., Hang, J., McDowell, M. T., VanBogelen, R. A., *Proteomics* 2003, 3, 1912–1919.
- [10] Moritz, B., Meyer, H. E., *Proteomics* 2003, 3, 2208–2220.

- [11] Gharbi, S., Gaffney, P., Yang, A., Zvelebil, M. J. *et al.*, *Mol. Cell. Proteomics* 2002, 2, 91–100.
- [12] Zhou, G., Li, H., DeCamps, D., Chen, S. *et al.*, *Proteomics* 2002, 1, 117–122.
- [13] Sechi, S., *Rapid Commun. Mass Spectrom.* 2002, 16, 1416–1421.
- [14] Gehanne, S., Cecconi, D., Carboni, L., Righetti, P. G. *et al.*, *Rapid Commun. Mass Spectrom.* 2002, 16, 1692–1700.
- [15] Garrells, J. I., *J. Biol. Chem.* 1979, 254, 7961–7977.
- [16] Garrells, J. I., Farrar, J. T., Burwell IV, C. B., in: Celis, J. E., Bravo, R. (Eds.), *Two-Dimensional Gel Electrophoresis of Proteins*, Academic Press, Orlando, FA, USA 1984, pp. 38–91.
- [17] Garrells, J. I., *J. Biol. Chem.* 1989, 264, 5269–5282.
- [18] Anderson, N. L., Taylor, J., Scandora, A. E., Coulter, B. P., Anderson, N. G., *Clin. Chem.* 1981, 27, 1807–1820.
- [19] Appell, R. D., Hochstrasser, D. F., Roch, C., Funk, M. *et al.*, *Electrophoresis*, 1988, 9, 136–142.
- [20] Appell, R. D., Hochstrasser, D. F., Funk, M., Vargas, J. R. *et al.*, *Electrophoresis*, 1991, 12, 722–735.
- [21] Pun, T., Hochstrasser, D. F., Appell, R. D. *et al.*, *Applied Theor. Electrophor.* 1988, 1, 3–9.
- [22] Lemkin, P. F., Lipknin, L. E., *Comput. Biomed. Res.* 1981, 14, 272–297.
- [23] Rowlands, D. G., Flook, A., Payne, P. I., van Hoff, A. *et al.*, *Electrophoresis* 1988, 9, 820–830.
- [24] Righetti, P. G., *Immobilized pH Gradients: Theory and Methodology*, Elsevier, Amsterdam, 1990, pp. 226–231.
- [25] Chackravarti, D. N., Chakravarti, B., Moutsatsos, I., *Computational Proteomics*, 2002, 32, S4–S15.
- [26] Marengo, E., Robotti, E., Gianotti, V., Righetti, P. G., *Ann. Chim.* 2003, 93, 105–116.
- [27] Marengo, E., Robotti, E., Righetti, P. G., Antonucci, F., *J. Chromatogr. A* 2003, 1004, 13–28.
- [28] Marengo, E., Robotti, E., Gianotti, V., Righetti, P. G. *et al.*, *Electrophoresis* 2003, 24, 225–236.
- [29] Marengo, E., Robotti, E., Cecconi, D., Scarpa, A., Righetti, P. G., FUZZ-IEEE 2004 Proceedings, Budapest, Hungary, 2004, 1, 359–364.
- [30] Otto, M., *Anal. Chim. Acta* 1993, 283, 500–507.
- [31] Pietrogrande, M. C., Marchetti, N., Dondi, F., Righetti, P. G., *Electrophoresis* 2002, 23, 283–291.
- [32] Pietrogrande, M. C., Marchetti, N., Dondi, F., Righetti, P. G., *Electrophoresis* 2003, 24, 217–224.
- [33] Brereton, R. G. (Ed.), *Chemometrics: Application of Mathematics and Statistics to Laboratory Systems*, Ellis Norwood, New York, USA 1990.
- [34] Massart, D. L., Vandeginste, B. G. M., Deming, S. N., Michotte, Y., Kaufman, L. (Eds.), *Chemometrics: a Textbook*. Elsevier, Amsterdam, The Netherlands 1988.
- [35] Eisenbeis, R. A. (Ed.), *Discriminant Analysis and Classification Procedures: Theory and Applications*. Lexington, USA 1972.
- [36] Klecka, W. R. (Ed.), *Discriminant Analysis*, Sage Publications, Beverly Hills, USA 1980.



Prof. Pier Giorgio Righetti studied chemistry at the University of Pavia and spent four years as a post-doc. at MIT and Harvard Medical School (USA). He is on the Editorial Board of several international scientific journals in the field of biochemistry and Proteome analysis, such as *Electrophoresis*, *Journal of Chromatography*, *Journal of Capillary Electrophoresis*, *Journal of Biochemical and Biophysical Methods*, *Proteomics* and *Journal of Proteome Research*. He has published >620 articles in international journals and written/edited ten books,

among them “The Proteome Revisited” (Elsevier, Amsterdam, 2001). He has developed a number of important methodologies in the field of biochemistry and proteome analysis, such as isoelectric focusing in immobilized pH gradients, multicompartment electrolyzers with isoelectric membranes, membrane-trapped enzyme reactors operating in an electric field, temperature-programmed capillary electrophoresis. Associated to these methodologies are a number of developments in the field of organic chemistry of monomers and polymers, such as the chemistries of Immobilines (acrylamido weak acids and bases) and novel, *N*-substituted polyacrylamides endowed with unique hydrophilicity and stability in alkaline milieux.

His research interests include protein purification and crystallization, proteome analysis, screening for genetic defects and DNA-based diagnosis. Currently his research focuses on a number of human cancer projects (neuroblastomas, mantle cell lymphomas, pancreatic and ovary tumors and their resistance to drugs), studied *via* proteomic tools (2-D maps, mass spectrometry). In addition, a number of neurological disorders are currently under active investigation (*e.g.*, sporadic Creutzfeldt-Jacob disease, prion protein diseases) as well as studies on food allergens. He is currently Full Professor of Biochemistry at the University of Verona and President of IPSo (Italian Proteome Society, founded in Verona in May 2004).

- [37] Shepard, R. N., *Psychometrika* 1962, 27, 219–246.
- [38] Kruskal, J. B., *Psychometrika* 1964, 29, 1–27.
- [39] Tucker, L. R., *Psychometrika* 1966, 31, 279–311.
- [40] Kroonenberg, P. M. (Ed.), *Three-mode Principal Component Analysis*, DSWO Press, Leiden, The Netherlands 1983.
- [41] Geladi, P., *Chemometrics Intell. Lab. Syst.* 1989, 7, 11–30.
- [42] Kroonenberg, P. M., *Acta Oecol.* 1989, 10, 245–256.
- [43] Henrion, R., *J. Chemometr.* 1993, 6, 477–494.
- [44] Henrion, R., *Chemometrics Intell. Lab. Syst.* 1994, 25, 1–23.
- [45] Henrion, R., Andersson, C. A., *Chemometrics Intell. Lab. Syst.* 1999, 47, 189–204.
- [46] Marengo, E., Leardi, R., Robotti, E., Righetti, P. G. *et al.*, *J. Proteome Res.* 2003, 2, 351–360.

- [47] Wold, S., *Kem. Kemi.* 1982, 9, 401–411.
- [48] Frank, I. E., Lanteri, S., *Chemometrics Intell. Lab. Syst.* 1989, 5, 247–257.
- [49] Marengo, E., Robotti, E., Righetti, P. G., Campostrini, N., Pascali, J., Ponzoni, M., Hamdan, M., Astner, H., *Clin. Chim. Acta* 2004, 345, 55–67.
- [50] Massart, D. L., Kaufman, L., in: Elving, P. J., Winefordner, J. D. (Eds.), *The interpretation of Analytical Chemical Data by the Use of Cluster Analysis*, Wiley, New York, USA 1983.
- [51] Marengo, E., Robotti, E., Cecconi, D., Hamdan, M. *et al.*, *Anal. Bioanal. Chem.* 2004, 379, 992–1003.
- [52] Kondo, T., Seike, M., Mori, Y., Fujii, K. *et al.*, *Proteomics* 2003, 3, 1758–1766.

REGULAR ARTICLE

Two-dimensional mapping as a tool for classification of green coffee bean species

Maria Teresa Gil-Agusti¹*, Natascia Campostrini²*, Lello Zolla³, Corrado Ciambella³, Carlo Invernizzi⁴ and Pier Giorgio Righetti²

¹ Area Química Analítica, CCEE, ESTCE, Universitat Jaume I, Castellon, Spain

² Laboratory of Proteome Analysis, University of Verona, Verona, Italy

³ Department of Environmental Sciences, University of Tuscia, Viterbo, Italy

⁴ Caffè Invernizzi, Via G. Compagnoni 29, Milano, Italy

Two species of the genus *Coffea*, *Coffea arabica* (Colombia) and *Coffea canephora* (Indiano Robusta) were analysed by two-dimensional (2-D) maps in order to obtain fingerprints of the expressed polypeptide chains and to determine which ones would characterize the two species. Green beans were milled under liquid nitrogen. A dry powder was produced by three different extraction protocols aimed at eliminating interfering substances (polyphenols). A reduced powder was produced by two successive extractions performed in acetone. Trichloroacetic acid (TCA; 10% w/v) and β -mercaptoethanol (0.07% v/v) in acetone were used for the first extraction (a) and 10% w/v TCA in acetone was used for the second extraction (b). Proteins were then solubilized in a solution (40 μ L per 1 mg powder) containing 7 M urea, 2 M thiourea, 3% w/v 3-(3-cholamidopropyl-dimethyl-amino)-1-propanesulfate, 1% v/v carrier ampholytes, 40 mM Tris, 5 mM tributylphosphine and 10 mM acrylamide as alkylating agent. Following incubation at room temperature for 1 hour and centrifugation (7000 rpm for 20 minutes), the supernatant was used for 2-D electrophoresis. The proteins were revealed by Sypro Ruby staining. Master maps of the five replicas of each species were compared by PDQuest analysis. The results of this differential proteome analysis were: sixteen proteins were expressed solely in *C. canephora* (var. Indiano Robusta) and five proteins were only found in *C. arabica* (var. Colombia). Another eight proteins were up-regulated in *C. canephora* (var. Indiano Robusta) in comparison to *C. arabica* (var. Colombia) and one was down-regulated in the same comparison. A number of these polypeptide chains were further characterized by mass spectrometry in the matrix-assisted laser desorption/ionisation-time of flight mode. Additionally, considering the low number of protein sequences of *Coffea* present in the databases we also investigated some spots with a more powerful tool, reversed phase-high-performance liquid chromatography-electrospray ionisation-tandem mass spectrometry, thus obtaining an internal peptide sequence. The general properties of the identified proteins are presented and discussed.

Received: July 14, 2004
Revised: September 9, 2004
Accepted: September 10, 2004

Keywords:

Green coffee beans / Two-dimensional maps

Correspondence: Professor Pier Giorgio Righetti, Department of Agricultural and Industrial Biotechnologies, University of Verona, Strada Le Grazie 15, 37134 Verona, Italy
E-mail: righetti@sci.univr.it
Fax: +39-045-8027929

Abbreviations: AA, acrylamide; AUC, Acetic acid-Urea-CTAB; CTAB, cetyltrimethylammonium bromide; FA, formic acid; TBP, tributylphosphine

1 Introduction

Ever since its introduction in Europe in the second half of the sixteen century [1] (probably spreading out from Venice, in those days aptly considered the emporium of Orient), coffee beans have been steadily growing in importance, due to their

* These authors contributed equally.

capillary diffusion first as an expensive beverage in coffee houses and then as a common household item. Today, this crop is believed to be one of the most important primary materials in the world, in terms of revenue generated, coming second only to petrol, in this respect [2]. The > 150 varieties grown around the world can be grouped into two species of the genus *Coffea* (family Rubiaceae), *Coffea arabica* (var. Colombia) and *Coffea canephora* (var. Indiano Robusta). Since the beginning of the sixties, with the advent of GC and MS, coffee has been extensively studied, with the aim in particular of isolating and characterizing the substances which impart to it its unique flavor. To date, > 850 compounds have been identified in the flavor of roasted coffee, and > 300 in the smell of green coffee [3], no doubt also due to the advent of the 2-D technique GC \times GC, which has greatly helped in isolating minor species hidden in major peaks [4].

However, very few studies have investigated the protein content of coffee beans, especially in regard to the use of 2-D maps, although it is well known that free amino acids, peptides and proteins are essential aroma precursors [5]. In this respect, perhaps the cornerstone study is the one by Rogers *et al.* [6], who reported the first detailed investigation of proteins in the *C. arabica* endosperm by 2-D map analysis, coupled to MS characterization of the most abundant polypeptide chains and mRNA amplification. In their study, the most abundant polypeptide spots observed on mature coffee grain 2-D profiles were found to be subunits of the same protein, which exist as multiple isoforms with varying pI values. Strong sequence similarities were found to the 11S family of plant storage proteins. Their structure is typical of the 11S type, which occurs as a precursor of a 55 kDa polypeptide, and is observed on 2-D profiles, under denaturing and reducing conditions, in the form of cleavage products at approximately 22 kDa (β arm) and 33 kDa (α arm). This 11S-like-protein accounts for about 45% of the total grain protein.

In another study, Ludwig *et al.* [7] investigated peptides and proteases in green coffee beans of both *C. arabica* and *C. canephora* by 2-D maps and *in situ* zymogramming. They were able to detect at least five to seven enzyme spots in all coffee samples tested, whose activities were stable for several months in intact beans. Moreover, by using Schagger and von Jagow's buffer system [8] in the second dimension gel, they could resolve peptides typically lost in conventional 2-D maps in their gel slabs which ranged in size from as low as 3500 Da up to 10 000 Da. By a model roasting carried out with these coffee peptides, as well as with peptide-dextrose mixtures, they could confirm the assumption that peptides contribute to the roasting flavor of coffee and build up a map of intensities of sensory impression.

In another line of research, Montavon *et al.* [9, 10] studied the evolution of green coffee protein profiles upon maturation and the changes in these protein profiles upon roasting, both *via* 2-D map analysis. In the first study [9], they confirmed, in both ripe and unripe samples, the presence of the α and β arms, as subunits of the 11S-like protein, as found by

Ludwig *et al.* [7], with no marked differences in the 2-D profiles between the two classes of beans. However, compared to ripe beans, unripe specimens were found to be more sensitive to oxidation of chlorogenic acids to such an extent that even upon incubation under anaerobic conditions, massive losses of α and β arm polypeptides could be detected in 2-D maps, with production of low M_r peptides. When investigating the changes in protein profiles upon roasting [10], however, not much could be deduced from 2-D map analyses, since, when reaching the high temperatures of a proper roasting process (240°C), essentially all polypeptide chains disappeared from the 2-D maps. In these studies no attempts were made to use 2-D maps for phylogenetic studies, *i.e.* for characterizing different coffee species *via* differential, quantitative proteomic analysis to classify the different varieties *via* polypeptide chains present in only single species and/or up- or down-regulated in the various cultivars. We report here the first such attempt, *via* a preliminary study on two species of the genus *Coffea*, *C. arabica* (var. Colombia) and *C. canephora* (var. Indiano Robusta).

2 Materials and methods

2.1 Chemicals and materials

Tris, mineral oil, DL-DTT, Tween 20, cetyltrimethylammonium bromide (CTAB) and EDTA were purchased from Sigma (Steinheim, Germany). Glycerol, methanol, ethanol, acetic acid and acetone were from Merck (Darmstadt, Germany). Acrylamide/Bis solution (40%), TEMED, acrylamide (AA), ammonium persulfate, the Protean IEF Cell, the GS-710 Densitometer, the Versa Doc Scanner, PDQuest Version 6.2 software as well as the linear Immobiline dry strips pH gradient 3–10 (17 cm long), were obtained from Bio-Rad (Hercules, CA, USA); glycine, SDS, iodoacetamide, urea, thiourea, tributylphosphine (TBP) and CHAPS were from Fluka (Buchs, Switzerland). Bromophenol blue, carrier ampholytes and agarose were purchased from Pharmacia Diagnostics (Uppsala, Sweden).

2.2 2-DE

2.2.1 Plant material

Indiano Robusta green coffee samples (*C. canephora* var. Indiano Robusta type Cherry AB) were obtained from Western India. Cherries were harvested in January 2002 and exported to Italy in March 2002. The beans were picked at full ripening (*ca.* 9–11 months on the plant). After removing the bean from the epicarp, it was subjected to natural desiccation (sun-dried) for *ca.* 12–15 days until its humidity reached 12%. Colombia green coffee samples (*C. arabica* var. Colombia type Supreme) were obtained from Medellin, Western Colombia. Cherries were harvested from the plantation on April 2002 and exported to Italy in June 2002. The

beans were picked at full ripening (*ca.* 6–8 months on the plant). After harvesting, the bean was removed from the epicarp by a machine treatment. After washing, the mucilage was removed by enzyme action. A subsequent washing step removed the enzymes and the pergamin, a thin skin covering the bean. Finally, the beans were dried. These samples were kindly provided by C. Invernizzi (Casa del Caffè, Milano, Italy). Green beans from *C. arabica* and *C. canephora*, divided in two classes, constituted the dataset used in this study: (i) green beans from *C. arabica* (var. Colombia; the pool used to perform six 2-D maps; Fig. 1a); (ii) green beans from *C. canephora* (var. Indiano Robusta; the pool used to perform six 2-D maps; Fig. 1b). The procedure was repeated three times, thus producing three genuine replicates for each map.

2.2.2 Protein extraction and preparation of samples for electrophoresis

Grains were finely ground in liquid nitrogen. A dry powder was produced by three different extraction protocols aimed at eliminating interfering substances (such as polyphenols).

2.2.2.1 Acetone/TCA/ β -mercaptoethanol solution extraction

A dry, reduced powder was produced by two successive extractions performed in acetone. TCA (10% w/v) and β -mercaptoethanol (0.07% v/v) in acetone were used for the first extraction (a) and 10% w/v TCA in acetone was used for the second extraction (b), according to the method of Damerval *et al.* [11]. Proteins were then solubilized in a solution (40 μ L per 1 mg powder) containing 7 M urea, 2 M thiourea, 3% w/v CHAPS, 1% v/v carrier ampholytes (pH 3–10), 40 mM Tris, 5 mM TBP, 10 mM AA [12]. Following incubation at room temperature for 1 h (alkylation was blocked with 10 mM DTT) and centrifugation (7000 rpm for 20 min, 4°C) the supernatant was stored at -20°C until use.

2.2.2.2 Lysis solution extraction

Green coffee samples were milled under nitrogen and extracted in solubilization/lysis buffer (1 mL per 200 mg powder) containing 7 M urea, 2 M thiourea, 3% w/v CHAPS, 1% v/v carrier ampholytes (pH 3–10), 40 mM Tris, 5 mM TBP, 10 mM AA [12]. Extraction was allowed to proceed for 1 h at room temperature under continuous shaking and alkylation was blocked with 10 mM DTT. Extracts were then centrifuged (7000 rpm, 20 min, 4°C) and supernatants were stored at -20°C until use.

2.2.2.3 Acetic acid/Urea/CTAB solution extraction

Green coffee samples were milled under nitrogen and extracted in a solution (1 mL per 200 mg powder) containing 0.1 M acetic acid, 3 M urea, 0.01% CTAB. Extraction was

allowed to proceed for 2 h at room temperature with continuous shaking. Extracts were then centrifuged (13 000 rpm, 25 min, 4°C) and supernatants were precipitated in an anhydrous solution of acetone and methanol in an 8:1 v/v ratio. The samples were kept at -20°C for 2 h and then centrifuged at 13 000 rpm for 30 min. The supernatant was discarded and the pellet was resuspended in a solution (40 μ L per 1 mg powder) containing 7 M urea, 2 M thiourea, 3% w/v CHAPS, 1% v/v carrier ampholytes (pH 3–10), 40 mM Tris, 5 mM TBP, 10 mM AA. Following incubation at room temperature for 1 h (alkylation was blocked with 10 mM DTT) and centrifugation (7000 rpm for 20 min, 4°C) the supernatant was stored at -20°C until use. We refer to this method as the acetic acid/urea/CTAB (AUC) extraction method.

2.2.3 IEF in IPG strips

The first dimension run was performed on strips (17 cm length, 0.5 mm thickness) with a linear pH gradient from pH 3 to pH 10 [13]. For analytical runs (used to find differentially expressed proteins), the IPG strips were rehydrated with 3 mg/mL of pooled samples (six strips with *C. arabica* samples and six strips with *C. canephora* samples) and traces of bromophenol blue for monitoring the electrophoretic run. For preparative runs (used to isolate proteins from the gel for identification with MS), the IPG strips were rehydrated with 4.5 mg/mL of pooled samples. Passive gel rehydration was allowed to continue for 8 h before the focusing step. IEF was carried out with a Protean IEF Cell (Bio-Rad), with a low initial voltage and then by applying a voltage gradient up to 10 000 V with a limiting current of 50 μ A/strip. The total product time \times voltage applied was 75 000 Vh for each strip and the temperature was set at 20°C .

2.2.4 Interfacing the IPG strips with the denaturing SDS solution and SDS-PAGE

Each strip was equilibrated with an SDS denaturing solution containing 6 M urea, 2% SDS, 20% glycerol, 0.375 M Tris-HCl (pH 8.8). The contact lasted for 30 min in tubes containing 20 mL each of the equilibration solution. Each strip was then interfaced with a gel slab using 0.5% agarose solubilized in cathodic buffer (192 mM glycine, 0.1% SDS, Tris to pH 8.3). The gel slabs were cast with a two-vessel gradient mixer, with total sufficient volumes for polymerizing 12 gel slabs of 1.5 mm thickness and with a porosity gradient from 7%T to 20%T. Polymerization took place overnight. The second dimension run was performed using a PROTEAN II xl Multi-Cell (Bio-Rad). The cathodic buffer was the same as stated above; the anodic buffer was a solution of 0.375 M Tris-HCl, pH 8.8. The electrophoretic run was performed by setting a current of 2 mA for each gel for 1 h, then 5 mA/gel for 2 h and 10 mA/gel until the end of the run for a total of 20–22 h. During the whole run the temperature was kept constant at 11°C .

2.3 Fluorescent staining

All gels were fixed with a solution containing 40% ethanol and 10% acetic acid for 30 min. After this step the gels were stained with Sypro Ruby (70 mL for each gel) overnight and then destained with 10% methanol, 7% acetic acid for 1 h [14]. The gels were then washed with milliQ water (Millipore, Bedford, MA, USA). Samples were digitized with a Versa Doc Scanner. The gels used to isolate proteins for identification with MS were stained with colloidal Coomassie blue.

2.4 Protein pattern and statistical analysis

Digitized images were acquired with PDQuest software (version 6.2) [15], which was used for cropping and orienting the images, detecting and identifying spots, comparing and matching spots, normalizing and analyzing the data and for preparing a report. A match set was created from the protein patterns of the two independent extracts (*C. arabica* and *C. canephora*). A standard gel was generated out of the image with the highest spot number. Spot quantities of all gels were normalized by removing nonexpression-related variations in spot intensity; for that, the raw quantity of each spot in a gel was divided by the total quantity of all the spots in that gel and included in the standard. The final synthetic image was a Gaussian scan image that contained all the Gaussian spots with a defined volume and quality. All subsequent spot matching and analysis steps in PDQuest software were performed on Gaussian spots. The results were evaluated in terms of spot OD. Statistical analysis (Student's *t*-test) via PDQuest allowed the study of proteins that were significantly increased or decreased in the two sets of samples.

2.5 *In situ* digestion and extraction of peptides

Spots were carefully cut out from 2-D Coomassie stained gels using a razor blade and subjected to in-gel trypsin digestion according to the method of Shevchenko *et al.* [16] with minor modifications. The gel pieces were swollen in a digestion buffer containing 50 mM NH_4HCO_3 and 12.5 ng/ μL trypsin (modified porcine trypsin, sequencing grade; Promega, Madison, WI, USA) in an ice bath. After 30 min the supernatant was removed and discarded and 20 μL of 50 mM NH_4HCO_3 was added to the gel pieces. Digestion was allowed to proceed at 37°C overnight. The supernatant containing tryptic peptides was dried by vacuum centrifugation. Prior to mass spectrometric analysis, the peptide mixtures were redissolved in 10 μL of 5% formic acid (FA).

2.6 MALDI-TOF MS

Peptides were desalted and concentrated according to the method of Gobom *et al.* [17]. Home-made 5 mm nanocolumns were packed with POROS R2 resin (PerSeptive Biosystems, Framingham, MA, USA) in a constricted GELoader tip (Eppendorf, Hamburg, Germany). A syringe

was used to force liquid through the columns by applying gentle air pressure. The columns were equilibrated with 20 μL of 5% FA and the analyte solutions were added. The columns were washed with 20 μL of 5% FA and the bound peptides subsequently eluted directly onto the MALDI target with 0.5 μL CHCA solution (20 $\mu\text{g}/\text{mL}$ in ACN/0.1% TFA, 70:30 vol/vol). MALDI-MS was performed using a Biflex III instrument (Bruker-Daltonics, Bremen, Germany), equipped with delayed extraction. Positively charged ions were analyzed in reflectron mode, using delayed extraction. Typically, 200 shots were averaged to improve the S/N ratio. Spectra were calibrated using trypsin autolysis products (m/z 842.51 and 2211.10) as internal standards. In a few cases this was not possible and external calibration was carried out by using peaks from a peptide calibration standard II (Bruker-Daltonics). Protein identification was performed by searching in a nonredundant protein sequence database (NCBIInr) via the MASCOT program (<http://www.matrixscience.com>). The following parameters were used for database searches: Monoisotopic mass accuracy 100 ppm, missed cleavages 2, complete carbamidomethylation of cysteines and partial oxidation of methionines. For positive identification, the score of the result of $(-10 \times \text{Log}(P))$ had to be over the significance threshold level ($p < 0.05$).

2.7 RP-HPLC-ESI-MS/MS

Peptide mixtures were separated using an LC packings system autosampler/nanoHPLC (Amsterdam, The Netherlands). A sample volume of 8 μL was loaded by the autosampler and concentrated on a commercial trap column (Zorbax 300 SB-C18, 5 μm , 300 μm I.D. \times 5 mm; Agilent Technologies, Palo Alto, CA, USA) at a flow rate of 10 $\mu\text{L}/\text{min}$. Separation was performed using a homemade capillary column (Zorbax 300 SB, C18, 3.5 μm , 75 μm id \times 10 cm) and a flow rate of 200 nL/min. HPLC solvents contained 0.1% FA and either 2% ACN (solvent A) or 98% ACN (solvent B). The column was pre-equilibrated with 100% solvent A. Elution was performed by a multisegment gradient with a first step from 0 to 5% solvent B in 5 min, a second one from 5 to 12% solvent B in 2 min, a third one from 12% to 50% solvent B in 30 min and a last segment from 50 to 100% solvent B in 8 min. Peptides were eluted directly into a Bruker ion trap Esquire 3000+ (Bruker-Daltonics). The capillary voltage was 1.5–2 kV and a dry gas flow rate of 3 L/min was used with a temperature of 230°C. The scan range used was from 300 to 1800 m/z . Typically, data were collected using an automatic data-dependent scanning mode (isolation width 4 m/z ; fragmentation amplitude 1.15 V). The spectrometer sequentially conducts MS/MS on the two most abundant precursor ions detected in the full scan. For the MS/MS spectra eight averages were allowed. For protein identification the following parameters were adopted: Complete carbamidomethylation of cysteines and partial oxidation of methionines, peptide mass tolerance \pm 1.2 Da, fragment mass tolerance \pm 0.9 Da, missed cleavages 2.

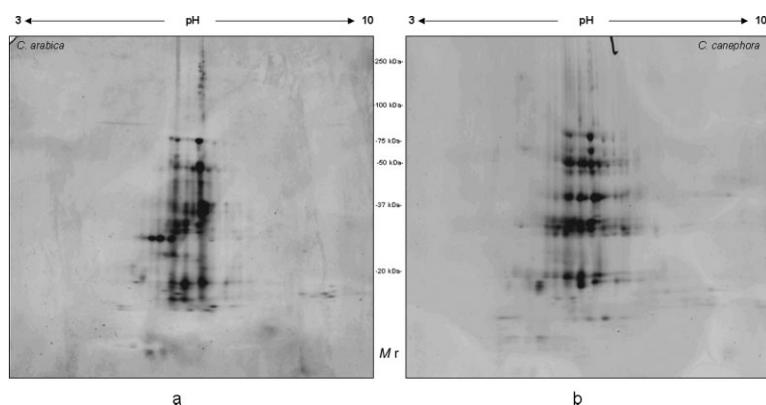


Figure 1. Representative maps of (a) *C. arabica* (var. Colombia) and (b) *C. canephora* (var. Indiano Robusta). First dimension: IPG pH 3–10 linear gradient. Second dimension: SDS-PAGE in a 7%T to 20%T porosity gradient.

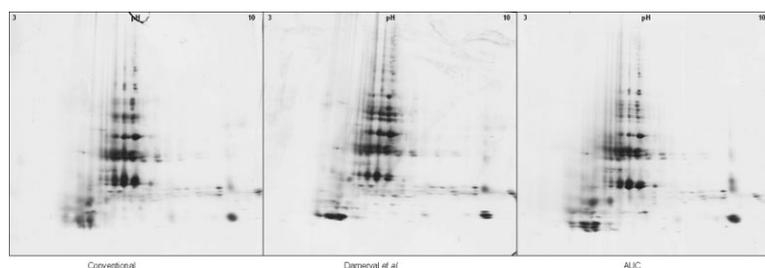


Figure 2. Examples of three 2-D maps of *C. canephora* (var. Indiano Robusta) obtained by extracting coffee grains with three different methods (from left to right: conventional method, method from Damerval *et al.* [11] and the AUC variant). The gels were stained with Sypro Ruby.

3 Results

Figure 1 gives representative maps of *C. arabica* (var. Colombia) (Fig. 1a) and *C. canephora* (var. Indiano Robusta) (Fig. 1b). The strings of spots characterizing the α and β arms, at 33 kDa and 22 kDa, respectively, are clearly visible, as well as their precursor at 55 kDa. The string of spots around 80 kDa could be oligomers (tetramers) of the β arm, as suggested by Rogers *et al.* [6]. Figure 2 shows three 2-D maps obtained by extracting the grains with the three different methods reported in section 2 (conventional method, method from Damerval *et al.* [11] and the AUC variant). In our hands, there does not seem to be any difference among the three protocols, therefore we adopted the well-engrained Damerval extraction procedure for all subsequent work. Figure 3 gives the results of the differential analysis obtained with PQQuest software. The 16 polypeptide chains present only in *C. canephora* (var. Indiano Robusta) are shown in A; panel B displays the five spots present only in *C. arabica* (var. Colombia); panel C shows eight polypeptide chains up-regulated (by factors of two to five) in *C. canephora* (var. Indiano Robusta) vs. *C. arabica* (var. Colombia) and panel D shows

the only protein we could detect up-regulated in *C. arabica* (var. Colombia). The pie chart in panel E summarizes the data on the differentially expressed proteins in the two green coffee species. A number of these proteins were identified by MALDI-TOF (Fig. 4) and ESI-MS when required. Their names and functions are given in Table 1.

4 Discussion

To our knowledge, this is the first attempt at a phylogenetic study of different coffee species *via* a differential proteomic approach. Given the number of proteins spots which were found to characterize, *via* their variation, the two species (a total of 30 proteins differentially expressed), it is hoped that, as more species are analyzed (even of closely-related lineages) it will be possible to differentiate them *via* this type of analysis. It should be remembered that this is the minimum number of spots that we found to be varied in the two species, since we adopted a medium-sensitivity stain (colloidal Coomassie Blue), which has about one order of magnitude less sensitivity than classical Silver-staining protocols [18]. We did not use Silver-staining since our aim was to detect spots which would be amenable to subsequent analysis by MS. It is well known that Silver-stains, due to the presence of formaldehyde or glutaraldehyde, are incompatible with MS analysis. Although recently aldehyde-free Silver-stains have been described [19–21], their sensitivity is much reduced, so that they do not seem to perform any better than the micellar Coomassie adopted here. Nevertheless, for identification purposes, when closely-related species are analyzed in the future, one could adopt well-controlled Silver-staining protocols or fluorescent stains to reveal many more polypeptide chains than presently reported and thus enhance the chances of detecting only subtle differences.

Our approach was also used by Rogers *et al.* [6]; however, their 2-D maps were only shown to highlight some differences between *C. arabica* and *C. canephora*, but no differential analysis with dedicated software was attempted. Interestingly, a completely different approach was attempted by Procida *et al.* [22] for obtaining fingerprints of native proteins from *C. arabica* and *C. canephora*, for differentiation of these two species as well as for characterizing differences among samples of the same variety from various plantations. They extensively dialyzed the green beans, ground them to a fine powder, mixed the powder with the MS matrix (sinapinic acid at saturation in ACN/water (50:50) in the presence of 0.1% TFA) and analyzed the entire extract by MALDI-TOF MS. In the analysis of 22 plantations of *C. arabica*, they constructed a matrix of 35 signals (peaks ranging in m/z from

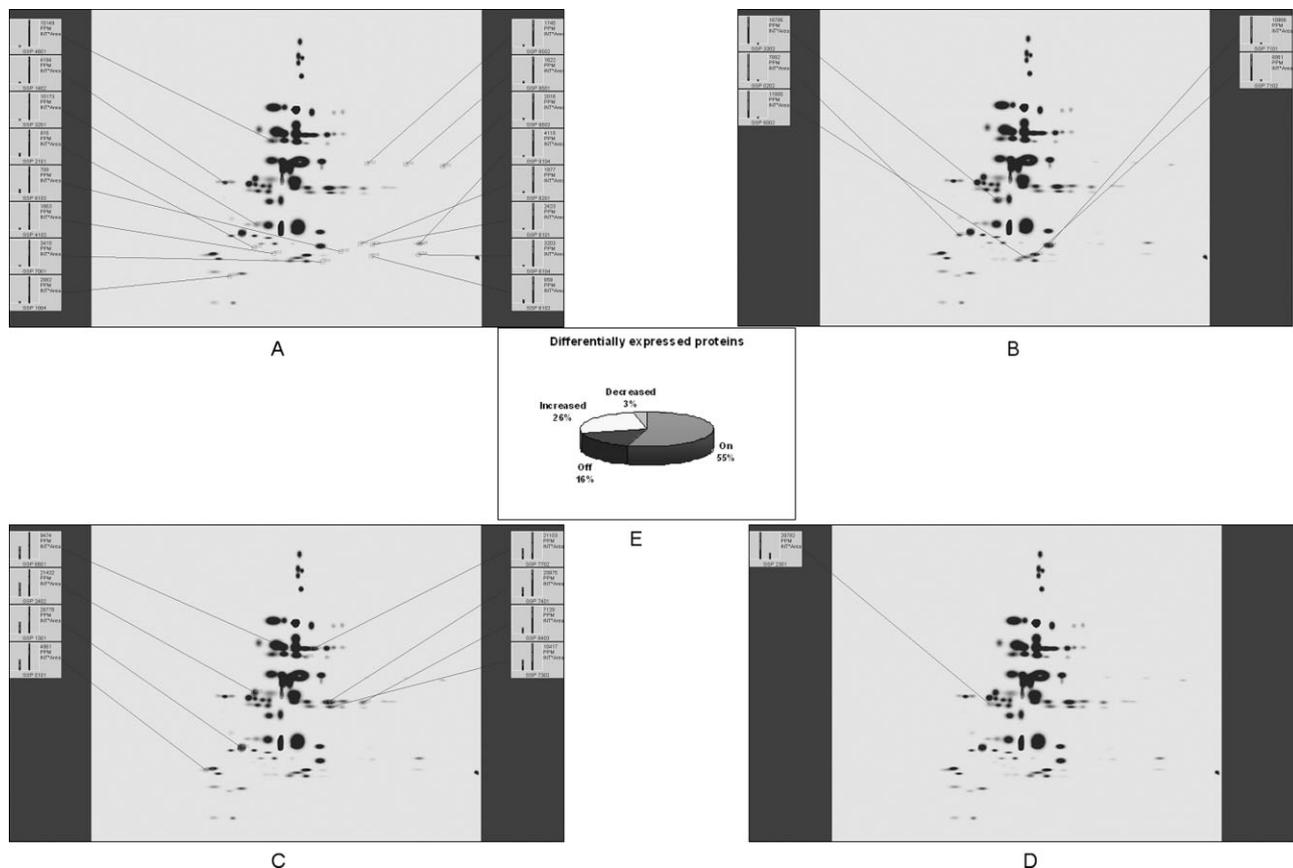


Figure 3. Variation in protein profiles between *C. arabica* (var. Colombia) and *C. canephora* (var. Indiano Robusta). (A) Visualization of the 16 polypeptide chains present only in *C. canephora* (var. Indiano Robusta). (B) Display of the five spots present only in *C. arabica* (var. Colombia). (C) Visualization of the eight polypeptide chains up-regulated (by factors of two to five) in *C. canephora* (var. Indiano Robusta) vs. *C. arabica* (var. Colombia) (increased). (D) Display of the only detectable protein up-regulated in *C. arabica* (var. Colombia) (decreased). (E) Summary of the data on the differentially expressed proteins in the green coffee species of *C. arabica* (var. Colombia) and *C. canephora* (var. Indiano Robusta).

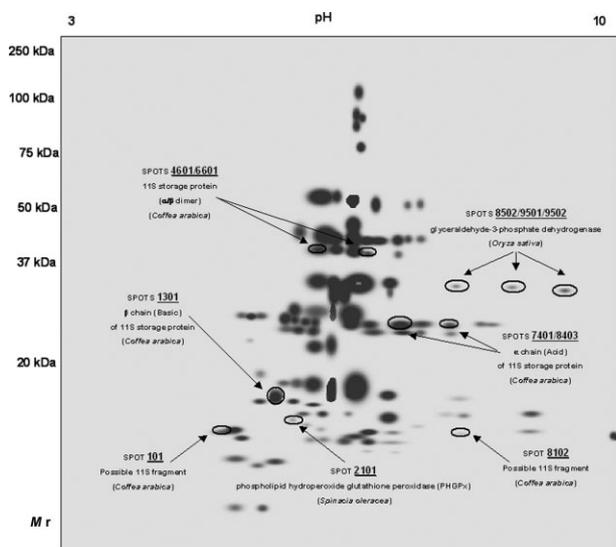


Figure 4. Master map of green coffee beans. The circled spots refer to 11 spots which could be identified by MS upon excision.

2438 up to 13 162) of which about a dozen appear to be present (with different intensities) in almost all of the 22 specimens. The other signals were characteristic of only a few cultivars. In the analysis of 17 plantations of *C. canephora*, they built a matrix of 30 signals (peaks ranging in *m/z* from 2515 up to 13 139), of which only five appear to be present (here too with different intensities) in almost all of the 17 specimens, the others being characteristic of just one or a few cultivars. Hierarchical classifications of *C. arabica* and *C. canephora* were obtained from these matrices by using the Ward's clustering method and presented in the forms of dendrograms. Even closely related plantations could be identified. This approach is certainly interesting and has the advantage of requiring a minimum of labor, since it exploits direct analysis of the entire powdered grains. However, it is impossible for us to compare the two sets of data for several reasons. First of all, Procida *et al.* [22] might also be looking at rather small peptides, which are in general not visible in 2-D maps, unless one uses special buffers in the SDS-PAGE dimension, as performed by Ludwig *et al.* [7]. Secondly, we

Table 1. Identified proteins from green coffee beans.

SSP number	Identified peptides		Theoretical M_r Calculated pI	NCBI nr data- base acces- sion number % Sequence coverage	Protein name and function	Variation (Indiano Robusta vs. Colombia)
	ESI-MS/MS ions	MALDI ions [M+H ⁺]				
8502	400.75 (+2) VVISAPSK	Not Analyzed	42 052.2 Da	gi 29150193	Putative glyceraldehyde- 3-phosphate dehydroge- nase (GAPD is a key enzyme in intermediary metabolism catalyzing the oxidation and sub- sequent phosphorylation of substrate aldehydes to acyl phosphate)	ON
9501	404.25 (+2) TLLFGEK		6.41	(<i>Oryza sativa</i>)		
9502	406.27 (+2) VLPALNGK			23%		
	464.8 (+2) KVISAPSK					
	653.25 (+2) DAPMFVVGVNEK					
	661.31 (+3) DAPMFVVGVNEK + Oxidation (M)					
	716.47 (+3) PDIDIVSNASCTTN					
	717.89 (+2) AASFNIIPSSTGAAK					
	749.95 (+2) VPTVDVSVVDLTVR					
2101	487.21 (+2) SVHEFVVR	Not Analyzed	19 047.6 Da	gi 20138151	Probable phospholipid hy- droperoxide glutathione peroxidase (Protects cells and enzymes from oxidative damage, by catalyzing the reduction of hydrogen peroxide, lipid peroxides and or- ganic hydroperoxide, by glutathione)	ON
	491.74 (+2) AEYPIFDK		5.92	(<i>Spinacia oleracea</i>)		
	562.48 (+2) GNDVDLSIYK			22%		
	674.34 (+2) VDVNGSNAAPIYK					
7401	Not Analyzed	1054.53 LPHYSNVPK 1061.55 LNAQEPSFR 1320.71 FFLAGNPQQGGGK 1876.94 TQCNIQKLNAQEPSFR 3299.64 LQLFLPEYSE QEQQPQQQQGQQQQGVGR	30 428.7 Da 6.93 ^{a)}	gi 2979526 (<i>Coffea arabica</i>) 24% ^{c)}	11S storage protein Acid chain (α) (This is a seed storage protein. He- xamer; each subunit is composed of an acidic and a basic chain derived from a single precursor and linked by a disulfide bond. Belongs to the 11S seed storage protein (globulins) family)	Up-regulated 3.7
8403	527.77 (+2) LPHYSNVPK 531.11 (+2) LNAQEPSFR 660.82 (+2) FFLAGNPQQGGGK 1121.53 (+2) NIFSGFDDQLLA DAFNVDLK 1250.74 (+3) LQLFLPEY- SEQVQQPQQQQE QQQHGVGRGWR	927.37 FPDRHQQ 1054.48 LPHYSNVPK 1061.75 LNAQEPSFR 1320.68 FFLAGNPQQGGGK 2242.20 NIFSGFDDQLLA DAFNVDLK	31 418.8 Da 7.03 ^{a)}	gi 4127631 (<i>Coffea arabica</i>) 31% ^{d)}	11S storage protein Acid chain (α)	Up-regulated 6.2
8102	483.60 (+3) KFLAGNPQQGGGK 660.84 (+2) FFLAGNPQQGGGK 690.92 (+2) GLRPLPHYSNVPK 1528.40.60 (+3) GDVLILLPGFTQW TYNDGDV PLVTVALLDVA NEANQLDLQSR	Not Analyzed	51 195.3 Da 6.65 ^{a)}	gi 2979526 (<i>Coffea arabica</i>) 15%	Possibile 11S fragment	ON
101	618.88 (+2) GNARIQVVDHK 666.32 (+2) AGNEGFYVAFK 707.63 (+2) TNDNAMINPLVGR 715.85 (+2) TNDNAMINPLVGR + Oxidation (M)	Not Analyzed	51 195.3 Da 6.65 ^{a, b)}	gi 2979526 (<i>Coffea arabica</i>) 8%	Possibile 11S fragment	Up-regulated 3.0

Table 1. Continued

SSP number	Identified peptides		Theoretical M_r Calculated pI	NCBI nr data- base acces- sion number % Sequence coverage	Protein name and function	Variation (Indiano Robusta vs. Colombia)
	ESI-MS/MS ions	MALDI ions [M+H ⁺]				
1301	Not Analyzed	1331.84 AGNEGFYVAFK 1427.05 IPILSSLQLSAER 1459.97 KAGNEGFYVAFK 1584.04 SSFQISSEAEELK 1717.29 QGQLIIVPQYFAVIK 1899.24 LSENIGLPQEADVFNPR 1960.21 SSFQISSEAEELKYGR	20 784.5 Da 6.3 ^{a, b)}	gi 2979526 (<i>Coffea arabica</i>) 40% ^{e)}	11S storage protein chain (β)	Basic Up-regulated 2.3
4601	531.12 (+2)LNAQEPSFR 660.59 (+2) FFLAGNPQQGGGK 666.37 (+2) AGNEGFYVAFK 715.85 (+2) TNDNAMINPLVGR 792.72 (+2)SSFQISSEAEELK	Not Analyzed	51 195.3 Da 6.65 ^{a)}	gi 2979526 (<i>Coffea arabica</i>) 13%	11S storage protein (α/β dimer)	ON
6601	Not Analyzed	1899.33 LSENIGLPQEADVFNPR 1427.08 IPILSSLQLSAER 717.32 QGQLIIVPQYFAVIK 584.03 SSFQISSEAEELK 061.47 LNAQEPSFR 380.56 LQLFLPEYS EQEQPQQQQEQQQHGVGR	964.1 Da 6.73 ^{a)}	gi 3641256 (<i>Coffea arabica</i>) 23%	11S storage protein (α/β dimer)	Up-regulated 2.4

a) Calculated pI value according to [6]

b) For possible 11S fragments we reported the calculated M_r and pI of the entire legumin

c) Calculated considering the amino acid sequence corresponding to the acid chain 28–300 aa of the entire precursor (gi|2979526)

d) Calculated considering the amino acid sequence corresponding to the acid chain 27–306 aa of the entire precursor (gi|4127631)

e) Calculated considering the amino acid sequence corresponding to the basic chain 301–487 aa of the entire precursor (gi|2979526)

The comparison is performed between the matched spots in *C. canephora* (var. Indiano Robusta) and *C. arabica* (var. Colombia). A threshold value of 2 (which corresponds to a variation of 100%) was chosen as a meaningful variation in the comparison of Indiano proteins vs. Colombia proteins. The numbers in the last column refer to folds of change.

might be seeing a set of quite different polypeptide chains, due to the fact that our samples are solubilized in strong cocktails containing the classical mixture of urea/thiourea and surfactants, whereas their solubilization mixture is just the solvent for sinapinic acid, which may not bring into solution whole classes of proteins. Thirdly, whereas our proteins, as extracted from a 2-D map, are characterized by a pI, an intact mass in MALDI-TOF (as well as peptide masses for identification purposes) and, when required, a lead sequence via ESI-MS, their peaks are only characterized by an m/z signal, which unfortunately cannot tell much about the properties of the polypeptide chain accompanying the signal.

5 Concluding remarks

Although preliminary in nature, these studies will be quite valuable in view of extending them to quality control: e.g. fraud detection (such as mixing of beans of different origins and of different commercial value), food adulteration (for example, if coffee beans are mixed with husks, hulls, or parchments (non-bean parts of the coffee cherry) the pro-

cessed soluble coffee will contain high levels of the sugar xylose), and food products containing material derived from a genetically modified organism (GMO).

Scientists in the food and beverage industry are faced with many different quality control tasks, such as making sure that flavors meet certain standards, identifying changes in process parameters that may lead to a change in quality, detecting adulterations in ingredients and identifying the geographical origin of raw materials. Food scientists who work for regulatory agencies, such as the Food and Drug Administration, are interested in detecting economic fraud due to product substitution and adulteration, as well as health risks from possible contamination. In this scenario, the proteomic approach could be a valuable method to assist in quality control.

P. G. R. is supported by grants from MIUR (COFIN 2003 and FIRB2001, no. RBNE01KJHT). M. T. G. A. would like to thank the Secretaría de Estado de Educación y Universidades and the European Social Fund for her post-doctoral fellowship. L. Z. is supported by Progetto Strategico 1999 (Area 2.2).

6 References

- [1] Capodici, S., Invernizzi, C., *Conoscere il Caffè*, Eusebianum, Milano, Italy 1983, pp. 27–30.
- [2] Charrier, A., Eskes, A. B., in: Charrier, A., Jacquot, M., Hamon, S., Nicolas, D. (Eds.), *L'Amélioration des plantes Tropicales*, Cirad Orstrom, Toulouse, France 1997, pp. 171–196.
- [3] Flament, I., *Coffee Flavour Chemistry*, Wiley, New York, USA 2002, pp. 79–99.
- [4] Brinkman, U. A., Vreuls, R. J. J., *J. Chromatogr. A* 2003, **1019**, 1–285.
- [5] Holscher, W., Steinhart, H., in: Parliament, T., Morello, M., McGorin, R. (Eds.), *Thermally Generated Flavours*, ACS Symposium Series 543, American Chemical Society, Washington, USA 1992, pp. 207–217.
- [6] Rogers, W. J., Bezdard, G., Deshayes, A., Meyer, I. *et al.*, *Plant. Physiol. Biochem.* 1999, **37**, 261–272.
- [7] Ludwig, E., Lipke, U., Raczek, U., Jäger, A., *Eur. Food Res. Technol.* 2000, **211**, 111–116.
- [8] Schägger, H., von Jagow, G., *Anal. Biochem.* 1987, **166**, 368–379.
- [9] Montavon, P., Duruz, E., Rumo, G., Pratz, G., *J. Agric. Food Chem.* 2003, **51**, 2328–2334.
- [10] Montavon, P., Mauron, A. F., Duruz, E., *J. Agric. Food Chem.* 2003, **51**, 2335–2343.
- [11] Damerval, C., De Vienne, D., Zivy, M., Thiellement, H., *Electrophoresis* 1986, **7**, 52–54.
- [12] Herbert, B., Galvani, M., Hamdan, M., Olivieri, E. *et al.*, *Electrophoresis* 2001, **22**, 2046–2057.
- [13] Righetti, P. G., *Immobilized pH Gradients: Theory and Methodology*, Elsevier, Amsterdam, The Netherlands 1990, pp. 206–220.
- [14] Steinberg, T. H., Chernokalskaya, E., Berggren, K., Lopez, M. F. *et al.*, *Electrophoresis* 2000, **21**, 486–496.
- [15] Garrells, J. I., Farrar, J. T., Burwell, C. B., in: Cedlis, J. E., Bravo, R. (Eds.), *Two-Dimensional Gel Electrophoresis of Proteins*, Academic Press, Orlando, USA 1984, pp. 37–91.
- [16] Shevchenko, A., Wilm, M., Vorm, O., Mann, M., *Anal. Chem.* 1996, **68**, 850–858.
- [17] Gobom, J., Nordhoff, E., Mirgorodskaya, E., Ekman, R., Roepstorff, P., *J. Mass Spectrom.* 1999, **34**, 105–116.
- [18] Merrill, C. R., Washart, K. M., in: Hames, B. D., (Ed.), *Gel Electrophoresis of Proteins: a Practical Approach*, Oxford University Press, Oxford, England 1998, pp. 53–91.
- [19] Mortz, E., Krogh, T. N., Vorum, H., Goerg, A., *Proteomics* 2001, **1**, 1359–1363.
- [20] Sinha, P., Poland, J., Schnoelzer, M., Rabilloud, T., *Proteomics* 2001, **1**, 835–840.
- [21] Yan, J. X., Harry, R. A., Spibey, C., Dunn, M. J., *Electrophoresis* 2000, **21**, 3657–3665.
- [22] Procida, G., Campisi, B., Seraglia, R., Traldi, P., *Rapid Commun. Mass Spectrom.* 2003, **17**, 140–148.

REGULAR ARTICLE

Spot overlapping in two-dimensional maps: A serious problem ignored for much too long

Natascia Campostrini¹, Liliana B. Areces², Juri Rappsilber²,
Maria Chiara Pietrogrande³, Francesco Dondi³, Fabio Pastorino⁴,
Mirco Ponzoni⁴ and Pier Giorgio Righetti¹

¹ Department of Agricultural and Industrial Biotechnologies, University of Verona, Verona, Italy

² FIRC Institute for Molecular Oncology Foundation (IFOM) and European Institute of Oncology (IEO), Milan, Italy

³ Department of Chemistry, University of Ferrara, Ferrara, Italy

⁴ Laboratory of Oncology, G. Gaslini Children's Hospital, Genoa, Italy

In the analysis of a neuroblastoma xenograft implanted in mice using two-dimensional maps, some 85 proteins were found to be up- or down-regulated (out of a total of 264 detected by a medium-sensitivity colloidal Coomassie stain). When these spots were eluted and analysed by mass spectrometry in a quadrupole time of flight mass spectrometer, a number of spots were found to be envelopes of different polypeptide chains. Out of a total of 74 proteins identified, 52 (71%) were found to be singlets, 14 (19%) were doublets, 6 (8%) were triplets, 1 was a quadruplet and 1 a quintuplet. Analysis of the ΔpI and ΔM_r of all species contained in a single gel segment eluted helped point out potential errors in protein identification. This was a unique case, in that very minute bioptic sample loads were applied to the gel. In normal cases, where sample loads of ca. 1 mg of total protein are applied and typically at least 1000 spots are visualised, the singlets will be the minority, rarely exceeding 30% of all spots analysed. The experimental data on the abundance of overlapping spots were in excellent agreement with theoretical data calculated on the basis of the statistical theory of spot overlapping, originally proposed by Davis and further developed by some of the authors. Ways and means for minimizing spot overlap and visualising a greater number of spots in a two-dimensional map are discussed.

Received: October 8, 2004

Revised: November 25, 2004

Accepted: January 16, 2005

Keywords:

Neuroblastoma / Spot overlap / Two-dimensional maps

1 Introduction

Until the early 1970s, the life of biochemists was perhaps not as complex as today. Their main concern was to purify to homogeneity their "pet" proteins and to spend the remain-

ing time studying their physico-chemical properties [1]. Thus, they had a simple way of harnessing the enormous complexity of a given biological system: to get rid of most of the material. Things changed abruptly in the mid-1970s, when someone made the bold move of trying to take a snapshot of all the components of the biological system under investigation [2–4]. The system that emerged and became popular was that of O'Farrell [2], perhaps because he devised the most elaborate "photographic camera". He was able to resolve and detect about 1100 different proteins from lysed *Escherichia coli* cells on a single 2-D map and suggested that the maximum resolution capability could have been as high as 5000 different proteins. Major reasons for the advance in

Correspondence: Professor Pier Giorgio Righetti, Department of Agricultural and Industrial Biotechnologies, Strada le Grazie 15, I-37134 Verona, Italy

E-mail: righetti@sci.univr.it

Fax: +39-045-8027929

Abbreviations: **IM**, interdistance model; **NB**, neuroblastoma; **SC**, single components; **SDO**, statistical degree of overlapping

resolution obtained by O'Farrell, compared to earlier workers, included the use of samples labelled with ^{14}C or ^{35}S to high specific activity and the use of thin (0.8 mm) gel slabs for the second dimension, which could then be easily dried for autoradiography. This detection method was able to reveal protein zones corresponding to one part in 10^7 of the sample (usually 1–20 μg was applied initially, since higher loads caused zone spreading, although up to 100 μg could be loaded). For the first IEF dimension, O'Farrell adopted gel rods of 13 cm in length and 2.5 mm in diameter. For the second SDS-PAGE dimension, he used the discontinuous buffer system of Laemmli [5], and for improved resolution, a concave exponential gradient of a polyacrylamide gel [6]. Since its very inception, O'Farrell carefully selected the best conditions available at the time; his system thus [7] survived almost unscathed for 30 years, with the exception of the introduction of IPGs [8] and of much improved solubilising cocktails [9].

Considering that the *E. coli* proteome could hardly comprise more than 4289 polypeptide chains [10] and that O'Farrell resolved some 1100 spots, while using the highest sensitivity detection method available, much better than silver staining [11], one wonders why he could not detect many more spots, although he claimed that his system could potentially resolve up to 5000 spots. One reason could be its use of low levels of sample loaded onto the gel, hindering the detection of low-abundance proteins (it is the current refrain of proteome analysis that we only see the tip of the iceberg!) [12]. The other reason, however, could be that the space allotted in the 2-D plane for accommodating the 4289 polypeptide chains was simply not sufficient (considering that the first IEF dimension was barely 13 cm and the second, M_r axis, was only 15 cm long).

With this problem in mind, in 2002 we began a long-term study on the possibility of spot overlapping in 2-D gels and its relevance to proteomics [13–19]. The results were not encouraging. If one explores the space allotted to most 2-D maps generated today (18 \times 20 cm, IEF \times SDS-PAGE) using a pH 3–10 gradient, one finds that the situation is far from perfect: for a sample containing barely 639 spots, only 32% are singlets (*i.e.*, homogenous spots comprising a single polypeptide chain), the remaining being doublets, triplets and even higher-order envelopes. In the case of 1500 spots loaded, the singlets would be only 27%; for a total of 3000 polypeptide chains, the singlets would amount to a mere 14%. This could very well have been the case of O'Farrell: given the meagre size of his 2-D gels, even if he had loaded much higher amounts, the singlets would have been less than 10%, and his chances of finding many more spots would have been slim. Thus, it would appear that in most conditions adopted today in 2-D map analysis, singlets are by far the rarest species, most other spots being envelopes of two or more chains.

Recently, when working on our project on neuroblastoma (NB) analysis [20, 21] and eluting spots for MS recognition, by MS techniques far better than the standard

MALDI-TOF we had been using, we were quite astonished to discover how substantial the number of overlapping spots was, even at relatively low sample loads.

We report here on what we believe is the first true comparison between the theoretical and experimental spot overlapping in present-day 2-D mapping.

2 Materials and methods

2.1 Chemicals and materials

Tris, mineral oil, DL-DTT, Tween 20 and EDTA were purchased from Sigma-Aldrich (Steinheim, Germany). Glycerol, methanol, ethanol, acetic acid and acetone were obtained from Merck (Darmstadt, Germany). Acrylamide (AA), 40% AA/Bis solution, TEMED, ammonium persulphate (APS), the Protean IEF Cell, a GS-710 densitometer, a Versa Doc Scanner, the software PDQuest version 6.2 as well as linear Immobiline dry strips pH gradient 3–10 (17 cm) were obtained from Bio-Rad (Hercules, CA, USA); glycine, SDS, iodoacetamide, urea, thiourea, tributylphosphine (TBP) and CHAPS were purchased from Fluka (Buchs, Switzerland). Bromophenol blue, carrier ampholytes and agarose were purchased from Amersham Biosciences (Uppsala, Sweden).

2.2 Xenograft neuroblastoma animal model

NB cells were injected in the adrenal glands of mice, as previously reported [22]. Briefly, 5-week-old female SCID mice purchased from Harlan Laboratories (Udine, Italy) were anaesthetised and injected with 2×10^6 SH-SY5Y NB cells in 20 μL of HEPES buffer in the capsule of the left adrenal gland after laparotomy. Mice were monitored at least two times weekly for evidence of tumour development, quantification of tumour size and evidence of tumour-associated morbidity. Mice were sacrificed and organs were frozen and stored in liquid nitrogen after being washed in PBS. All experiments involving animals were reviewed and approved by the licensing and ethical committee of the National Cancer Research Institute and by the Italian Ministry of Health.

2.3 Two-dimensional gel electrophoresis

2.3.1 Sample treatment

Adrenal glands from SCID mice carrying primary NB tumour and NB liver metastases were stored at -80°C . All samples were homogenized (5% homogenate) with a lysis solution containing 7 M urea, 2 M thiourea, 3% CHAPS, 40 mM Tris, 5 mM TBP, 0.5% carrier ampholytes pH 3–10, 5 mM EDTA and 1 mM PMSF. After 1 h of lysis, samples were centrifuged at 4°C at 6000 rpm to eliminate all residual particles. The collected supernatant was reduced and alkylated as previously described [23]. Interfering substances

(lipids, salts) were removed by a precipitation step with a cold mix of acetone and methanol (8:1). Pellets were finally resuspended in sample solution containing urea, thiourea, CHAPS and Tris. The first-dimension run was performed on strips (17 cm length, 0.5 mm thickness) with a linear pH gradient from pH 3 to 10 [8]. For analytical runs (used to find differentially expressed proteins, gels stained with SYPRO Ruby), the IPG strips were rehydrated with 500 μg of pooled samples (four strips with pooled adrenal glands carrying primary NB tumour and four strips with pooled liver metastasis samples) containing traces of bromophenol blue for monitoring the electrophoretic run. For preparative runs (used to isolate proteins from the gel for identification with MS, gels stained with colloidal Coomassie blue), the IPG strips were rehydrated with 600 μg of pooled samples. For all subsequent steps (SDS-PAGE, staining and differential analysis with PDQuest) see previous reports [20, 21], where all these steps are described in detail. The images of gels stained with SYPRO Ruby and colloidal Coomassie blue were superimposed; the spot patterns obtained with the two stain methods, regarding the spots to be cut, were identical.

2.4 Protein identification by mass spectrometry

2.4.1 Digestion of proteins and sample preparation

Classic “in-gel digestion” was performed according to the standard protocol [24], but in a 96-well plate for parallel processing (Arecas and Rappsilber, manuscript in preparation). In brief, the excised gel pieces were transferred manually into an individual well of a V-bottom polypropylene 96-well plate (Pink Microtiter plates-pierced; Genomic Solutions, Huntingdon, Cambs, UK) and washed sequentially once with 100 μL of water and 100 μL of 25 mM NH_4HCO_3 ; they were dehydrated by adding 100 μL of ACN. The gel pieces were then swelled in 100 μL of 10 mM DTT for 30 min at room temperature and subsequently shrunk with ACN. The proteins were alkylated with 100 μL of 55 mM iodoacetamide for 20 min at room temperature in the dark. All the liquid material was transferred by centrifugation at 1000 rpm for 2 min. After drying completely, the gel pieces were reswollen with 20–25 μL of digestion buffer (12.5 ng/ μL trypsin in 25 mM NH_4HCO_3) for 30 min at 4°C. Afterwards, the digestion buffer was replaced by the same buffer without trypsin to ensure that the gel plugs were kept wet during the digestion process. The plate was covered and incubated at 37°C for 12–18 h.

Following digestion, the peptide mixture was desalted and concentrated using home-made C_{18} -Stage tips [25]. The peptides were eluted directly onto a collection plate using 10 μL of 80% ACN/1% TFA by vacuum (4–6 Hg). Following elution, the volume of the extracted peptides was reduced to 2 μL using a centrifugal vacuum concentrator (Concentrator 5301, Eppendorf, Albany, NY, USA).

2.4.2 MS and data analysis

The total tryptic digest samples were analysed by RP nano-LC-MS/MS using an Agilent 1100 nanoflow LC system (Agilent Technologies, Palo Alto, CA, USA). The LC system was coupled to a QSTAR XL hybrid QqTOF mass spectrometer (Applied Biosystems/MDS Sciex, Concord, Ontario, Canada). Binding and chromatographic separation of peptides were achieved in a 15-cm-long fused silica emitter (100 μm id; Picotip New Objective, Woburn, MA, USA) packed in-house with reverse-phase C18 resin (Maisch, Germany).

The tryptic peptide mixtures were loaded onto the packed column using an auto-sampler at a flow rate of 700 nL/min. The peptides were separated in a linear gradient of 5 to 24% ACN in 0.5% acetic acid over 15 min and then over 5 min in 81% ACN at 300 nL/min. The QSTAR was operated in data-dependent acquisition mode to automatically switch between MS and MS/MS.

Protein identification was made using MASCOT (Matrix Science, London, UK) to search the mammalian NCBI non-redundant protein database. Search parameters were: MS and MS/MS tolerance 0.2 Da; tryptic specificity allowing for up to one missed cleavage; fixed modification, carbamidomethylation of cysteine; variable modification, oxidation of methionine.

2.5 Application of the spot overlapping theory

The statistical procedure “quantitative theory of peak overlapping” [15, 16] was applied to experimental data in order to statistically estimate the overlapping degree present in the 2-D PAGE separation. A map was constructed by using the pI and M_r coordinates (pI and M_r of the barycentre of each of the 74 spots detected). The third intensity coordinate is described by the abundance of each spot as retrieved from the PDQuest software. The 2-D PAGE map was divided into a proper number of strips (4) so that each strip contained at least 20 single components (SCs). On each 1-D separation, the statistical procedure was applied to estimate the number of proteins present in each strip and their sum yielded the total number in the map.

Data concerning each strip are as followings:

p	m	α
21	28	0.26
18	21	0.20
19	31	0.33
16	23	0.19

where p is the total number of detectable spots (comprising singlets, s , doublets, d , triplets, t , etc.) and m the total number of SCs (thus, $m = s + 2d + 3t + \dots$). Their sum yielded the total number in the map: 103 components, whereas 74 spots were detected. This means that the degree of overlapping γ , *i.e.*, p/m (in%), present in the map was 71% (note that in an ideal map where a complete separation has

SPOT OVERLAPPING

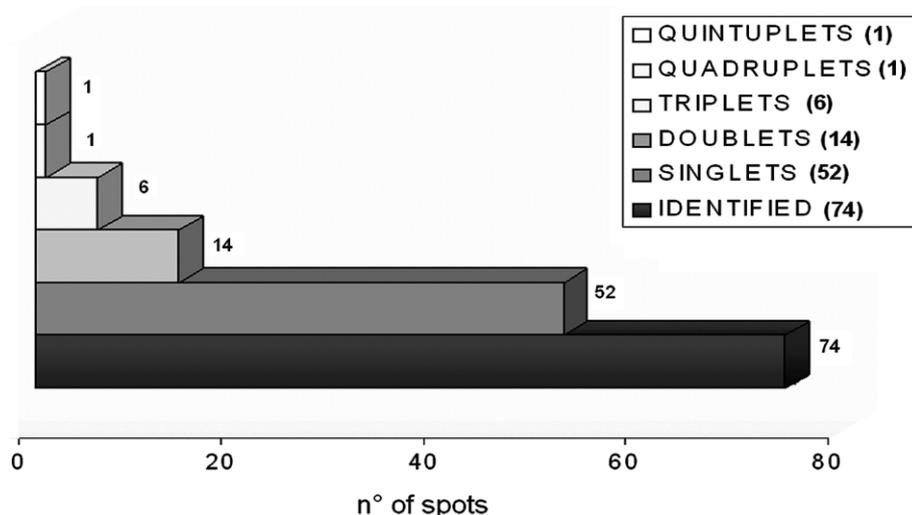


Figure 2. Graphic representation of the number of singlets, doublets, triplets, quadruplets and quintuplets found in the spots analysed by MS (the first bar indicates the total number of proteins identified).

Table 1. Total number of spots identified by MS and divided into singlets, doublets, triplets, quadruplets and quintuplets

SINGLETs (52 SPOTS)				
SSP number	2-D Gel Liver Metastasis-Primary Tumour	pI, M _r (Da) (Swiss-Prot)	Δ pI, Δ M _r	Sample
8004	P12710 FABL_MOUSE Fatty acid-binding protein, liver	8.59–14246		Liver metastasis
8007	P31786 ACBP_MOUSE Acyl-CoA-binding protein (ACBP)	8.78–9869		Liver metastasis
7003	P26772 CH10_RAT 10 kDa heat shock protein, mitochondrial (Hsp10)	8.91–10770		Liver metastasis
5009	P80254 DOPD_RAT D-dopachrome tautomerase	6.15–13002		Liver metastasis
5011	P16436 HNT1_BOVIN Histidine triad nucleotide-binding protein 1	6.31–13648		Liver metastasis
4410	Q9QXD6 F16P_MOUSE (Q9QXD6) Fructose-1,6-bisphosphatase (EC 3.1.3.11)	6.18–36781		Liver metastasis
7204	P19157 GTP2_MOUSE Glutathione S-transferase P 2	8.13–23478		Liver metastasis
7208	P35700 PDX1_MOUSE Peroxiredoxin 1 (EC 1.11.1.-)	8.26–22176		Liver metastasis
6205	P24472 GTA4_MOUSE Glutathione S-transferase 5.7 (EC 2.5.1.18)	6.77–25564		Liver metastasis
6203	P35700 PDX1_MOUSE Peroxiredoxin 1 (EC 1.11.1.-)	8.26–22176		Liver metastasis
4310	O08709 PDX6_MOUSE Peroxiredoxin 6 (EC 1.11.1.-)	5.72–24739		Liver metastasis
1208	O00264 PGC1_HUMAN Membrane associated progesterone receptor component	4.56–21540		Liver metastasis
601	P16991 YB1_HUMAN) Nuclease sensitive element binding protein 1	9.87–35924		Liver metastasis
6309	P13255 GLMT_RAT Glycine N-methyltransferase	7.21–32418		Liver metastasis
7603	P54869 HMCM_MOUSE Hydroxymethylglutaryl-CoA synthase	8.02–53786		Liver metastasis
6408	P00480 OTC_HUMAN Ornithine carbamoyltransferase, mitochondrialprecursor	8.75–39901		Liver metastasis
5613	P47738 DHAM_MOUSE Aldehyde dehydrogenase, mitochondrial precursor	7.53–56537		Liver metastasis
5006	P70349 HNT1_MOUSE Histidine triad nucleotide-binding protein 1	6.31–13648		Liver metastasis
1306	P34022 RANG_MOUSE Ran-specific GTPase-activating protein	5.15–23596		Liver metastasis
2206	Q9CPU0 LGUL_MOUSE Lactoylglutathione lyase	5.25–20678		Liver metastasis
2409	Q64374 SM30_MOUSE Senescence marker protein-30	5.16–33407		Liver metastasis
2311	Q64374 SM30_MOUSE Senescence marker protein-30	5.16–33407		Liver metastasis
7610	P04762 CATA_RAT Catalase (EC 1.11.1.6)	7.15–59626		Liver metastasis
6504	Q99J08 S142_MOUSE SEC14-like protein 2	6.68–46300		Liver metastasis
7704	P30038 PUT2_HUMAN Delta-1-pyrroline-5-carboxylate dehydrogenase	8.25–61751		Liver metastasis
6607	P10860 DHE3_RAT Glutamate dehydrogenase, mitochondrial precursor	8.05–61428		Liver metastasis

Table 1. Continued

SINGLETs (52 SPOTS)				
SSP number	2-D Gel Liver Metastasis-Primary Tumour	pI, M _r (Da) (Swiss-Prot)	Δ pI, Δ M _r	Sample
6508	P35505 FAAA_MOUSE Fumarylacetoacetase	6.92–46103		Liver metastasis
7503	P17764 THIL_RAT Acetyl-CoA acetyltransferase, mitochondrial precursor	8.92–44695		Liver metastasis
5404	P14152 MDHC_MOUSE Malate dehydrogenase, cytoplasmic (EC 1.1.1.37)	6.16–36346		Liver metastasis
5614	P47738 DHAM_MOUSE Aldehyde dehydrogenase, mitochondrial precursor	7.53–56537		Liver metastasis
6410	P13255 GLMT_RAT Glycine N-methyltransferase (EC 2.1.1.20)	7.21–32418		Liver metastasis
7312	P29410 KAD2_RAT Adenylate kinase isoenzyme 2, mitochondrial	6.32–26248		Liver metastasis
7705	P30038 PUT2_HUMAN Delta-1-pyrroline-5-carboxylate dehydrogenase	8.25–61751		Liver metastasis
7508	O09171 BHMT_RAT Betaine-homocysteine S-methyltransferase	8.02–44976		Liver metastasis
7509	P13437 THIM_RAT 3-ketoacyl-CoA thiolase, mitochondrial	8.09–41871		Liver metastasis
6705	Q63342 M2GD_RAT Dimethylglycine dehydrogenase, mitochondrial precursor	6.91–96047		Liver metastasis
7609	P24270 CATA_MOUSE Catalase (EC 1.11.1.6)	7.72–59626		Liver metastasis
4201	Q00623 Apolipoprotein A-I precursor	5.64–30587		Primary tumour
3306	P09936 UBL1_HUMAN Ubiquitin carboxyl-terminal hydrolase isozyme	5.33–24824		Primary tumour
7202	P30086 PEBP_HUMAN Phosphatidylethanolamine-binding protein (PEBP)	7.43–20925		Primary tumour
5003	P02248 UBIQ_HUMAN (P02248) Ubiquitin	6.56–8565		Primary tumour
5004	P31949 S111_HUMAN Calcizzarin (S100C protein)	6.56–11740		Primary tumour
1004	P46656 ADX_MOUSE Adrenodoxin, mitochondrial precursor	5.37–20123		Primary tumour
1302	P17080 RAN_HUMAN GTP-binding nuclear protein RAN (TC4)	7.01–23596		Primary tumour
4001	P02695 RET3_MOUSE Retinoic acid-binding protein I	5.31–15460		Primary tumour
6302	O70250 PMG2_MOUSE Phosphoglycerate mutase	8.65–28696		Primary tumour
1702	P27797 CRTG_HUMAN (P27797) Calreticulin precursor (CRP55) (Calregulin)	4.29–48111		Primary tumour
3401	Q15691 MAE1_HUMAN (Q15691) Microtubule-associated protein RP/EB family member	5.02–29999		Primary tumour
8507	P09411 PGK1_MOUSE Phosphoglycerate kinase 1	7.52–44405		Primary tumour
5505	Q15019 SEP2_HUMAN Septin 2 (NEDD5 protein homolog)	6.15–41487		Primary tumour
4101	P13668 STN1_RAT Stathmin (Phosphoprotein p19)	6–17117		Primary tumour
4108	P13668 STN1_RAT Stathmin (Phosphoprotein p19) (pp19)	6–17117		Primary tumour
DOUBLETS (14 SPOTS)				
5103	P08228 SODC_MOUSE Superoxide dismutase [Cu-Zn] (EC 1.15.1.1)	6.03–15811 10.32–13755	4.02–2056	Liver metastasis
3303	P02278 H2BA_HUMAN (P02278) Histone H2B.a/g/k Q64374 SM30_MOUSE Senescence marker protein-30 (SMP-30) Q99426 TBCB_HUMAN (Q99426) Tubulin-specific chaperone B	5.16–33407 5.06–27325	0.1–6082	Liver metastasis
6405	Q61176 ARG1_MOUSE Arginase 1 (Liver-type arginase) P00481 OTC_RAT Ornithine carbamoyltransferase, mitochondrial	6.52–34808 9.12–39886	2.6–5078	Liver metastasis
7506	Q92524 PRSX_HUMAN 26S protease regulatory subunit S10B	7.09–44173	1.27–2411	Liver metastasis
7501	P16460 ASSY_MOUSE Argininosuccinate synthase P16460 ASSY_MOUSE Argininosuccinate synthase	8.36–46584 8.36–46584	0.27–4713	Liver metastasis
7311	P13437 THIM_RAT 3-ketoacyl-CoA thiolase, mitochondrial P14604 ECHM_RAT Enoyl-CoA hydratase, mitochondrial precursor	8.09–41871 8.40–31516	0.66–1335	Liver metastasis
7611	Q8R164 BPHL_MOUSE Valacyclovir hydrolase Q03265 ATPA_MOUSE ATP synthase alpha chain, mitochondrial precursor P24549 DHA1_MOUSE Aldehyde dehydrogenase 1A1	9.06–32851 9.22–59752 7.91–54336	1.31–5416	Liver metastasis

Table 1. Continued

DOUBLETS (14 SPOTS)				
SSP number	2-D Gel Liver Metastasis-Primary Tumour	pI, M _r (Da) (Swiss-Prot)	Δ pI, Δ M _r	Sample
6510	Q07417 ACDS_MOUSE Acyl-CoA dehydrogenase, short-chain	8.96–44946	0.4–869	Liver metastasis
	O35423 SPYA_MOUSE Serine-pyruvate aminotransferase	8.57–45815		
8510	P13437 THIM_RAT 3-ketoacyl-CoA thiolase, mitochondrial	8.09–41871	1.17–6363	Liver metastasis
	Q9DB77 UCR2_MOUSE Ubiquinol-cytochrome C reductase complex core	9.26–48234		
8509	P16460 ASSY_MOUSE Argininosuccinate synthase	8.36–46584	0.06–1988	Liver metastasis
	P00558 PGK1_HUMAN Phosphoglycerate kinase	8.3–44596		
4506	P55262 ADK_CRIGR Adenosine kinase (EC 2.7.1.20)	6.12–40285	0.51–3413	Liver metastasis
	P13444 METL_RAT S-adenosylmethionine synthetase alpha and beta	5.61–43698		
6501	P42024 ACTZ_HUMAN Alpha-centractin (Centractin)	6.19–42613	0.06–7505	Primary tumour
	P26641 EF1G_HUMAN Elongation factor 1-gamma (EF-1-gamma)	6.25–50118		
2710	P08238 HS9B_HUMAN Heat shock protein HSP 90-beta	4.97–83133	0.19–33033	Primary tumour
	P05218 TBB5_HUMAN Tubulin beta-5	4.78–50100		
6001	P16436 HNT1_BOVIN Histidine triad nucleotide-binding protein 1	6.31–13648	0.95–2061	Primary tumour
	P02088 HBB1_MOUSE Hemoglobin beta-1 chain (B1)	7.26–15709		
TRIPLETS (6 SPOTS)				
7507	P16460 ASSY_MOUSE Argininosuccinate synthase	8.36–46584	0.84–2179	Liver metastasis
	P09411 PGK1_MOUSE Phosphoglycerate kinase	7.52–44405		
	P00558 PGK1_HUMAN (P00558) Phosphoglycerate kinase	8.30–44596		
2106	P00173 CYB5_RAT Cytochrome b5	4.9–15224	5.42–1264	Liver metastasis
	P02278 H2BA_HUMAN Histone H2B.a/g/k	10.32–13960		
	P02261 H2AC_HUMAN Histone H2A.c/d/i/n/p	10.32–13960		
5510	Q64442 DHSO_MOUSE Sorbitol dehydrogenase	6.60–40091	0.31–8263	Liver metastasis
	P32754 HPPD_HUMAN 4-hydroxyphenylpyruvate dioxygenase	6.50–44803		
	P26516 PSD7_MOUSE 26S proteasome non-ATPase regulatory subunit	6.29–36540		
6509	P15650 ACDL_RAT Acyl-CoA Dehydrogenase	7.63–47873	0.99–5605	Liver metastasis
	P09606 GLNA_RAT Glutamine synthetase	6.64–42268		
	P35505 FAAA_MOUSE Fumarylacetoacetase	6.92–46103		
3204	O75947 ATPQ_HUMAN ATP synthase D chain, mitochondrial	5.22–18360	4.87–3824	Primary tumour
	O2362 RS7_XENLA 40S ribosomal protein S7	10.09–22184		
	P32119 PDX2_HUMAN Peroxiredoxin 2	5.66–21892		
6301	P60174 TPIS_HUMAN Triosephosphate isomerase rase (EC 5.3.1.1) (TIM)	6.51–26538	0.49–2495	Primary tumour
	P30041 PDX6_HUMAN Peroxiredoxin 6	6.02–24904		
	P34062 PSA6_HUMAN Proteasome subunit alpha type 6	6.35–27399		
QUADRUPLETS (1 SPOT)				
5307	P46953 3HAO_RAT 3-hydroxyanthranilate 3,4-dioxygenase	5.57–32582	0.53–9138	Liver metastasis
	P82979 HCC1_HUMAN Nuclear protein Hcc-1	6.10–23671		
	P97532 THTM_RAT 3-mercaptopyruvate sulfurtransferase	5.88–32809		
	P46952 3HAO_HUMAN (P46952) 3-hydroxyanthranilate 3,4-dioxygenase	5.62–32542		
QUINTUPLETS (1 SPOT)				
8508	P17764 THIL_RAT Acetyl-CoA acetyltransferase	8.92–44695	0.96–5988	Liver metastasis
	P24752 THIL_HUMAN	8.98–45199		
	P13437 THIM_RAT 3-ketoacyl-CoA thiolase	8.09–41871		
	P00883 ALFA_RABIT Fructose-bisphosphate aldolase	8.40–39211		
	O09171 BHMT_RAT Betaine-homocysteine S-methyltransferase	8.02–44976		

Table 2. Results obtained from the Quantitative Theory of Peak Overlapping (*m* values) and SDO model (*s*, *d*, *t*, *q*, *q* values): comparison between estimated and experimental data.

	Counted spots (p)	Number of proteins (m)	Number of detectable peaks (p)	s :number of singlets	d :number of doublets	t :number of triplets	q :number of quadruplets	q :number of quintuplets
1 st strip	21	20	20.3	14.7	4.0	1.1	0.3	0.1
2 nd strip	18	21	15.7	11.8	2.9	0.7	0.2	0.0
3 rd strip	19	31	20.5	13.6	4.6	1.5	0.5	0.2
4 th strip	16	23	18.2	14.4	3.0	0.6	0.1	0.0
Total map	74	103	74.7	54.5	14.6	4.0	1.1	0.3
Experimental			74	52	14	6	1	1

The number of pure peaks, singlets, doublets, triplets, quadruplets, quintuplets were computed for a critical interdistance $x_0 = 0.05$ pH along the *pI* axis.

first-dimension IEF separation, *i.e.*, the range of the IPGs adopted (in this case pH 3–10); (iii) Scanner resolution, *i.e.*, standard (1 mm) or high (0.5 mm); (iv) Sensitivity of the detection system; the ability of the MS analytical system to identify the presence of two or more components in one spot, *i.e.*, the contamination level of the major component in a spot. It is evident that any information on the purity degree of each spot depends on these experimental data.

4 Discussion

4.1 The theory of spot overlapping at work

This is the first instance in which our theory of spot overlapping has been applied to a real experimental 2-D PAGE gel after its development and validation on computer-generated maps [17, 18]. The excellent agreement with data obtained from MS analysis of a real 2-D map proves that it is a powerful and robust method for accurately predicting the overlapping degree present in a map. It must be emphasized that the present data represent more of an exception than the rule, due to the very minute sample applied and the presence of only 264 total spots detectable. Thus, this is one of the few cases in which the number of singlets would be in the majority over the other spots (71%). In crowded conditions, as is typical of a tissue homogenate under normal loading conditions (*ca.* 1 mg total protein) and standard gel sizes (18 × 20 cm, IEF × SDS-PAGE), the singlets would be by far the least abundant species. The other surprising result is the fact that the only quintuplet found was in a *pI* region (pH *ca.* 9) less crowded with protein spots (statistical analysis has shown that most proteins crowd around the pH 4–6 region, *ca.* 60% of the total) [28].

4.2 How to minimize spot overlap

There are three major ways to increase the total number of spots seen, while diminishing spot overlapping: either increase the gel size, or exploit narrower and narrower pH gradients over the same separation distance, or both. The first avenue was explored long ago by Young [29, 30], who first adopted giant gels (37 cm in the first, 39 cm in the second dimension) out of frustration in his inability to detect protein changes in cells stimulated with adrenal steroid hormones. He had been using conventional size 2-D gels and had been able to resolve and detect only 700 to 900 spots. When increasing the gel area by a factor of 6, and loading concomitantly an amount of protein *ca.* 100-fold greater, he could resolve and detect a total of *ca.* 5000 polypeptide chains and was thus able to monitor many hormone-induced changes in protein levels. Young [30] stated that “many of the major spots visible on the smaller gels, and especially the streaks, were resolved into multiple spots on the larger gels. Thus, spots presumed to represent single proteins on smaller gels often in fact represent several”. The other outstanding result came from the work of Klose and Zeindl [31], who reported the resolution and detection of at least 10 000 different polypeptides in an epithelial-like human larynx carcinoma cell line. Here too the data were obtained after substantial experimental effort: the cell line had been fully labelled with ¹⁴C-amino acids, the gels used were of giant size (30 cm in the first, 42 cm in the second dimension) and detection had to be done sequentially by using multiple exposures for revealing the faintest spots.

The other remedy to adopt is to run the so-called “zoom gels”, *i.e.*, a series of narrow-range IPG strips (covering no more than 1 pH unit) [32, 33]. The third path was explored by Oguri *et al.* [34], in what they call “cybergel” technology. Not only did they run “zoom gels” by using seven different narrow-range (1 pH unit) IPG gels, covering the entire pH 3–10 range, over the standard 18-cm length of the pH axis, but

they also greatly increased the length of the SDS-PAGE dimension by running gels of either 35 cm or even 55 cm length, instead of the standard 20 cm length. All the gel images were then “stitched” electronically into a 70 × 67 cm “cybergel”. The increment in spot detection (and resolution) was impressive: from a total of 853 spots in a standard-size gel up to 6677 spots, an increase of 783% in total spot count.

However, even these remedies do not seem immune from some surprising results. A case in point is represented by a paper of Gygi *et al.* [35], who, in search of the “missing” low-abundance proteins (those with a codon bias value <0.1) in a total yeast lysate, prepared a narrow IPG range (4.5–5.5) over a 25 cm gel length (resulting in a pH slope of 0.004/mm, 1 order of magnitude shallower than the pH gradients reported here!). By loading high sample levels (0.5 mg total protein), they could visualise no fewer than 1500 spots, a truly remarkable resolution/detection capability. Yet, when excising a 4 cm² gel area and analysing 50 proteins therein, they found a spot to be composed of a “sextuplet”!

4.3 On the ΔpI and ΔM_r spread in overlapping spots

Some other interesting conclusions can be drawn by a close inspection of Table 1, where in the subsections listing doublets, triplets, quadruplets and quintuplets, we have listed also the ΔpI and ΔM_r values of all proteins found in a “single spot”. Considering that when cutting out a gel segment for MS analysis we generally excise an ellipsoid of *ca.* 3 mm (pH axis) × 2 mm (M_r dimension), we would expect to collect on average, in each single gel segment, a pI interval of 0.2 pH units maximum and 1000 Da maximum, respectively. This is respected in only a few cases, *e.g.*, spots no. 8509, 3303, 7501, 2710 and 6501 among the doublets. In the vast majority of cases, the pI spread is as wide as 1 pH unit. What does this mean, considering the high precision of the IPG technology and the excellent correlation between theoretical pI s of pro-

teins and their expected position along the IPG gradient, as demonstrated by Bjellqvist *et al.* [36]? In some cases it might simply mean that the theoretically predicted pI is wrong, due to some erroneous sequences stored in databases (as in fact elegantly found in [36]). But in the majority of cases it might simply mean that the proteins of diverging pI values collected in a single spot might represent post-translationally modified proteins, such as glycosylated, phosphorylated and deamidated proteins and any modification altering the surface charge.

The other surprise comes when analysing the ΔM_r data. Here too the interval very rarely encompasses only 1000 Da; in most cases the M_r spread goes from 2000 to 7000 Da and even higher. What does this mean? In a few cases, one might consider that some of the spots with the largest ΔM_r values could represent cleaved products. But it cannot be excluded that in a number of other cases, this large spread might be due to an aberrant binding of SDS. Although Pitt-Rivers and Impiombato [37] demonstrated long ago that the classic SDS/protein ratio is 1.4:1 mg, which accounts for the fact that there is a very good relationship between migration behaviour and M_r , quite a few exceptions have been found, notably for histones [38], glycoproteins [39] and others [40].

4.4 On the significance of apparently “aberrant” spots

Some proteins, however, were found in spots that did not match the predicted M_r or pI by a divergence that cannot be explained by either of the ways discussed so far. Table 3 presents these proteins together with others identified in the respective spots. As can be seen from the number of sequenced peptides and the resulting protein MASCOT score, all proteins were identified unambiguously. When investigating the M_r and pI , however, it is obvious that tubulin has too low an M_r for spot 2710, while histone 2A and 2B have

Table 3. Spots that did not match the predicted M_r or pI

SSP number	pID	MASCOT Score	peptide	M_r (kDa)	pI	comment
2710	P08238	869	25	83.4	4.97	
	P05218	161	6	50.1	4.78	M_r low
2106	P00173	571	15	15.2	4.9	
	P02278	105	3	13.7	10	pI high
	P02261	65	2	13.9	10.9	pI high
	gil16877384	304	9	44 (23.7)	8.8 (6.1)	(gi32129199)
3204	P23821	239	8	22.1	10	pI high
	O75947	270	7	18.4	5.2	
	P32119	152	4	22	5.6	
5103	P08228	319	5	15.9	6	
	P02278	55	2	13.1	10.3	pI high

too high a *pI* for spot 2106. The same is true for the ribosomal protein S7 in spot 3204 and histone H2B in spot 5103. The best explanation may come from the fact that they only have odd behaviour in one dimension, either M_r or *pI*, and that they are very abundant proteins: tubulin, ribosomal protein S7, and histones H2A and H2B. While we prepared our gels with utmost care, we still had remnants of streaking visible in Fig. 1. It is therefore conceivable that the sensitivity was so high that we started picking up background in the 2-D gel. In line with this, we found histone H2A in two spots, 5103 and 2106, which migrated at nearly identical M_r .

We encountered one other error when using NCBI, a richer database than Swiss-Prot, but also less annotated. In spot 2106 we found an additional protein gi16877384 that has a high M_r and *pI* for spot 2106. When reinvestigating our data we realised that for gi16877384 all peptides match the C-terminal region of this protein. We then actually found another database entry, gi32129199, to which all peptides match and that covers only the C-terminal part of gi16877384. This entry fits with *pI* 6.1, in good agreement with the observed value. It has a slightly lower mass than expected from the position of the spot. This may indicate that yet another isoform of the protein is present that has not yet been deposited in the database.

Thus, paradoxically, careful examination of the types of proteins found in overlapping spots can yield some information on the validity of spot detection by MS (in our case, the error seems to be of the order of only 1–2%). By the same token, however, the identity of at least one protein found in the sextuplet of Gygi *et al.* [35] (see their Table 2) must also be discussed: protein VMA1 (M_r 67.7 kDa) cannot simply be found in an envelope of spots all having M_r values between 48 and 50 kDa.

4.5 Protein analysis by mass spectrometry

Spots excised from the gel were analysed by LC-MS to get a better coverage of potential mixtures than is possible by MALDI-TOF PMF. LC allows for separation of peptides and therefore simplifies spectra and furthermore extends analysis time. From the peptides that had overlapping elution from the column, peaks were selected for fragmentation by intensity, *i.e.*, the most intense first. Nevertheless, this did not bias the analysis towards more abundant factors at the expense of less abundant proteins above the sensitivity threshold of our technology. During our LC-MS analysis there was ample time to sequence a lot more peptides than we actually did. This means that the analysis was not limited by time, *i.e.*, exhaustive with respect to how many proteins could be identified with our technology. Using better equipment or improved analysis procedures would allow improvement of the detection limits and hence would likely result in more proteins to be identified. If consequently thought through, one easily realises that singlets do not exist. In fact, even “empty” spaces on the gel are likely to contain many proteins, just below the detection limit of the staining

procedure. Today, MS supersedes the sensitivity of staining procedures such as Coomassie or silver staining. However, radioactive labelling allows proteins to be spotted even when nothing can be detected by MS. In the end, looking at maps of protein on 2-D gels resembles looking at the sky full of stars. Every improvement in technology allows one to map as yet unseen objects. The key message is therefore qualitative and not quantitative. It is quantitative, however, with respect to the current state of the art and will have to be adjusted with improving technology.

Supported by grants from AIRC (Associazione Italiana per la Ricerca sul Cancro) to P.G.R. and J.R., and from MIUR, Roma (PRIN-2003, FIRB-2001, No. RBNE01KJHT) and Cariverona (2004) to P.G.R. Additionally supported by the Italian University and Scientific Research Ministry (grants no. 2001033797_001 and 2003039537_005) Rome, and by the University of Ferrara. We thank Dr. N. Marchetti for algorithm development.

5 References

- [1] Mahler, H. R., Cordes, E. H., *Biological Chemistry*, Harper & Row, New York 1971.
- [2] O'Farrell, P. H., *J. Biol. Chem.* 1975, 250, 4007–4021.
- [3] Anderson, N. L., Anderson, N. G., *Proc. Natl. Acad. Sci. USA* 1977, 74, 5421–5426.
- [4] Scheele, J. A., *J. Biol. Chem.* 1975, 250, 5375–5380.
- [5] Laemmli, U. K., *Nature* 1970, 227, 680–685.
- [6] Margolis, J., Kenrick, K. G., *Anal. Biochem.* 1968, 25, 347–354.
- [7] Righetti, P. G., *Isoelectric Focusing: Theory, Methodology and Applications*, Elsevier, Amsterdam 1983, pp. 1–386.
- [8] Righetti, P. G. *Immobilized pH Gradients: Theory and Methodology*. Elsevier, Amsterdam 1990, pp. 1–400.
- [9] Righetti, P. G., Stoyanov, A. L., Zhukov, M. Y., *The Proteome Revisited*, Elsevier, Amsterdam 2001, pp. 1–395.
- [10] Fountoulakis, M., Takacs, M. F., Berndt, P., Langen, H. *et al.*, *Electrophoresis* 2000, 20, 2181–2195.
- [11] Herbert, B. R., Sanchez, J. C., Bini, L., in: Wilkins, M. R., Williams, K. L., Appel, R. D., Hochstrasser, D. F. (Eds.), *Proteome Research: New Frontiers in Functional Genomics*, Springer, Berlin 1997, pp. 13–33.
- [12] Pedersen, S. K., Harry, J. L., Sebastian, L., Baker, J. *et al.*, *J. Proteome Res.* 2003, 2, 303–312.
- [13] Dondi, F., Bassi, A., Cavazzini, A., Pietrogrande, M. C., *Anal. Chem.*, 1998, 70, 766–773.
- [14] Pietrogrande, M. C., Cavazzini, A., Dondi, F., *Rev. Anal. Chem.* 2000, 19, 123–155.
- [15] Pietrogrande, M. C., Marchetti, N., Dondi, F., Righetti, P. G., *Electrophoresis* 2002, 23, 283–291.
- [16] Pietrogrande, M. C., Marchetti, N., Dondi, F., Righetti, P. G., *Electrophoresis* 2003, 24, 217–224.
- [17] Davis, J. M., *Anal. Chem.* 1991, 63, 2141–2152.
- [18] Davis, J. M., *Anal. Chem.* 1993, 65, 2014–2023.

- [19] Giddings, J. C., *J. Chromatogr. A* 1995, 703, 3–15.
- [20] Marengo, E., Robotti, E., Righetti, P. G., Campostrini, N. *et al.*, *Clin. Chim. Acta* 2004, 345, 55–67.
- [21] Campostrini, N., Pascali, J., Hamdan, M., Aster, H. *et al.*, *J. Chromatogr. B* 2004, 808, 279–286.
- [22] Pastorino, F., Brignole, C., Marimpietri, D., Cilli, M. *et al.*, *Cancer Res.* 2003, 63, 7400–7409.
- [23] Herbert, B., Galvani, M., Hamdan, M., Olivieri, E. *et al.*, *Electrophoresis* 2001, 22, 2046.
- [24] Shevchenko, A., Wilm, M., Vorm, O., Mann, M. *Anal. Chem.* 1996, 68, 850–858.
- [25] Rappsilber, J., Ishihama, Y., Mann, M., *Anal. Chem.* 2003, 75, 663–670.
- [26] Neuhoff, V., Arold, N., Taube, D., Ehrhardt, W., *Electrophoresis* 1988, 9, 255–262.
- [27] Merrill, C. R., Washart, K. M., in: Hames, B. D. (Ed.), *Gel Electrophoresis of Proteins: A Practical Approach*, Oxford University Press, Oxford 1998, pp. 53–91.
- [28] Gianazza, E., Righetti, P. G., *J. Chromatogr.* 1980, 193, 1–8.
- [29] Voris, B. P., Young, D. A., *Anal. Biochem.* 1980, 104, 478–484.
- [30] Young, D. A., *Clin. Chem.* 1984, 30, 2104–2108.
- [31] Klose, J., Zeindl, E., *Clin. Chem.* 1984, 30, 2014–2020.
- [32] Hoving, S., Voshol, H., van Oostrum, J., *Electrophoresis* 2000, 21, 2617–2621.
- [33] Westbrook, J. A., Yan, J. X., Wait, R., Welson, S. Y., Dunn, M. J., *Electrophoresis* 2001, 22, 2865–2871.
- [34] Oguri, T., Takahata, I., Katsuta, K., Nomura, E. *et al.*, *Proteomics* 2002, 2, 666–672.
- [35] Gygi, S.P., Corthals, G.L., Zhang, Y., Rochon, Y., Aebersold, R., *Proc. Natl. Acad. Sci. USA* 2000, 97, 9390–9395.
- [36] Bjellqvist, B., Hughes, G., Pasquali, C., Paquet, N., Ravier, F. *et al.*, *Electrophoresis* 1993, 14, 1023–1031.
- [37] Pitt-Rivers, R., Impiombato, F. S. A., *Biochem. J.* 1968, 109, 825–833.
- [38] Panyim, S., Chalkley, R., *Arch. Biochem. Biophys.* 1969, 130, 337–346.
- [39] Segrest, J. P., Jackson, R. L., Andrews, E. P., Marchesi, V. T., *Biochem. Biophys. Res. Commun.* 1971, 44, 390–395.
- [40] Nelson, C. A., *J. Biol. Chem.* 1971, 246, 3895–3901.



Proteomic approaches for studying chemoresistance in cancer

Pier Giorgio Righetti[†], Annalisa Castagna, Paolo Antonioli, Daniela Cecconi, Natascia Campostrini and Sabina Carla Righetti

The role of various proteins involved in drug resistance in tumor cells is discussed in this review. Two types of studies are covered: those performed in the preproteomics era and those carried out with modern proteomic tools, namely 2D (electrophoretic) maps and 2D chromatography. In the preproteomic studies, one protein had generally been held responsible for a given chemoresistance. However, analysis via proteomic tools may reveal entire sets of proteins that are up- or downregulated (or switched on/off) in chemoresistant tumor cell lines compared with parental tumor lines. Therefore, it appears more realistic to expect that exposure of cells to drugs results in the activation of different mechanisms of resistance. Such investigations have led to the broadly shared opinion that exposure of cells to drugs results in the activation of different mechanisms of resistance, and that a specific drug-resistant phenotype consists of several molecular mechanisms that are simultaneously active. The proteomic papers reviewed clearly support the hypothesis that many metabolic pathways are affected during the resistance process. Although the modulation of expression levels of such proteins is not clear proof of their role in drug resistance per se, at least some of the themes are very likely to be involved in the resistance phenotype, and thus may be potential targets for new drugs. It is hoped that this review will bring new insight in this field and will stimulate novel and deeper searches with proteomic tools (including prefractionation of subcellular organelles, such as nuclei, to bring to the fore low-abundance proteins that might be responsible for the onset of drug resistance).

Expert Rev. Proteomics 2(2), 215–228 (2005)

Proteomics & oncology: a general survey

Cancer is a complex disease that represents the end product of a multistep biologic process including growth of neoplastic cells at the primary tumor site, invasion of the host tissue, angiogenesis, intravasation in the circulatory system, arrest and extravasation into a new tissue and finally growth at a secondary site [1]. Despite the knowledge of genetic mechanisms driving these events at the tissue level, the biochemical end points are not fully characterized and treatment is not optimized. The ability to accurately profile a cancer would have a profound effect on the quality of the treatment received by a patient. The identification of molecular markers assisting in the subtyping of cancer could be useful for an appropriate and uniform classification of cancer for improving the diagnosis and prognosis of disease.

Presently used tumor markers are often not sufficient for diagnosing cancer, since their levels can be elevated in people with benign conditions or with nontumoral ailments; additionally, many tumor markers are not specific for a particular type of cancer.

Currently, the main application of tumor markers is for assessing a cancer's response to treatment and checking its recurrence. In the past 20 years, research activities have mainly focused on the genetic characteristics of cancer cells [2,3]. However, in recent years it has been realized that this strategy, despite having produced insight into the biology of cancer cells, would be insufficient for a deeper and global comprehension of the mechanisms involved in cancer, and the real key players of all physiologic and pathologic processes had to be called in the arena, namely the proteins [4]. The term

CONTENTS

Proteomics & oncology:
a general survey

Chemoresistance studies
in the pre- (early)
proteomic era

Chemoresistance studies
in the proteomic era

Expert opinion

Five-year view

Key issues

References

Affiliations

[†] Author for correspondence
University of Verona, Department
of Industrial & Agricultural
Biotechnologies, Laboratory of
Proteome Science, Strada Le
Grazie 15, Verona 37134, Italy
Tel.: +39 045 802 7901
Fax: +39 045 802 7929
righetti@sci.univr.it

KEYWORDS:

2D chromatography, 2D maps,
14-3-3 proteins, biomarkers,
cancer, chemoresistance, heat
shock proteins, molecular
diagnostics, oncoproteomics,
protein profiling

cancer proteome refers to the collection of proteins expressed by a given cancer cell and should be considered a highly dynamic entity within the cell, causative of cellular activities [5,6]. As such, proteomics aims not only to identify, catalog and characterize proteins, but also to understand how they interact to affect the overall metastatic progression. Studies of global protein expression in human tumors (e.g., liver, prostate, breast, bladder and esophagus) have led to the identification of various polypeptide markers, that can potentially useful as diagnostic tools [7]. In the last few years, a new subdiscipline of proteomics, referred to as clinical proteomics, has entered the arena of cancer research, where, along with new technologies currently under development, it will focus on the discovery of the next generation of targets and imaging biomarkers [8–20].

An exciting approach that has recently emerged attempts to replace single molecule discovery efforts with serum proteomic pattern diagnostics [8–10]. The concept behind pattern diagnostics is that the blood plasma proteome reflects tissue and organ pathology, generating patterns of protein changes that have diagnostic potential without even knowing the identity of the individual proteins. Since mass spectrometry (MS)-based approaches provide a pattern of peaks, the idea is that these patterns can discriminate certain diseases. The diagnostic tool is thus represented by the pattern or signature of the proteins, rather than their identities [9]. For this approach, researchers have used surface-enhanced laser desorption/ionization (SELDI) time-of-flight (TOF)-MS. SELDI is similar to matrix-assisted laser desorption/ionization (MALDI) as the

target molecules are laser desorbed and ionized for analysis by MS. However, SELDI molecules are present on the surface of protein chips, which have an active surface chemistry (hydrophobicity, anion and cation exchangers, metal or biologic affinity) to retain proteins with complementary properties. For serum proteome pattern diagnostics, samples from affected and control persons are individually applied to a chip and the retained proteins are subsequently ionized and detected by TOF-MS. Sophisticated bioinformatic software is then used to compare spectra and determine discriminatory patterns of peaks within samples of unhealthy individuals. In the first proof-of-principle study, sera from 100 females were screened for ovarian cancer: the proteomic pattern that was used to identify ovarian cancer appeared to exhibit a predictive value of 94%, a specificity of 95% and a sensitivity of 100% [21]. Such a proteomic signature, exploiting SELDI-MS, has also been developed for discriminating individuals with prostate cancer from healthy specimens [22]. However, whether or not SELDI-MS and bioinformatics will live up to such expectations in the near future is the subject of an ongoing and lively debate [23–25]. Some researchers contest that SELDI is not sensitive enough and captures only high-abundance proteins, and thus it is not suitable for measuring true cancer biomarkers. Of equal concern is the reproducibility of the technique.

Indeed, characterizing the human plasma proteome has become a main goal in the proteomics arena. The plasma proteome is perhaps the most complex in the human body: it consists not only of the resident, hemostatic proteins, but also of

Table 1. Differentially expressed proteins in chemoresistant melanoma cell lines in the IPG pH 4–8 range (modified from [53]).

Protein	Molecular mass (kDa)	Isoelectric point	Expression
HSP60	57.6	5.2	Up
HSP60 variant (chaperonin)	57.8	5.1	Up
HSX70 variant	64.3	5.4	Up
HSP27	28.3	6.0	Up
HSP27 variant/hydroxyacyl-coenzyme A dehydrogenase II	26.9	7.7	Up
β1 subunit of G-proteins	39.1	5.5	Up
β2 subunit of G-proteins	38.1	5.4	Up
Nicotinamide <i>N</i> -methyltransferase	29.5	5.6	Up
Peroxiredoxin 1	22.1	8.7	Up
Proteasome subunit β-type 3	22.9	6.1	Up
Hypothetical protein DKFZp566J2046	24.9	7.0	Up
HSP90	89.6	4.9	Down
Nucleophosmin	37.8	4.6	Down
Phosphoglyceromutase	28.7	6.9	Down

HSP: Heat shock protein; IPG: Immobilized pH gradient.

Table 2. Differentially expressed proteins in chemoresistant melanoma cell lines in the IPG pH 8–11 range (11 out of 20 spots) (modified from [53]).

Protein	Molecular mass (kDa)	Isoelectric point	Expression
Collagen-binding protein 1 and 2	47	9.5	Up
Human aldolase A	40	8.1	Up
Electron transfer flavoprotein	28	8.2	Up
Voltage-dependent anion-selective channel protein (porin)	30	8.4	Up
GTP:AMP phosphotransferase	25	9.6	Up
Calcyclin-binding protein	26	8.3	Up
ATP synthase α -chain	55	9.9	Up
Glutathione reductase	52	7.7	Up
Heterogenous nuclear ribonucleoprotein A2/B1	41	8.3	Up
Biphenylhydrolase related protein	32	8.8	Up
Galectin 3	25	7.9	Up

IPG: Immobilized pH gradient.

immunoglobulins, cytokines, protein hormones, secreted and foreign proteins, indicative of infection. In addition, blood circulates through almost all tissues of the body and thus contains tissue-leakage proteins, including those released from damaged or dying cells. The blood should, therefore, contain information on the physiologic state of all tissues in the body. This, combined with its accessibility, makes the blood proteome invaluable for medical purposes. However, characterizing the proteome of blood plasma will be a daunting challenge. In addition to its immense repertoire of proteins present, the dynamic range of these proteins is on the order of 10^9 , with serum albumin being the most abundant (40–50 mg/ml) and low-level proteins such as interleukin-6 present at 0–5 pg/ml [26]. In an attempt to mine the deep plasma proteome, Pieper and coworkers, via a depletion technique, used immunoaffinity subtraction chromatography to remove the nine most prominent species prior to 2D chromatography (anion-exchange followed by size-exclusion chromatography) [27]. The resulting sample was then analyzed by electrophoretic 2D maps. Although it cannot be excluded that several protein species could be lost via spurious adsorption onto the various chromatographic sorbents in all these prefractionation steps, this operation reveals a relatively large number of proteins (~100 species) previously not described in sera. Nevertheless, notwithstanding the impressive amount of work involved, no new markers for disease could be found as yet. The problem with this depletion approach could be the codepletion of other proteins in sera bound to the immunosubtracted ones, or recognition of some epitopes of the antisera utilized. In addition, subfractionation onto chromatographic columns could further lead to protein losses. Aware of these limitations, the authors are now proposing a quite revolutionary strategy, based on the principle of equalizing the various species present in sera. Equalization is achieved via beads containing a library of linear hexamers made

using 20 amino acids, which can contain up to 64 million different chemistries. When a complex protein mixture of a first variance, such as human serum, is incubated with this library, representatives of each component of the mixture will bind to these individual ligands. In large overloading conditions, high-abundance proteins saturate their specific affinity ligand and excess is removed during the washing step, while low-abundance proteins continue to concentrate on their specific affinity ligands. After processing, representatives of all the original components can be eluted to produce a sample with a second variance where each representative is still present but in vastly different concentrations. Proteins of high abundance are significantly diluted while low-abundance species are concentrated [28–30]. It is anticipated that this approach will bring a genuine revolution in detecting even the most dilute serum components, thus hopefully exposing the much sought after disease biomarkers.

Having provided a panoramic survey on the relationship between oncology and proteomics, focus will now be placed on the problem of resistance of tumor cells to drug treatment. Cancer chemotherapy has gradually improved over the last 30 years with the development of novel antitumor drugs of higher efficiency and decreased toxicity. While treatment of certain malignancies with chemotherapy has been successful and encouraging, the effectiveness has often been limited by drug resistance of tumor and by side effects on normal tissues and cells. Many tumors appear to be intrinsically resistant to several of the more potent cytotoxic agents used in cancer therapy. Other tumors, although initially sensitive, recur and develop resistance not only to initial therapeutic agents, but also to other drugs not used in the treatment, a phenomenon referred to as multidrug resistance (MDR) [31,32]. Although this poses a serious problem in medical treatment, the molecular mechanisms of drug resistance are not yet fully understood.

Table 3. Differentially expressed proteins in thermoresistant stomach cancer cell lines in the IPG pH 4–8 range (modified from [54]).

Protein	Molecular mass (kDa)	Isoelectric point	Expression
14-3-3 σ (stratifin)	27.8	4.7	Up
BIP, glucose-regulated protein 78	72.3	5.1	Up
Calnexin	67.6	4.5	Up
Calreticulin	48.1	4.3	Up
Cancer oncogene (homologous to)	–	–	Up
Cyclin D2	33.1	5.1	Up
Cyclin D2 variant	–	–	Up
Cytokeratin 7	51.3	5.5	Up
Cytokeratin 8	53.5	5.5	Up
Cytokeratin 19	44.1	5.1	Up
FK506-binding protein 4	51.7	5.4	Up
Glutathione transferase M3	26.4	5.4	Up
HSP27	28.3	6.0	Up
HSP27 variant	–	–	Up
HSX70 variant	–	–	Up
HSP90	83.1	5.0	Up
Nucleophosmin (nucleolar protein B23, numatrin)	32.6	4.6	Up
Reticulocalbin (high Mr)	38.9	4.9	Up
Thioredoxin peroxidase (1 or 2? [§])	21.9/22.1	5.7/8.3	Up
14-3-3 η	28.1	4.8	Down
Aldehyde dehydrogenase 1	54.7	6.3	Down
Annexin 1, lipocortin 1	38.6	6.6	Down
Annexin 1, lipocortin 1 variant	–	–	Down
β 1 subunit of G-proteins	–	–	Down
β 2 subunit of G-proteins	–	–	Down
Initiation factor 5A (eIF-5A)	16.7	5.1	Down
Initiation factor 5A (eIF-5A) variant	20.2	6.5	Down
Prohibitin	29.8	5.6	Down
Reticulocalbin (low Mr)	36.9	4.3	Down
Rho GDI, GDP dissociation inhibitor (homologous to) (1 or 2 or 3? [§])	23.2/23.0/25.1	5.0/5.1/5.5	Down
Transgelin 2/neuropolypeptide h3	22.4/20.9	8.4/7.4	Down
Translationally controlled human tumor protein	19.6	4.8	Down
Vimentin	53.5	5.1	Down

[§]The question mark refers to different forms of a protein, not specified in the original article. The authors provide the molecular mass and isoelectric point values of the various forms.

HSP: Heat shock protein; IPG: Immobilized pH gradient; Mr: Molecular mass.

In this review, several mechanisms, including drug transporters, cellular stress response, apoptosis regulation and cellular survival signals, that have relevance to drug resistance will be discussed. These pathways could provide new targets for effective cancer therapy. The discussion will be divided into two parts: those findings obtained in the pre- (or early) proteomic era and those related to proteomic tools, via either differential 2D map analysis [33] or its chromatographic counterpart [34].

Chemoresistance studies in the pre- (early) proteomic era

These studies are characterized by the fact that the various proteins held responsible for chemoresistance have been revealed by traditional means, such as enzyme activity measurements, western blots followed by immunodetection, measurement of mRNA levels and gene manipulations. In general, these studies have pointed out a single protein held responsible for a given phenomenon of chemoresistance in a particular type of tumor. In fact, some interesting studies on MDR based on 2D polyacrylamide gel electrophoresis (PAGE) had already been published in the mid-1980s [35]; however, these very early studies could not provide large amounts of information, as the MS technique was not yet applied to protein identification.

Conversely, proteomic studies, due to their ability to simultaneously explore thousands of polypeptide chains, have resulted in the detection of a whole panel of proteins either up- or downregulated, or turned on or off, in chemoresistance studies. The section that follows has been made possible by an exhaustive review of such earlier studies written by Tsuruo and coworkers [36].

The ATP-binding cassette transporters

Studies on the MDR phenotype have led to the discovery of ATP-binding cassette (ABC) transporters, such as P-glycoprotein (P-gp). Overexpression of P-gp, encoded by the *MDR1* gene, confers resistance to a variety of structurally and functionally unrelated antitumor drugs such as vinblastine, vincristine, doxorubicin, daunorubicin, etoposide, paclitaxel and quite a few others [37–39]. P-gp is a 180-kDa glycoprotein localized on the plasma membrane of resistant cancer cells. It can bind antitumor drugs as soon as they are delivered inside the cell or even earlier, during the crossing of the plasma membrane, and expel them outside in an ATP-dependent manner [40]. Expression of P-gp is elevated in intrinsically drug-resistant cancers of the colon, kidney and adrenal glands, as well as in some tumors that acquire drug resistance after chemotherapy. Conversely, various compounds, including calcium channel blockers and calmodulin inhibitors, such as verapamil, PSC-833 and MS-209, known to be potent P-gp inhibitors, have been shown to enhance the cytotoxic activity of various agents. In addition to small-molecular-mass P-gp inhibitors, monoclonal antibodies against P-gp have been shown to be quite potent in reversing MDR in drug-resistant cells.

Proteasome involvement in stress-mediated resistance of solid tumors

The proteasome is a huge, barrel-shaped supramolecular complex, ca. 870 kDa, that is formed by the assembly of α and β subunits, delimiting an empty central cavity where proteolytic processes take place. It is the major site for degradation of abnormal or

Table 4. Differentially expressed proteins in doxorubicin-resistant breast cancer cells (modified from [55]).

Protein	Molecular mass (kDa)	Isoelectric point	Expression
Cyclophilin A	18.24	7.83	+2.1
FK-506 binding protein	12.01	8.42	+4.1
Glyceraldehyde-3-phosphate dehydrogenase	36.22	8.73	+2.5
Heat shock protein 27	22.42	8.14	+6.0
Metallothionein	7.21	7.80	+10.7
40S ribosomal protein S28	7.90	11.32	+2.5
40S ribosomal protein S4	29.83	10.95	+6.0
Stathmin	17.30	5.87	+2.0
Superoxide dismutase	16.16	6.08	+2.0
Telomerase binding protein	18.98	4.16	+2.7
Thioredoxin	12.02	4.65	+6.4
Thymosin- β -10	5.03	5.14	+4.2
Prothymosin α	12.20	3.49	+2.0
Ubiquitin	8.56	7.53	+3.3
Ubiquitin-like protein NEDD8	9.07	9.12	+7.2

irrelevant and regulatory proteins such as transcription factors, oncoproteins and cyclins. Solid tumor cells often grow under environmental stress conditions, such as glucose deprivation, hypoxia, low pH and nutrient deprivation, due to inadequate vascularization. Such a microenvironment, by itself, could be involved in drug resistance as it hinders drug access to tumor cells and reduces the oxygen radicals generated by antitumor drugs. Recent research has demonstrated that stress conditions indeed stimulate nuclear accumulation of proteasome in HT-human colon as well as ovarian cancer cells [41]. Such stressed cells were shown to become resistant to etoposide and doxorubicin, both of which are antitumor topoisomerase II poisons. Proteasome inhibitors (e.g., lactacystin) were found to attenuate the inducible chemoresistance by inhibiting the topoisomerase II depletion induced by glucose starvation and hypoxia. Since topoisomerase II restoration was seen only at the protein level, this indicates that depletion of this protein occurred through a proteasome-mediated degradation mechanism. This suggests that proteasome inhibition may be useful for overcoming various types of drug resistance.

The proteasome is involved in the activation of nuclear factor (NF)- κ B, the central transcription factor of the immune system; limited research suggests that NF- κ B, which has antiapoptotic properties, may play a major role in inducible chemoresistance [42]. The proteasome also controls the levels of proteins that are important for cell cycle progression and apoptosis in normal and malignant cells; for example, Bcl-2, cyclins and caspases [43].

Involvement of glyoxalase I in apoptosis resistance

Apoptosis is an active cell death mechanism that plays a role in several biologic processes. Various antitumor agents have been reported to elicit apoptosis in cancer cells [44]. This implies

that blocking of apoptosis signaling could be another mechanism for MDR in chemotherapy. Glyoxalase I is an essential component in pathways leading to the detoxification of methylglyoxal, a side product of glycolysis. Methylglyoxal (due to its aldehyde group) is involved in a variety of detrimental processes, such as covalent adducts with proteins and DNA. In tumor cells, the high glycolytic activity caused by rapid deregulated growth increases the intracellular levels of methylglyoxal. Detoxification systems are required for eliminating this toxicity. The principal route for methylglyoxal catabolism is the glyoxalase pathway, which consists of the enzymes glyoxalase I and II. The enzymatic activity as well as the mRNA expression of glyoxalase I is significantly elevated in several drug-resistant cells, such as UK711, UK110 and K562/ADM, as compared with their parental cells. Since these mutant cell lines were populations that survived after treatments with etoposide or adriamycin, it appears that the development of drug resistance is accompanied by overexpression of the enzyme. In fact, when overexpressed in human leukemia cells, glyoxalase inhibited apoptosis normally induced by these two drugs, indicating a direct involvement of this enzyme in apoptosis suppression. These results indicate that glyoxalase I inhibitors are effective drug resistance-reversing agents in different types of cancer.

Akt & heat shock protein 90 as promising targets for cancer chemotherapy

The susceptibility of cells to apoptosis appears to depend on the balance between pro- and antiapoptotic (survival) signals. The fact that diverse chemotherapeutic agents induce apoptosis, whilst engaging different intracellular targets, raises the possibility

Table 5. Mass spectrometry-identified proteins in the A431 versus A431/Pt comparison.

Protein	Molecular mass (Da) – isoelectric point	Trend in A431 vs. A431/Pt comparison
Calmodulin	16,706 – 4.1	Threefold increased in A431
Calumenin	37,107 – 4.5	Off in A431
Microtubule-associated protein RP/EB1	29,999 – 5.1	Twofold increased in A431
Stathmin (phosphoprotein pp18)	17,171 – 6.0	2.2-fold increased in A431
Mitochondrial ATPase inhibitor (IF1)	12,249 – 9.7	Off in A431
Heat shock cognate 71 kDa protein (isoform 1)	70,898 – 5.4	Off in A431
T-complex protein 1, β subunit	57,488 – 6.4	Off in A431
GTP-binding nuclear protein RAN	24,423 – 7.6	5.5-fold decreased in A431
Calponin and maspin	36,414 – 5.9 and 42,139 – 5.7	4.7-fold decreased in A431
Phosphoglycerate kinase 1	44,597 – 8.5	3.5-fold decreased in A431
Potent heat-stable protein phosphatase 2A inhibitor (I1PP2A)	28,585 – 4.0	2.7-fold increased in A431

Each protein in the table represents a processed spot of A431 vs. A431/Pt and is identified by protein name, isoelectric point and molecular mass attributes and trend in the comparison under object (modified from [57]).

that anticancer drugs may induce apoptosis by decreasing survival signals, such as the serine/threonine kinase (Akt)-mediated survival-signaling pathway. When activated, Akt (also known as PKB or RAC-PK) binds to the plasma membrane, where it is phosphorylated at two key regulatory sites, Thr3⁰⁸ (by the enzyme 3-phosphoinositide-dependent protein kinase [PDK]-1) and Ser4⁷³ (by PDK2). Phosphorylation at both residues is necessary for full activation of Akt and its subsequent control of biologic responses, including apoptosis inhibition and cell cycle progression. Recent evidence also suggests that heat shock protein (HSP)90 could play an important role in the Akt signaling pathways by binding to both Akt and PDK1 [45–46]. HSP90 is an abundant and highly conserved protein involved in a diverse array of cellular processes. HSP90 acts as a chaperone for unstable signal transducers and keeps them poised for activation until they are stabilized by conformational changes associated with signal transduction. HSP, when bound to Akt, protects it from dephosphorylation. This is mediated by protein phosphatase (PP)2A, thus stabilizing its

kinase activity. HSP90 may thus function as a scaffold for Akt and its substrates. On the basis of these observations, it has been suggested that both Akt and HSP90 could be promising targets for developing new drugs suppressing chemoresistance in tumor cells.

HSP90-directed agents could affect molecules upon which tumors depend for their proliferation and survival. Recent works suggest that the HSP90 chaperone complex could be a novel target for cancer therapy [47]. Several small molecule inhibitors of HSP90 have been identified that can deplete cellular levels of multiple oncogenic client proteins simultaneously by enhancing their ubiquitination and proteasome-mediated degradation. The activity of HSP90 inhibitors has been well validated in preclinical breast cancer models, both in single-agent studies and in combination with conventional chemotherapy. One of these inhibitors, 17-allylamino, 17-demethoxygeldanamycin (17-AAG) has recently completed Phase I testing. This drug binds to a highly conserved 25-kDa N-terminal domain of HSP90, a binding site for ATP, which results in disruption of the HSP90 chaperone complex. The

Table 6. Mass spectrometry-identified proteins in the A431 versus A431 after cisplatin exposure.

Protein	Molecular mass (Da) – isoelectric point	Trend in A431 vs. A431 treated comparison
Calmodulin	16,706 – 4.1	Threefold decreased in A431 treated
F-actin capping protein	32,923 – 5.6	Fivefold increased in A431 treated
Actin (isoform γ)	41,793 – 5.4	On in A431 treated
Tropomyosin (α 3 chain)	32,819 – 4.7	Threefold increased in A431 treated
Tropomyosin (α 4 chain)	28,522 – 4.7	Threefold increased in A431 treated
Microtubule-associated protein RP/EB	29,999 – 5.1	Threefold increased in A431 treated
Annexin V	35,806 – 5.0	Twofold increased in A431 treated
14-3-3 protein epsilon	29,174 – 4.7	On in A431 treated
14-3-3 protein ζ/δ	27,745 – 4.8	Fourfold increased in A431 treated
14-3-3 protein ζ/δ	27,745 – 4.8	On in A431 treated
Proteasome activator complex	27,362 – 5.5	Fourfold increased in A431 treated
Heat shock cognate 71 kDa protein	70,898 – 5.4	Off in A431 treated
Heat shock cognate 71 kDa protein	70,898 – 5.4	Eightfold decreased in A431 treated
Heat shock cognate 71 kDa protein	70,898 – 5.4	On in A431 treated
Protein disulfide isomerase A3	56,782 – 6.3	Twofold decreased in A431 treated
Stathmin	17,171 – 6.0	Threefold decreased in A431 treated
Stathmin	17,171 – 6.0	Twofold decreased in A431 treated
Peroxiredoxin 2	21,892 – 5.9	Fourfold decreased in A431 treated
60 kDa heat shock protein	61,055 – 5.8	On in A431 treated
Peptidyl-prolyl <i>cis-trans</i> isomerase A	17,881 – 8.2	Threefold decreased in A431 treated

Each protein in the table represents a processed spot of A431 vs. A431 after 1 h cisplatin exposure and is identified by protein name, isoelectric point and molecular mass attributes and trend in the comparison under object (modified from [57]).

Table 7. Mass spectrometry-identified proteins in the A431/Pt versus A431/Pt after cisplatin exposure.

Protein	Molecular mass (Da) – isoelectric point	Trend in A431/Pt vs. A431/Pt treated
Voltage-dependent anion-selective channel protein	30,641 – 8.8	Threefold decreased in A431/Pt treated
Protein-L-isoaspartate (D-aspartate) O-methyltransferase	24,519 – 7.2	Twofold decreased in A431/Pt treated
Peroxiredoxin 6	24,904 – 6.3	On in A431/Pt treated

Each protein in the table represents a processed spot of A431/Pt vs. A431/Pt after 1 h cisplatin exposure and is identified by protein name, isoelectric point and molecular mass attributes and trend in the comparison under object (modified from [57]).

agent was well tolerated at drug exposures that were shown to cause modulation of HSP90 client protein levels [48]. HSP90 inhibitors have a synergistic effect with another class of antitumor agent: the proteasome inhibitors (e.g., PS-341, a boronic acid dipeptide). Both classes of antitumor agents block NF- κ B activation; however, since they target separate distinct levels of regulation of NF- κ B activity, combined use of subtoxic concentrations of these drugs may achieve a synergistic inhibitory effect on NF- κ B and cell survival. The strong synergistic interaction between the HSP90 inhibitor and PS-341 (bortezomib) confirms the functional significance of the upregulation of HSPs, in general, and, in particular, HSP90, as a protective mechanism against PS-341-induced apoptosis and provides the framework for combination treatments that will include HSP90 inhibitors in an effort to augment clinical efficacy and overcome clinical refractoriness to PS-341 [49].

Chemoresistance studies in the proteomic era

While in previous studies one protein has generally been held responsible for a given chemoresistance, by 2D map analysis entire sets of proteins have been found to be up- or downregulated in chemoresistant tumor cell lines compared with parental tumor lines grown in the absence of the particular drug that had induced chemoresistance. Some of the few examples of such studies that the authors could find in present-day literature are discussed and evaluated below. The experimental scenario in these studies is quite similar: a given parental cancer cell line and its drug-resistant subline (in general cultivated for a few months in low doses of a given drug, until induction of chemoresistance) are lysed in strongly solubilizing cocktails and their total protein asset displayed by 2D maps, coupling orthogonally a charge (today mostly isoelectric focusing [IEF] in immobilized pH gradients [IPG]) to a mass fractionation (sodium dodecyl sulfate [SDS]-PAGE) [16]. Master maps are created and matched via image analysis programs, such as PDQuest; the spots found to be modulated are subsequently eluted and identified via different MS methods, such as MALDI-TOF or electrospray ionization (ESI) ion-trap (IT)-MS.

Chemoresistance in human adenocarcinoma of the pancreas

Sinha and coworkers have selected two sublines of a cultured cell line derived from the adenocarcinoma of the pancreas: an MDR-resistant subline obtained in the presence of daunorubicin, and another selected in the presence of mitoxantrone

[50]. Three proteins, epidermal fatty acid-binding protein (E-FABP), cofilin and stratifin (14-3-3 σ), were found to be overexpressed in the chemoresistant cell lines, one of them (cofilin) being common to both cell sublines, the other two being modulated only in the mitoxantrone-resistant subline.

Chemoresistance in colorectal cancer & fibrosarcoma cells

As stated above, the classic mechanism of MDR is accompanied by the overexpression of P-gp, which leads to reduced accumulation of drug in the cell sap mediated by an increased ATP-dependent drug efflux. However, there are atypical MDR forms that do not depend on the P-gp mechanism of resistance. Sinha and coworkers have studied, by 2D map analysis, proteins that were overexpressed in colorectal and fibrosarcoma cell lines made resistant towards mitoxantrone, a drug known to induce atypical MDR [51]. Two proteins, adenine phosphoribosyl transferase and breast cancer-specific gene 1 (BCSG1), were found overexpressed in the resistant colorectal tumor cell line; by the same token, another two proteins, Rho-GDP dissociation inhibitor and another unknown one (having sequence homologies with yeast protein yer-7), were overexpressed in the chemoresistant fibrosarcoma cell line. The putative role of these proteins is discussed in [51].

Chemoresistance in malignant melanoma cells

In the two previous papers from Sinha's group, the proteomic pattern displayed was from 2D gels in an IPG pH 4.0–8.0 interval, stained with a medium-sensitivity colloidal Coomassie blue. In the following papers, the picture changes considerably: the proteomic pattern is developed by exploiting two IPG gels, one in pH 2.8–5.0, the other in pH 4.0–8.0 intervals. Additionally, these gels are stained not only with micellar Coomassie, but also with silver stains. By combining the information from the two pH intervals, well over 1000 polypeptide chains are visualized (as against, probably, no more than 500 with conventional stains). As a result, quite a few dozen spots are found to be up- or downregulated or switched on/off (contrary to the meagre one to two spots in the previous papers [50,51]). Nevertheless, the authors here have chosen to focus only on those spots exhibiting an increase or decrease by at least a factor of four (normally, in differential PDQuest analysis, colour intensity changes of a factor of two, i.e., 100% change in colour uptake, are considered significant and accepted for further analysis).

Table 8. Mass spectrometry-identified proteins in the A431 after cisplatin exposure versus A431/Pt after cisplatin exposure.

Protein	Molecular mass (Da) – isoelectric point	Trend in A431 treated vs. A431/Pt treated
Voltage-dependent anion-selective channel protein	30,641 – 8.8	Fourfold increased in A431 treated
Src substrate cortactin (amplixin, oncogene EMS1)	61,636 – 5.3	On in A431 treated
Heat shock protein 90-β	83,133 – 5.0	On in A431 treated
Translational endoplasmic reticulum ATPase	89,322 – 5.2	On in A431 treated
Tubulin α-2 chain	49,895 – 5.0	On in A431 treated
Glucose-6-phosphate dehydrogenase	56,651 – 6.7	On in A431 treated
Nucleoside diphosphate kinase A	17,149 – 6.1	Fivefold increased in A431
Heat shock cognate 71 kDa protein	70,898 – 5.4	Threefold increased in A431 treated
ATP synthase β chain	56,560 – 5.3	On in A431 treated
T-complex protein 1 and chromatin assembly factor 1	57,488 – 6.4 and 47,656 – 4.8	On in A431 treated
Phosphatidylethanolamine-binding protein	20,926 – 7.6	On in A431 treated
Ubiquitin carboxyl-terminal hydrolase	26,183 – 4.9	On in A431 treated
Lactoglutathione lyase	20,575 – 5.2	On in A431 treated
Phosphoglycerate mutase 1	28,673 – 7.2	Threefold decreased in A431 treated
Glutathione S-transferase	23,225 – 5.5	Twofold increased in A431 treated
Proteasome subunit α Type 2	25,767 – 7.5	Threefold decreased in A431 treated
Stathmin	17,171 – 6.0	Twofold decreased in A431 treated
Peroxiredoxin 1	22,110 – 8.5	On in A431 treated
Profilin I	14,914 – 8.7	On in A431 treated
Proteasome subunit α Type 5	26,469 – 4.8	On in A431 treated
Ubiquitin-conjugating enzyme E2 N	17,138 – 6.5	Threefold decreased in A431 treated
Nucleoside diphosphate kinase A and ubiquitin-conjugating enzyme E2 N	17,149 – 6.1 and 17,138 – 6.5	Threefold decreased in A431 treated

Each protein in the table represents a processed spot of A431 vs. A431/Pt, both after 1 h cisplatin exposure, and is identified by protein name, isoelectric point and molecular mass attributes and trend in the comparison under object (modified from [57]).

Sinha and coworkers have selected a panel of sublines of human melanoma cells (MeWo), rendered separately resistant to the following cytotoxic drugs: vindesine, cisplatin, fotemustine and etoposide [52]. By differential 2D map analysis, this group could demonstrate that four proteins (all in the pH 2.8–5.0 range), namely the translationally controlled tumor protein, the human elongation factor 1-δ, the tetratricopeptide repeat protein and the isoform 14-3-3γ of the 14-3-3 family, were overexpressed (by the restricted cut-off criterion of > fourfold) in the chemoresistant melanoma lines. This same system was revisited by Sinha and coworkers more recently, with yet another profound change in spot pattern [53]. This time, the alkaline part of the pH interval was explored by abandoning the IPG pH 2.8–5.0 range, still keeping the mid-pH 4.0–8.0 range and adding to it another, alkaline IPG interval, pH 8.0–11.0. This group also report detection of some 1500 spots

in the IPG pH 4.0–8.0 range, and an additional 500 spots in the IPG pH 8.0–11.0 interval. The score now changes considerably: in the neutral-to-weakly acidic milieu (pH 4.0–8.0) a total number of 14 proteins (TABLE 1) showed alteration in expression, whereas 20 proteins (TABLE 2 lists only the 11 spots identified) were differentially expressed in the basic milieu (pH 8.0–11.0). The data provided in TABLES 1 & 2 were confirmed in an additional report by the same group [54].

Thermoresistance in stomach cancer cell lines

In an attempt to establish potential interactions between chemo- and thermal-resistance and at identifying common pathways between the two mechanisms of resistance, Poland and coworkers have grown stomach cancer cell lines at progressively elevated temperatures (0.5°C increments) until a subline able to grow at 39.4°C was established [54]. 2D map analysis,

spot excision and MS determination identified 19 proteins that were upregulated and 15 that were downregulated in the thermoresistant cell lines, as listed in TABLE 3. It can be appreciated that quite a few, particularly the HSPs, are in common with those polypeptide chains differentially expressed in chemoresistant cell lines, indicating some common general pathways of regulation.

Chemoresistance to doxorubicin in breast cancer cells

The approach used by Brown and Fenselau to investigate chemoresistance in breast cancer cells is quite different from those reported above [55]: the cytosolic extracts from lysed cells were subjected to reversed phase high-performance liquid chromatography on a C-4 column [56]. The collected fractions were then digested with trypsin and differentially labeled with either ^{18}O (resistant cells) or with ^{16}O (control, susceptible cells). The peptides obtained from the two cell populations were mixed and analyzed by quadrupole TOF-MS, the ratio of the areas of the $^{18}\text{O}/^{16}\text{O}$ peaks giving the expression ratio of a given protein in the two cell populations. By this method, these authors identified 15 proteins that were upregulated in the doxorubicin-resistant cell line, as listed in TABLE 4 (here a change greater than twofold was considered significant). Their potential role on the insurgence of chemoresistance was discussed and evaluated [55].

Chemoresistance to cisplatin in cervix squamous cell carcinomas

Castagna and coworkers have recently published a vast study on chemoresistance, by taking quite a unique approach of unprecedented complexity to the problem [57]. The cervix squamous cell carcinoma cell line A431 and its cisplatin-resistant subline were used as a model system. However, the

experimental set-up involved not just a two-way comparison of the control versus the drug-resistant cell line, as performed in all other papers reviewed in this article, but also an acute cisplatin treatment of both cell lines, so as to obtain an insight into the proteome changes related to mechanisms for immediate survival. This led to a four-way comparison, as follows:

- A431 versus A431/Pt cells
- A431 versus A431 cisplatin-exposed cells
- A431/Pt versus A431/Pt cisplatin-exposed cells
- A431 cisplatin-exposed cells versus A431/Pt cisplatin-exposed cells

The complex proteomic changes detected are summarized in TABLES 5–8 (note in TABLE 7 that, although only three proteins could be characterized, another ten polypeptide spots, strongly modulated, could not be identified via current databases, although they produced good spectra). These authors found modulation of proteins that could be classified under various categories such as [57]:

- Molecular chaperones (e.g., HSP60, HSC71 and HSP90)
- Ca^{2+} -binding proteins (e.g., calmodulin and calumenin)
- Proteins involved in drug detoxification (e.g., peroxiredoxins 2 and 6 and glutathione-S-transferase [GST])
- Antiapoptotic proteins (e.g., 14-3-3 switched on in cisplatin-exposed cells)
- Ion channels (e.g., voltage-dependent anion-selective channel [VDAC]-1)
- Cytoskeletal proteins and their regulators, for example, stathmin and EB1)

Table 9. Proteins detected in more than one comparison.

Protein	Molecular mass (Da) – isoelectric point	Ref.
HSP60	57.6 – 5.2	[53,57]
HSP27	28.3 – 6.0	[53,I,III*,55]
HSP90	89.6 – 4.9	[53,54,57,VIII]
Heat shock cognate 71 kDa and HSX70	70.9 – 5.4	[53,57,VI]
Voltage dependent anion-selective channel protein	30 – 8.4	[53,57,VII,VIII]
14-3-3 γ , - ϵ , - η , - σ , - ζ/δ	29 – 4.7, 4.9	[50,52,53,III,57,VI]
Stathmin	17.3 – 5.87	[55,57,V,VI,VIII]
Proteasome subunits α (types 2, 5), β	25.7, 22.9 – 7.5, 6.1	[53,57,VIII]
Peroxiredoxin 1, 2 and 6	22.1, 24.9 – 6.1, 8.7	[53,57,VI,VII]
Annexin I and V	35.8 – 5.0	[54,57,VI]

*The roman numbers in parentheses refer to the table in which the various proteins are listed.
HSP: Heat shock protein.

In particular, the basal levels of HSC71 and HSP60 were increased in A431/Pt cells compared with A431 cells, and cisplatin exposure resulted in upregulation of HSP60 and HSP90 only in A431 cells. Furthermore, cisplatin exposure upregulated the antiapoptotic 14-3-3 protein in both cell lines, GST in sensitive cells and PRX6 in A431/Pt cells. These findings are consistent with a constitutive expression of defence factors by resistant cells and with activation by cisplatin of mechanisms acting to protect cells from drug-induced damage. Conversely, in a differential proteomic study of human gastric cancer cell lines resistant to cisplatin, Yoo and coworkers found only one protein modulated, namely pyruvate kinase M2, whose enzyme activity and protein level was strongly downregulated in chemoresistant cells (however, by a thorough inspection of their paper, one gets the impression that their approach was not heuristic, but deterministic) [58].

Expert opinion

What emerges from these data is that the preproteomic era view of 'one protein, one disease' (as reviewed in [31–32,36]) cannot hold any longer. It appears more realistic to think that exposure of cells to drugs results in the activation of different mechanisms of resistance. The current concept of drug resistance is based on the hypothesis that a specific drug-resistant phenotype consists of several molecular mechanisms that are simultaneously active, as demonstrated by the proteomic papers reviewed in this article and amply discussed therein [51–55,57]. That being the case, it is obvious that, in a 2D map display, the proteins that are modulated as a result (up- or downregulated, switched on/off) in the drug-resistant cells must be counted by the dozens, as in fact seen in all proteomic reports screened. The families of proteins involved in chemoresistance can be classified into at least the following groups:

- Molecular chaperones
- Cytoskeletal proteins and their regulators
- Proteins involved in drug detoxification, for example, the thioredoxin system
- Proteins in the ubiquitin–proteasome pathway
- Ion channels and ion-binding proteins
- Apoptosis inhibitors

As a rule of thumb, it would appear from the data presented that the total proteins modulated in drug-resistant phenotypes could be as high as 1–2% of the total protein asset of a cell. If it is true that a single cell could express well above 12,000 polypeptide chains [59] (since this is the number separated and detected via giant 2D gels, it is suspected that this number could very well be three- to ten-times higher, due to the noxious problem of spot overlap and to the difficulty in detecting low-abundance proteins), this would pose a formidable problem to any pharmacologist trying to target any specific protein for reversal of drug resistance [60]. In order to

simplify this search, the authors propose a new supertable, listing only those proteins that have been found modulated by different research groups in different types of cells (TABLE 9). Perhaps a few of them will turn out to be the likely target for a new pharmacologic approach. With one proviso, though: it is still not known whether these proteins represent early or late events that would appear only after the phenotype of the drug-resistant cell line has been fully established. Perhaps the key early events are still hidden from us, most likely due to the fact that they involve low-abundance nuclear proteins. The next generation of proteomic studies will have to rely heavily on prefractionation techniques [29], so as to bring to the surface those 'hidden proteins'! Meanwhile, some proteins worth looking at are well evidenced in TABLE 9. Among those, of particular interest appears to be the family of 14-3-3 proteins (just about most of them), a well known family of antiapoptotic proteins. In both chemoresistant cells and cells surviving acute cisplatin treatment, the 14-3-3 proteins are strongly upregulated, as also shown by the levels of mRNA [57]. Also, many members of the family of HSPs are strongly modulated; of particular importance, among them, could be HSP27, since this protein, as reviewed by Sarto and coworkers, shows increased levels in a large number of cancers, associated with different expression of phosphorylated isoforms [61]. This protein has also been highlighted as an important factor for chemoresistance in the study of Urbani and coworkers [62]. Interestingly, HSP90 is just about the only protein in common among the studies on the preproteomic and proteomic era, found to be modulated in just about all studies reported here and already highlighted as a potential target for new chemotherapeutic drugs [36]. For additional overviews on proteomic studies on chemoresistance, see also Hutter and Sinha [63], Lage [64] and Verrills and coworkers [65].

Five-year view

It is difficult to predict what will happen during the next 5 years, since proteomic studies related to chemoresistance are just too few and not well correlated. The authors envision that quite a few more groups will enter this field and that, perhaps, co-ordinated exchange of information and collaborative efforts among several research groups will be operative. Additionally, biochemists, those that are mostly involved in these proteomic studies, will have to join forces and expertise with pharmacologists, both in university research laboratories and with the pharmaceutical industry, if some specific proteins and pathways are to be targeted with novel, more powerful and less toxic drugs able to reverse or strongly quench the phenomenon of drug resistance, as well as thermoresistance [66], in tumors. From this point of view, the new field of pharmacoproteomics (i.e., the study of chemoresistance from the point of view of the pharmacologist) might be particularly helpful [67]. Furthermore, it is quite probable that, here too, we might only be seeing the tip of the iceberg and that the real proteins originating drug resistance might still be missing.

Thus, the next generation of studies will have to focus on nuclear proteins and other subcellular organelles, by exploiting prefractionation techniques able to bring to the limelight the proteins responsible for the early events in the process of drug resistance. To this end, not only will classic 2D map analysis need to be used, but also 2D chromatographic techniques, able to see a window of proteins, such as highly hydrophobic, very small and very large ones, that escape detection with the conventional 2D electrophoretic approach [56]. From this point of view, it is quite surprising that the only study performed by a chromatographic approach screened here appears to be strongly biased towards small-size proteins (TABLE 4) and missing all species above 30 kDa [55], as seen, on the contrary, in all other studies [50–54,57].

The impact of proteomics on the study of chemoresistance, and on drug discovery and clinical trial design, will also have effects on the future of personalized medicine for creating personalized, designed therapies for individual patients or groups of patients with similar diseases based on expression profiling [18]. Personalized therapy is financially desirable, as it will reduce the costs of drug development by shortening the drug development cycle.

Acknowledgements

Partially supported by the Associazione Italiana Ricerca sul Cancro (AIRC) Milan, Italy, by the MIUR (FIRB 2001, Cod. RBNE01KJHT; PRIN, 40%, 2003), Rome, Italy, and by a grant from Cariverona Bank (2004).

Key issues

- 2D (electrophoretic) maps and 2D chromatography are among the technologies that will have the greatest impact on drug resistance studies in cancer therapy.
- Ideally, a well-equipped laboratory should be skilled in both techniques, since they are complementary and they see different portions of the total proteomic asset of a cell.
- Prefractionation techniques (e.g., isolation of pure nuclei) will be essential, since they will permit detection of early events on insurgence of drug resistance, most likely due to low-abundance nuclear proteins. Such a study is yet to be performed.
- Interdisciplinary approaches, by which scientists in the field of proteomics (biochemists) will exchange data and information with pharmacologists, molecular biologists, geneticists and experts in informatics for data mining, are sorely needed.

References

Papers of special note have been highlighted as:

- of interest
 - of considerable interest
- 1 Fidler IJ. The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. *Nature Rev. Cancer* 3, 453–458 (2003).
 - 2 McCormick F. Signalling networks that cause cancer. *Trends Cell. Biol.* 9, M53–M56 (1999).
 - 3 Blume-Jensen P, Hunter T. Oncogenic kinase signalling. *Nature* 411, 355–365 (2001).
 - 4 Wiesner A. Detection of tumour markers with ProteinChip technologies. *Curr. Pharm. Biotechnol.* 5, 45–67 (2004).
 - 5 Herrmann PC, Liotta LA, Petricoin EF. Cancer proteomics: the state of the art. *Dis. Markers* 17, 49–57 (2001).
 - 6 Petricoin EF, Zoon KC, Kohn EC, Barrett JC, Liotta LA. Clinical proteomics: translating benchside promise into bedside reality. *Nature Rev. Drug Discov.* 1, 683–695 (2002).
 - 7 Jain KK. Applications of proteomics in oncology. *Pharmacogenomics* 1, 385–393 (2000).
 - 8 Krieg RC, Paweletz CP, Liotta LA, Petricoin EF. Clinical proteomics for cancer biomarker discovery and therapeutic targeting. *Technol. Cancer Res. Treat.* 1, 263–272 (2002).
 - 9 Petricoin EF, Liotta PA. Clinical applications of proteomics. *J. Nutr.* 133, 2476S–2484S (2003).
 - 10 Petricoin EF, Wulfkuhle J, Espina V, Liotta LA. Clinical proteomics: revolutionizing disease detection and patient tailoring therapy. *J. Proteome Res.* 3, 209–217 (2004).
 - 'An irresistible kaleidoscope of clinical proteomics' *St. Louis Post-Dispatch*.
 - 11 Marko-Varga G, Fehniger TE. Proteomics and disease: the challenges for technology and discovery. *J. Proteome Res.* 3, 167–178 (2004).
 - 12 Weston AD, Hood L. Systems biology, proteomics and the future of healthcare: towards predictive, preventive and personalized medicine. *J. Proteome Res.* 3, 179–196 (2004).
 - 'Nihil obstat quominus imprimatur', recommended even by the Holy See!
 - 13 Chaurand P, Schwartz SA, Caprioli RM. Assessing protein patterns in disease using imaging mass spectrometry. *J. Proteome Res.* 3, 245–252 (2004).
 - 14 Domon B, Broder S. Implications of new proteomics strategies for biology and medicine. *J. Proteome Res.* 3, 253–260 (2004).
 - 15 Qiu J, Madoz-Gurpide J, Misek DE *et al.* Development of natural protein microarrays for diagnosing cancer based on an antibody response to tumor antigens. *J. Proteome Res.* 3, 261–267 (2004).
 - 16 Hamdan H, Righetti PG. *Proteomics Today: Protein Assessment and Biomarkers Using Mass Spectrometry, 2D Electrophoresis and Microarray Technology*. Wiley, Hoboken, USA, 2005.
 - 'Everything you wanted to know about proteomics but were afraid to ask' (W Allen, unpublished).
 - 17 Sanchez JC, Corthals GL, Hochstrasser DF (Eds). *Biomedical Applications of Proteomics*, Wiley-VCH, Weinheim, Germany 2004.
 - 18 Jain KK. Role of oncoproteomics in the personalized management of cancer. *Expert Rev. Proteomics* 1, 49–55 (2004).
 - 19 Martin DB, Nelson PS. From genomics to proteomics: techniques and applications in cancer research. *Trends Cell. Biol.* 11, 560–565 (2001).
 - 20 Reif DM, White BC, Moore JZH. Integrated analysis of genetic, genomic and proteomic data. *Expert Rev. Proteomics* 1, 67–75 (2004).

- 21 Petricoin EF, Ardekani AM, Hitt BA *et al.* Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 359, 572–577 (2002).
- 22 Petricoin EF, Ornstein DK, Paweletz CP *et al.* Serum proteomic patterns for detection of prostate cancer. *J. Natl Cancer Inst.* 94, 1576–1578 (2002).
- 23 Cottingham K. Clinical proteomics: are we there yet? *Anal. Chem.* 75, 472A–476A (2003).
- 24 Diamandis EP. Analysis of serum proteomic patterns for early cancer diagnosis: drawing attention to potential problems. *J. Natl Cancer Inst.* 96, 353–356 (2004).
- 25 Diamandis EP. Proteomic patterns in biological fluids: do they represent the future in cancer diagnostics? *Clin. Chem.* 49, 1272–1275 (2003).
- 26 Anderson LN, Anderson NG. The human plasma proteome: history, character and diagnostic prospects. *Mol. Cell. Proteomics* 1, 845–867 (2002).
- 27 Pieper R, Gatlin CL, Makusky *et al.* The human serum proteome: display of nearly 3700 chromatographically separated protein spots on two-dimensional electrophoresis gels and identification of 325 distinct proteins. *Proteomics* 3, 1345–1364 (2003).
- 28 Thulasiraman V, Lin S, Lathrop J, Lomas L, Hammond D, Boschetti E. Reduction of concentration difference of proteins from biological fluids using combinatorial ligands. *Electrophoresis* 26 (2005) (In Press).
- 29 Righetti PG, Castagna A, Antonioli P, Boschetti E. Pre-fractionation techniques in proteome analysis: the mining tools of the third millennium. *Electrophoresis* 26, 297–319 (2005).
- **A gripping tale of prefractionation science, *The Daily Tar Hill* (Chapel Hill).**
- 30 Righetti PG, Castagna A, Antonucci F *et al.* Proteome analysis in the clinical chemistry laboratory: myth or reality? *Clin. Chim. Acta* (2005) (In Press).
- 31 Gottesman MM, Fojo T, Bates SE. Multi-drug resistance in cancer: role of ATP-dependent transporters. *Nature Rev. Cancer* 2, 48–58 (2002).
- 32 Stein WD, Bates SE, Fojo T. Intractable cancers: the many faces of multi-drug resistance and the many targets it presents for therapeutic attack. *Curr. Drug Targets* 5, 333–346 (2004).
- 33 Hamdan M, Righetti PG. Modern strategies for protein quantification in proteome analysis: advantages and limitations. *Mass Spectrom. Rev.* 21, 287–302 (2002).
- 34 Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnol.* 17, 994–999 (1999).
- 35 Shen DW, Cardarelli C, Hwang J *et al.* Multiple drug-resistant human KB carcinoma cells independently selected for high-level resistance to colchicine, adriamycin, or vinblastine show changes in expression of specific proteins. *J. Biol. Chem.* 261, 7762–7770 (1986).
- 36 Tsuruo T, Naito M, Tomida A *et al.* Molecular targeting therapy of cancer: drug resistance, apoptosis and survival signals. *Cancer Sci.* 94, 15–21 (2003).
- **‘NUNC HIC OUT NUMQUAM...’, or ‘IT IS NOW OR NEVER...’ that you will take a chance at learning about the panoply of macromolecular targets for potential therapy of cancer (Elvis Presley, unpublished).**
- 37 Sugimoto Y, Tsuruo T. DNA-mediated transfer and cloning of a human multi-drug resistant gene of adriamycin-resistant myelogenous leukaemia K562. *Cancer Res.* 47, 2620–2625 (1987).
- 38 Levchenko A, Mehta BM, Niu X *et al.* Intercellular transfer of P-glycoprotein mediates acquired multi-drug resistance in tumor cells. *Proc. Natl Acad. Sci. USA* 102, 1933–1938 (2005).
- 39 Arora A, Seth K, Kalra N *et al.* Modulation of P-glycoprotein-mediated multi-drug resistance in K562 leukemic cells by indole-3-carbinol. *Toxicol. Appl. Pharmacol.* 202, 237–243 (2005).
- 40 Sun J, He ZG, Cheng G *et al.* Multi-drug resistance P-glycoprotein: crucial significance in drug disposition and interaction. *Med. Sci. Monit.* 10, RA5–RA14 (2004).
- 41 Ogiso Y, Tomida A, Kim HD, Tsuruo T. Glucose starvation and hypoxia induce nuclear accumulation of proteasome in cancer cells. *Biochem. Biophys. Res. Commun.* 258, 448–452 (1999).
- 42 Voboril R, Hochwald SN, Li J *et al.* Inhibition of NF-kappa B augments sensitivity to 5-fluorouracil/folinic acid in colon cancer. *J. Surg. Res.* 120(2), 178–188 (2004).
- 43 Julian A. The proteasome: a suitable antineoplastic target. *Nature Rev. Cancer* 4, 349–360 (2004).
- 44 Kaufmann SH, Earnshaw WC. Induction of apoptosis by cancer chemotherapy. *Exp. Cell. Res.* 256, 42–49 (2000).
- 45 Sato S, Fujita N, Tsuruo T. Modulation of Akt kinase activity by binding to Hsp90. *Proc. Natl Acad. Sci USA* 97, 10832–10837 (2000).
- 46 Fujita N, Sato S, Ishida A, Tsuruo T. Involvement of Hsp90 in signalling and stability of PDK1. *J. Biol. Chem.* 277, 10346–10353 (2002).
- 47 Goetz MP, Toft DO, Ames MM, Erlichman C. The Hsp90 chaperone complex as a novel target for cancer therapy. *Ann. Oncol.* 14, 1169–1176 (2003).
- 48 Beliakoff J, Whitesell L. Hsp90: an emerging target for breast cancer therapy. *AntiCancer Drugs* 15, 651–662 (2004).
- 49 Mitsiades N, Mitsiades CS, Poulaki V *et al.* Molecular sequelae of proteasome inhibition in human multiple myeloma cells. *Proc. Natl Acad. Sci. USA* 99, 14374–14379 (2002).
- 50 Sinha P, Hütter G, Köttgen E, Diemel M, Schadendorf H, Lage H. Increased expression of epidermal fatty acid binding protein, cofilin, and 14–3–3-σ (stratifin) detected by two-dimensional gel electrophoresis, mass spectrometry and micro-sequencing of drug-resistant human adenocarcinoma of the pancreas. *Electrophoresis* 20, 2952–2960 (1999).
- 51 Sinha P, Hütter G, Köttgen E, Diemel M, Schadendorf H, Lage H. Search for novel proteins involved in the development of chemoresistance in colorectal cancer and fibrosarcoma cells *in vitro* using two-dimensional electrophoresis, mass spectrometry and microsequencing. *Electrophoresis* 20, 2961–2969 (1999).
- 52 Sinha P, Kohl S, Fischer J *et al.* Identification of novel proteins associated with the development of chemoresistance in malignant melanoma using two-dimensional electrophoresis. *Electrophoresis* 21, 3048–3057 (2000).
- 53 Sinha P, Poland J, Kohl S *et al.* Study of the development of chemoresistance in melanoma cell lines using proteome analysis. *Electrophoresis* 23, 2386–2404 (2003).
- 54 Poland J, Schadendorf D, Lage H, Schnolzer M, Celis JE, Sinha P. Study of therapy resistance in cancer cells with functional proteome analysis. *Clin. Chem. Lab. Med.* 40, 221–234 (2002).
- 55 Brown KJ, Fenselau C. Investigation of doxorubicin resistance in MCF-7 breast cancer cells using shot-gun comparative proteomics with proteolytic ¹⁸O labelling. *J. Proteome Res.* 3, 455–462 (2004).
- 56 Nägele E, Vollmer M, Hörth P, Vad C. 2D-LC/MS techniques for the identification of proteins in highly complex mixtures. *Expert Rev. Proteomics* 1, 37–46 (2004).

- 57 Castagna A, Antonioli P, Astner H *et al.* A proteomic approach to cisplatin resistance in the cervix squamous cell carcinoma cell line A431. *Proteomics* 4, 3246–3267 (2004).
- **The most famous 'J'accuse...!' ever launched against chemoresistance in the scientific literature.** Émile Zola, *L'Aurore* (Paris), January 13, 1898.
- 58 Yoo BC, Ku JL, Hong SH *et al.* Decreased pyruvate kinase M2 activity linked to cisplatin resistance in human gastric carcinoma cell lines. *Int. J. Cancer* 108, 532–539 (2004).
- 59 Klose J, Zeindl E. An attempt to resolve all the various proteins in a single human cell type by two-dimensional electrophoresis. I: extraction of all cell proteins. *Clin. Chem.* 30, 2014–2020 (1984).
- 60 Pietrogrande MC, Marchetti N, Dondi F, Righetti PG. Spot overlapping in two-dimensional polyacrylamide gel electrophoresis separations: a statistical study of complex protein maps. *Electrophoresis* 23, 283–291 (2002).
- 61 Sarto C, Magni F, Valsecchi C, Mocarelli P. Heat shock protein 27 in cancer. In: *Biomedical Applications of Proteomics*. Sanchez JC, Corthals GL, Hochstrasser DF (Eds), Wiley-VCH, Weinheim, Germany, 97–109 (2004).
- 62 Urbani A, Poland J, Bernardini S *et al.* A proteomic investigation into etoposide chemo-resistance of neuroblastoma cell lines. *Proteomics* 5, 796–804 (2005).
- 63 Hutter G, Sinha P. Proteomics for studying cancer cells and the development of chemoresistance. *Proteomics* 1, 1233–1248 (2001).
- 64 Lage H. Proteomics in cancer cell research: an analysis of therapy resistance. *Patbol. Res. Pract.* 200, 105–117 (2004).
- 65 Verrills NM, Walsh B J, Cobon GS, Hains PG, Kavallaris M. Proteome analysis of Vinca alkaloid response and resistance in acute lymphoblastic leukemia reveals novel cytoskeletal alterations. *J. Biol. Chem.* 278, 45082–45093 (2003).
- 66 Poland J, Urbani A, Lage H, Schnolzer M, Sinha P. Study of development of thermoresistance in human pancreatic carcinoma cell lines using proteome analysis. *Electrophoresis* 25, 173–183 (2004).
- 67 Chapal N, Molina L, Molina F, Laplanche M, Pau B, Petit P. Pharmacoproteomic approach to the study of drug mode of action, toxicity, and resistance: applications in diabetes and cancers. *Fundam. Clin. Pharmacol.* 18, 413–422 (2004).

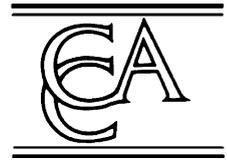
Affiliations

- Pier Giorgio Righetti, PhD
University of Verona, Department of Industrial & Agricultural Biotechnologies, Laboratory of Proteome Science, Strada Le Grazie 15, Verona 37134, Italy
Tel.: +39 045 802 7901
Fax: +39 045 802 7929
righetti@sci.univr.it
- Annalisa Castagna, PhD
University of Verona, Department of Industrial & Agricultural Biotechnologies, Laboratory of Proteome Science, Strada Le Grazie 15, Verona 37134, Italy
Tel.: +39 045 802 7086
Fax: +39 045 802 7929
castagna@sci.univr.it
- Paolo Antonioli, PhD
University of Verona, Department of Industrial & Agricultural Biotechnologies, Laboratory of Proteome Science, Strada Le Grazie 15, Verona 37134, Italy
Tel.: +39 045 802 7087
Fax: +39 045 802 7929
antonioli@sci.univr.it
- Daniela Cecconi, PhD
University of Verona, Department of Industrial & Agricultural Biotechnologies, Laboratory of Proteome Science, Strada Le Grazie 15, Verona 37134, Italy
Tel.: +39 045 802 7086
Fax: +39 045 802 7929
cecconi@sci.univr.it
- Natascia Campostrini, PhD
University of Verona, Department of Industrial & Agricultural Biotechnologies, Laboratory of Proteome Science, Strada Le Grazie 15, Verona 37134, Italy
Tel.: +39 045 802 7087
Fax: +39 045 802 7929
campostrini@sci.univr.it
- Sabina Carla Righetti, PhD
Istituto Nazionale per lo Studio e la Cura dei Tumori, Milano, Italy
Tel.: +39 022 390 2202
Fax: +39 022 390 2692
sabina.righetti@istitutotumori.mi.it



ELSEVIER

Clinica Chimica Acta 357 (2005) 123–139



www.elsevier.com/locate/clinchim

Proteome analysis in the clinical chemistry laboratory: Myth or reality?

Pier Giorgio Righetti^{a,*}, Annalisa Castagna^a, Francesca Antonucci^a, Chiara Piubelli^{b,1},
Daniela Cecconi^a, Natascia Campostrini^a, Chiara Rustichelli^a, Paolo Antonioli^a,
Gianluigi Zanusso^b, Salvatore Monaco^b, Lee Lomas^c, Egisto Boschetti^c

^aDepartment of Agricultural and Industrial Biotechnologies, University of Verona, Strada Le Grazie No. 15, Verona 37134, Italy

^bSection of Clinical Neurology, Department of Neurological and Visual Sciences, University of Verona, Verona, Italy

^cCiphergen Biosystems Inc, Fremont, CA 94555, USA

Received 4 March 2005; accepted 9 March 2005

Abstract

Background: We review here modern aspects of proteomic analysis, as displayed via orthogonal mass/charge analysis (isoelectric focusing in the first dimension, followed by sodium dodecyl sulphate electrophoresis in polyacrylamide gels, SDS-PAGE, at right angles, in the second dimension).

Methods: This technique is capable of displaying a few thousand polypeptide chains, characterized by a single *pI* and *M_r* value as coordinates, and recognized via elution, digestion and mass spectrometry analysis. Although, up to the present, this technique has been used mostly for advanced research, with no immediate applications in the clinical chemistry laboratory, there are hints that such applications will soon become a reality.

Results and conclusions: In the field of cancer research, it is here shown that stathmin (Op18) becomes heavily phosphorylated in cancerous mantle cell lymphomas and that the progression of the disease can be followed by the progression of phosphorylation of Op18 and by the appearance of additional phosphorylated spots. Also chemoresistance of different tumors has been evaluated via 2D-PAGE through quantitative, differential proteomics: among up- and down-regulated proteins in a human cervix squamous cell carcinoma cell line (A413), rendered resistant to cisplatin, one particular protein was found to appear in large quantities by de novo synthesis: 14–3–3, a protein known to impart resistance to apoptosis to cells.

In the field of brain disorders, we could set up an easy test for detecting pathological prions in sporadic Creutzfeldt–Jakob disease (sCJD), by simply searching for those pathological forms in the olfactory mucosa (up to this finding, diagnosis could only be confirmed post-mortem). We are currently working on a test for differentiating sCJD from all the other degenerative dementias. Upon 2D mapping of cerebrospinal fluid (CSF) and immunoblot analysis, we could identify a major spot (*pI* 4.8, *M_r* 30 kDa) followed by some two–three minor spots (*pI*s 5.0–6.0, same *M_r* value) of the same 14–3–3 anti-apoptotic protein

* Corresponding author. Tel.: +39 045 8027901; fax: +39 045 8027929.

E-mail address: righetti@sci.univr.it (P.G. Righetti).

¹ Current address: GlaxoSmithKline, Via Fleming 4, 37135 Verona, Italy.

involved in chemoresistance. By this test, sCJD could be differentiated from all the other degenerative dementias, which are 14–3–3 negative (in sCJD, the rapid and massive brain cell damage releases large quantities of 14–3–3 in the cerebrospinal fluid). Another protein that appears very promising as a marker for sCJD is cystatin C, that is strongly up-regulated in this pathology.

Human sera should also be mined for discovery of many more markers for disease. Up to the present, no one could be found, but this was due to the presence of several major proteins, obscuring all rare ones. Via several immuno-subtraction steps, followed by ion exchange and size exclusion chromatography, one can now detect proteins and peptides present in sera at levels below 10 ng/mL, highlighting the road to discovery of novel markers of disease. Another technique that could revolutionize biomarker discovery in biological fluids consists in the use of combinatorial beads to reduce the dynamic range. They consist in a library of combinatorial ligands coupled to small beads. Such a library comprises hexameric ligands composed of amino acids, resulting in millions different structures. When these beads are impregnated with complex proteomes (e.g., human sera, CSF, urines) of widely differing protein compositions, they are able to significantly reduce the concentration differences, thus greatly enhancing the possibility of evidencing low-abundance species.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Proteome analysis; Stathmin; Mass spectrometry analysis

1. Introduction

Proteomics has recently received strong attention for gaining a wider and comprehensive understanding of the relevance of genomic information generated via numerous genomic sequencing initiatives [1,2]. However, correlating proteomic with genomic information is not an easy task: not only expression levels of proteins vary widely within a cell, but also their correlation with mRNA levels is rather poor [3]. In addition, especially in higher organisms, post-translational modifications [4,5] and differential splicing [6] can significantly blur the information gained at the DNA level. As an additional burden, despite the complete sequencing of several genomes, the functions of most of the putatively expressed proteins remains to be established, although computational methods are being developed for assigning predicted functions to proteins, based on genomic information [7].

The drive to analyze proteins at a genomic scale has led to the development of a number of parallel approaches. Among them, a well-established technique is two-dimensional polyacrylamide gel electrophoresis (2D-PAGE). 2D-PAGE is an open assay approach, which permits the separation and detection of proteins from a wide variety of sources without the need of any prior knowledge of function [8]. The technique has been refined, over the years, by the introduction of immobilized pH gradients (IPG), which provide increased reproducibility and resolu-

tion [9] and by the development of more efficient solubilizing cocktails [10]. As a result of such advances, 2D-PAGE is capable of profiling many thousands of proteins on a single matrix with exquisite resolution, separating discrete isoforms differing by post-translational modifications as subtle as a single deamidation event [11]. Other post-translational modifications detectable by 2D-PAGE include phosphorylation [12] and glycosylation [13], to name just a few. In conjunction with major advances in image analysis and improved statistical tools, it is possible to define those proteins whose expression is altered under defined conditions [14]. The other major event in protein analysis via 2D mapping has been its coupling to mass spectrometry (MS). MS has become the method of choice for protein identification and characterization following separation by 2D-PAGE [15]. This is achieved by either protein mass fingerprinting using time-of-flight mass spectrometry (MS) [16,17] or de novo sequencing using electrospray MS [18,19].

In recent years, alternative approaches to the classical 2D-PAGE methodology have come to the limelight. For example, Washburn et al. [20], Washburn and Yates [21], Liu et al. [22] and Davis et al. [23] demonstrated the use of shotgun identification of protein mixtures by proteolytic digestion followed by multidimensional liquid chromatography. The technique, dubbed MudPIT (multidimensional protein identification technology, a most unfortunate acronym, since, according to the Webster's New Word

Dictionary, it conveys the following meaning: Mud: “wet, soft, sticky earth”, or even, “defamatory remarks, libel or slander”; Pit: “a hole or cavity in the ground”, “an abyss”, “hell”, “any concealed danger, trap, snare”) consists in strong cation exchange chromatography coupled with reversed phase chromatography and thus, although pompously called “multidimensional”, is just as two-dimensional as 2D-PAGE. There is a fundamental distinction, though, between the two approaches: whereas 2D-PAGE gives a display, in the charge/mass plane, of intact polypeptide chains, just as present in a cell or tissue, MudPIT exhibits their tryptic digests, i.e., the constituent peptides in which an intact protein is broken down prior to analysis. With benefits and disadvantages claimed on both sides: whereas MudPIT allows analysis also of hydrophobic and very large proteins, 2D-PAGE, in turns, permits easy identification of all post-translational modifications, by which a single gene product gives rise to a number of different proteins with different functions. There exist also hybrid techniques, by which isoelectric focusing is integrated with reversed phase liquid chromatography [24,25]. Currently, proteomic research embraces two complementary strategies. The first one, known as cell-mapping proteomics, aims at defining protein–protein interactions for building a picture of the complex networks that constitute intracellular signalling pathways. The second strategy, protein expression proteomics, monitors global expression of large numbers of proteins within a cell type or tissue, and quantitatively identifies how patterns of expression change under different circumstances. The approaches here described are valid only in the second case, since, for cell-mapping proteomics, one can only deal with native proteins in order to observe interacting events, which can involve dozens of proteins at a time and require full integrity of tridimensional structure. Another technology that deserves to be mentioned, associating selectively adsorbing surfaces and mass spectrometry, is the so-called SELDI (surface enhanced laser desorption ionization) with more than 350 publications to date (for details see specific section below).

Although the field of proteomics embraces essentially all fields of biology and medicine [26,27], we will restrict this review to some interesting examples of practical relevance in laboratory medicine.

2. Proteomics and cancer

Cancer research has greatly benefited from the proteomic approach, as highlighted in a number of reviews [28–32]. It has become apparent, in fact, that, after the completion of the human genome project, the practice of medicine had to be extended not only to the identification of genes involved in the appearance, progression and treatment of a disease (genomics), as done up to the present time, but also to the question of what these specific genes do and how they interact in communication networks (functional genomics) as well as to the roles played by their protein products in molecular pathways (proteomics). Since the literature in this field is literally burgeoning, we will limit this treatise to some interesting examples from ongoing research in our laboratory. We have recently engaged in quantitative, differential protein profiling in mantle cell lymphoma (MCL), a class of non-Hodgkin lymphomas [33,34]. During the course of these studies, we noticed that normal reactive lymph nodes (as shown by histology), upon 2D-PAGE and immunoblot analysis of stathmin, exhibited only one strong spot of this protein, followed by two very faint, more acidic spots, attributed to phosphorylated derivatives (Fig. 1, upper and lower panels to the left). Comparative analysis of 2D maps of two different mantle cell lymphoma (MCL) biopsies and reactive (healthy) tissues pool lead us to establish that stathmin (Op18, Op—oncoprotein) is overexpressed in MCL samples. As tumor progressed (25% neoplastic cells), we observed that the phosphorylated spots became more intense (Fig. 1, upper and lower central panels). When the tumor had extended to almost all the lymph node (90% neoplastic cells, as evidenced by histology), not only the train of phosphorylated stathmin isoforms had become more intense, but also an additional phosphorylated spot was evidenced (Fig. 1, upper and lower panels to the right). Thus, for the first time, we could have a diagnostic tool that would correlate the extent of tumor invasion with stathmin overexpression and progression of phosphorylation. Op18 is a phosphoprotein known to be a relay protein integrating various intracellular pathways regulating the proliferation, differentiation and other cellular functions [35,36]. Recent data, additionally, indicate a role of stathmin as a potential molecular target for new anti-neoplastic drugs. A potentially predictive

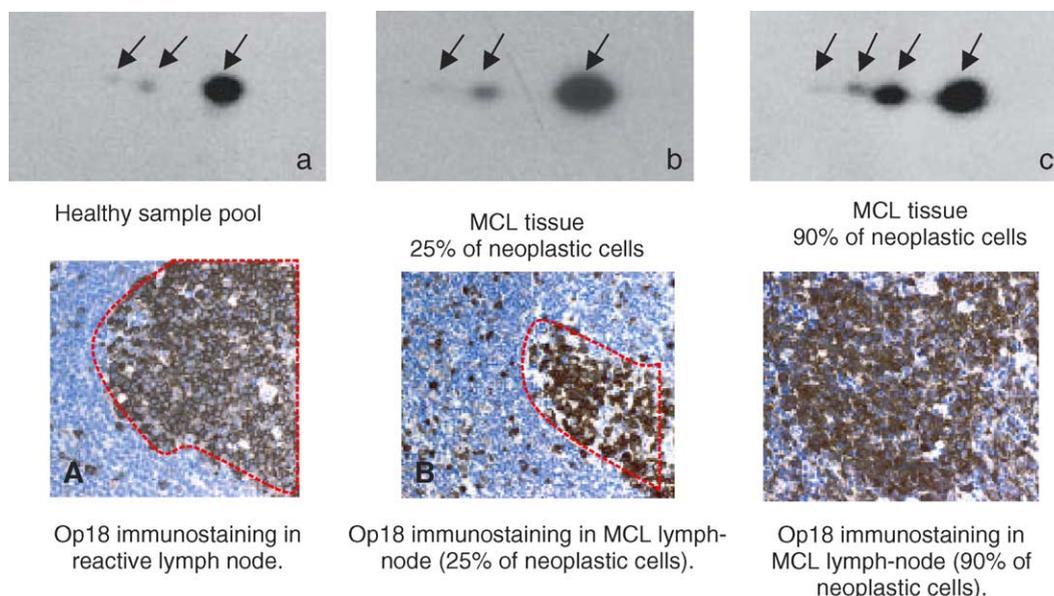


Fig. 1. Three upper panels—stathmin/Op18 phosphorylation pattern. Proteins from total tissue lysates were separated on linear pH 4–7 IPG strips, followed by 8–18% SDS–PAGE analysis. Combined with 2D electrophoresis, Western blot analysis was used for detecting Op18 isoforms in reactive lymph node (a) and MCL lymphoma samples (b and c). Lower panels—Op18 immunostaining. Left panel—reactive lymph node. Only proliferating centrofollicular cells are positive (area inside the red dotted line). Central panel (25% of neoplastic cells)—proliferating centrofollicular cells are positive (area inside red dotted line) and also the neoplastic mantle zone shows scattered positivity. Right panel (90% of neoplastic cells)—not only the proliferating centrofollicular cells are positive (area inside red dotted line), but also the neoplastic mantle zone shows massive positivity all throughout the stained tissue section (modified from Ref. [34], by permission).

role of stathmin has been suggested, since its over-expression can decrease the sensitivity of neoplastic cells to anti-microtubule drugs [37].

3. Chemoresistance in tumor treatment

Chemotherapy of cancer may fail for various reasons. Among these, drug resistance is the most important one. During treatment, resistance towards therapy, especially towards anticancer drugs, develops in ca. 1/3 of the cases. Such resistance may be primary (intrinsic, i.e., the tumor cells do not respond from the start), or it may be secondary (acquired; i.e., the tumor initially responds to therapy, but eventually tumor growth resumes and the patient relapses). It is likely that the biochemical mechanisms involved in primary and secondary resistance are largely similar, but we know far more the second one, since it is easier to study in the laboratory, due to the fact that resistance can be induced in cultured cells and one can then apply

differential proteomic techniques to investigate up- and down-regulated protein expression in the resistant cell line. Acquired drug resistance in cultured cells is nearly always due to a genetic change in these cells, i.e., resistant cells are mutant cells. Mechanisms that confer drug resistance in tumoral cells include increased drug transport, modulation of apoptotic pathways and changes in metabolic mechanisms including detoxification. We have recently started a program aimed at unravelling the mechanism of drug resistance in a human cervix squamous cell carcinoma cell line (A431), rendered resistant to cisplatin upon long-term exposure to low levels of this drug (the antitumor effect of cisplatin issues from the drug-mediated formation of DNA adducts). In addition, control and resistant cells were further treated with acute, massive doses of cisplatin and surviving cells also analyzed by the 2D-PAGE protocol. As a result of all comparisons among the variously-treated and control cells, a number of protein spots (about 40) were found to have experienced up- and down-regulation.

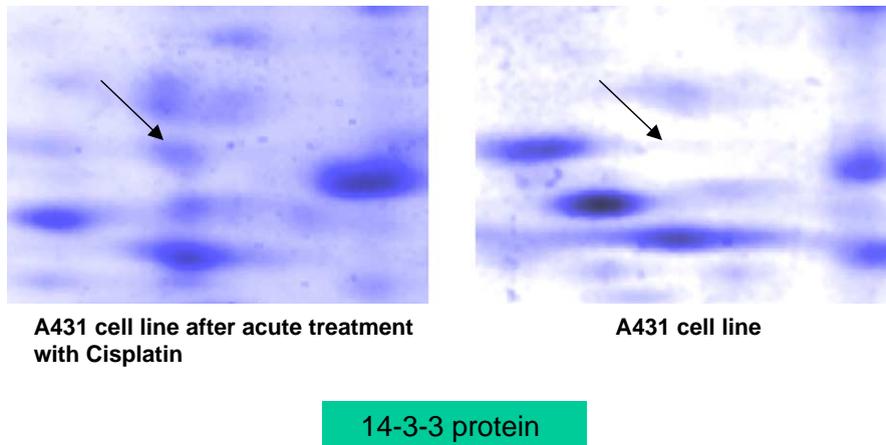


Fig. 2. Two-dimensional PAGE of total lysates of control A431 cervical cancer cells (right panel) vs. the same cells surviving after an acute, massive-dose treatment with cisplatin. First dimension: IPG pH 3–10; second dimensions: 8–18%T polyacrylamide gel slab impregnated with SDS. Stain: micellar Coomassie Brilliant Blue. The two panels represent a small (about 2 cm²) section of the 2D map centred on the 14–3–3 spot (marked by an arrow in the two panels), recognized via peptide mass fingerprinting (modified from Ref. [38], by permission).

Although we hypothesize that a specific drug-resistant phenotype should consist of several molecular mechanisms that are simultaneously active, at least one major difference was immediately apparent, as shown in Fig. 2: the cells that had survived acute treatment with massive doses of cisplatin had synthesized, *de novo*, a protein spot, found by mass spectrometry to be the protein called “14–3–3”, a member of a family of at least seven distinct forms, known to be an anti-apoptotic protein, thus able to prevent, or strongly slow down, cell death [38].

4. Proteomics of brain disorders

Our group has been among the first ones to introduce 2D mapping, coupled with Western blots, in the study of cerebral disorders [39,40]. The results were quite unique. The prion protein, which typically, in mono-dimensional SDS–PAGE, was resolved into 3 major bands (the diglycosylated form, M_r ca. 34 kDa, the mono-glycosylated, M_r ca. 30 kDa and the unglycosylated species, M_r ca. 27 kDa), accompanied by some minor truncated forms, (M_r in the range 18–22 kDa), in 2D mapping gave origin to strings of some 60 spots, covering the *pI* 4–8 range, in the same M_r interval, indicating a most intriguing and variegated glycosylation pattern. In addition, via plain SDS–PAGE, we were able to set

up an easy test for detecting pathological prions in sporadic Creutzfeldt–Jakob disease (sCJD), by simply searching for those pathological forms in the olfactory mucosa (up to this finding, diagnosis could only be confirmed post-mortem) [41]. Although in its infancy in clinical medicine, this novel, simple and most efficient test might find an immediate application for large scale screening for prion infection in cattle, which up to the present had to be slaughtered in order to detect the carriers of the “mad cow” disease. Work from our group is reported in Fig. 3, which confirms the usefulness of this test for differentiating sCJD from all the other degenerative dementias. Upon 2D mapping of cerebrospinal fluid (CSF) and immunoblot analysis, one can easily identify a major spot (*pI* 4.8, M_r 30 kDa) followed by some two–three minor spots (*pI*s 5.0–6.0, same M_r value) of the same 14–3–3 anti-apoptotic protein discussed in the previous section. An easy, mono-dimensional test can be performed via a simple SDS–PAGE followed by immunoblot (Fig. 3, bottom panel). By this test, sCJD can be differentiated from all the other degenerative dementias, which are 14–3–3 negative (in sCJD, the rapid and massive brain cell damage releases large quantities of 14–3–3 in the cerebrospinal fluid). In addition, we have observed another potential marker, in CSF, for CJD, again via extensive differential proteomic analysis via 2D maps of control and disease samples: cystatin C, which appears almost as *de novo*

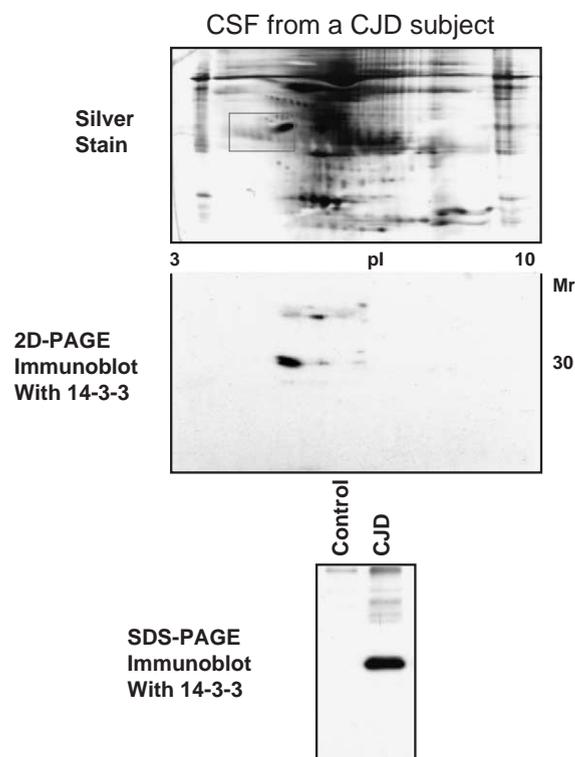


Fig. 3. Analysis of 14–3–3 in cerebrospinal fluid. Upper panel—2D map of CSF from a patient with CJD (electrophoretic conditions as shown in this figure). The 14–3–3 protein is within the area limited by the red square (pI 4.8–6.0 and molecular mass of 30 kDa), as detected via an immunoblot with anti-14–3–3 antibodies (intermediate panel). Bottom panel—mono-dimensional SDS–PAGE followed by an immunoblot with anti-14–3–3 antibodies for differentiating CJD from all the other degenerative dementias, which are 14–3–3 negative (unpublished experiments).

spot in CJD specimens (see Fig. 4A and B; Piubelli et al., manuscript in preparation). Interestingly, when we finished this project and started collecting literature data, we found out that identical conclusions had been reached by Sanchez et al. [42], simply by using SELDI-TOF, via capture of proteins with affinity for copper on the surface of chips containing immobilized copper (IMAC-Cu⁺⁺).

5. Human serum: mining below the tip of the iceberg

It seems only yesterday when Tiselius, in his monumental moving boundary apparatus, was trying

to analyze human sera for early diagnosis of different diseases (the technique was so complicated and the instrument of such a mammoth size that the technique soon fell into oblivion). Yet, he was able to observe the boundaries of 5–6 major components of sera. Considering that today's cellulose acetate electrophoresis (CAE) does not perform any better, that was quite an accomplishment, for the years 1940–1945. When biochemists started analysing sera via 2D mapping, the picture changed drastically: now at least 600 spots could be revealed scattered in the pI/M_r plane, via silver staining, thus originating the joke that the best thing one could do with CAE results would be to engrave them on the marble stone of the patient's tomb, considering the total lack of resolution and sensitivity of the method. Yet, even with the advent of 2D mapping, no new markers of diseases, in addition to those presently available, could be found. The reasons are outlined in a remarkable paper by Pieper et al. [43]: it is nearly impossible to detect any early potential marker of a disease (and there should be plenty of them in sera) as long as whole serum is analyzed, since the most abundant proteins will simply obliterate the signal of the rare ones (the dynamic range of plasma proteins concentrations comprising some ten orders of magnitude!). These authors had to go through the painstaking process of removing the nine most abundant human serum proteins (albumin, IgG, haptoglobin, transferrin, transthyretin, α_1 -antitrypsin, α_1 -acid glycoprotein, hemopexin and α_2 -macroglobulin) via capture with nine different antisera. But this exercise was only for warming up the muscles! The immuno-affinity-subtracted serum proteins were concentrated and subjected to anion-exchange chromatography, in turn followed by size exclusion chromatography. The various fractions were finally analyzed via an impressive number of 2D maps, and the final 3700 collected spots from all possible maps reduced to 1800 distinct serum proteins, as identified by mass spectrometry. They finally collapsed into 325 distinct proteins, after sequence homology and similarity searches had been carried out for eliminating redundant protein annotations. This was truly a large-scale digging below the tip of the iceberg. Yet, much to their dismay, the end results were not so exciting: even after such an extensive pre-cleaning, no new markers of disease could be found. This does not

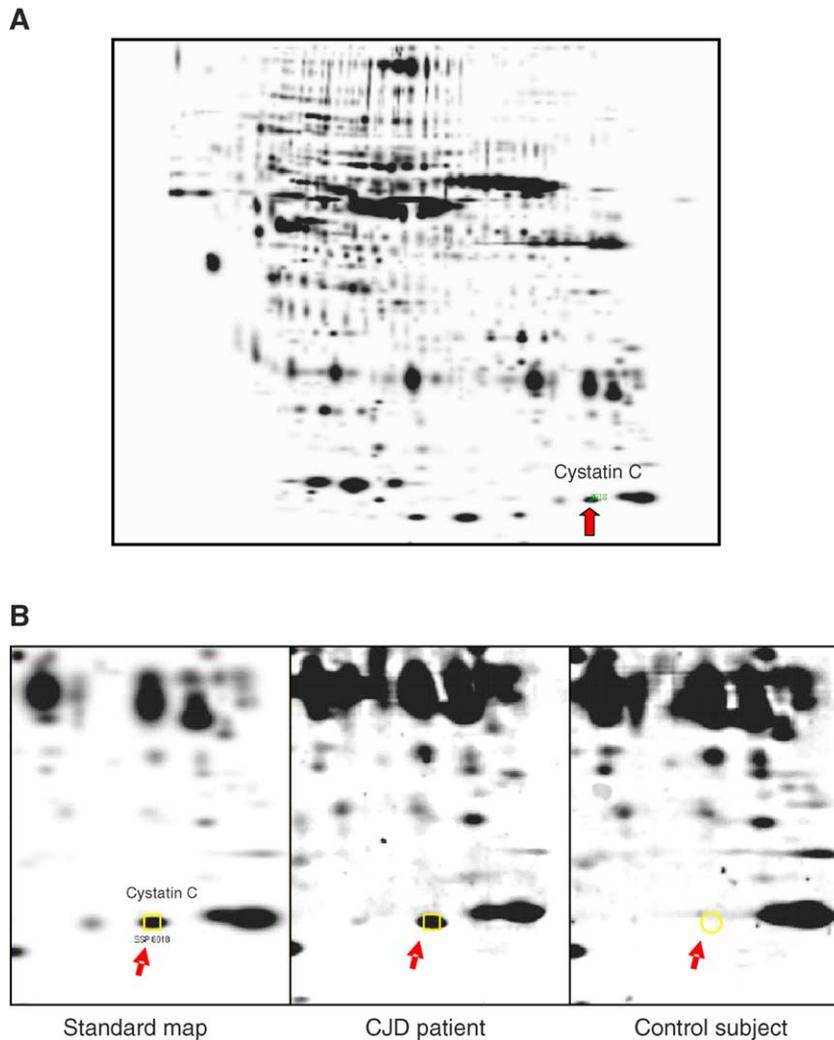


Fig. 4. (A) Reference two-dimensional map of cerebrospinal fluid resulting from the matchset of 6 CJD and 6 control samples. Among a number of other spots differentially modulated, the spot marked with an arrow (found to be cystatin C) was markedly up-regulated, to the point that it was essentially undetectable in normal CSF, as shown in the detailed maps of panel B (Piubelli et al., unpublished experiments).

seem a defeat, though: via this process, these authors could finally visualize, even with a medium-sensitivity stain, several proteins present in sera in <10 ng/mL concentrations, such as interleukin 6, cathepsin and peptide hormones. That pre-fractionation would be the key to success we had also fully understood a few years ago [44], but we could not compete with the efforts and resources of such a large group, aptly known in the market as “Large Scale Biology Corporation” [43]. Actually, as shown in Fig. 5, even their approach might not be enough for

gaining access to the “deep proteome,” in that by this procedure, they remove 90% of the serum proteins. At the recent HUPO International meeting in Beijing (October 24–27, 2004), it has been suggested that one should further treat this depleted serum with an additional set of 12 antisera able to remove the next 12 most abundant proteins: this would subtract an additional 9% serum proteins, finally “exposing to view” the truly hidden proteome, the remaining 1% buried underneath this mound of major species. Yet, this approach might not be immune from some severe

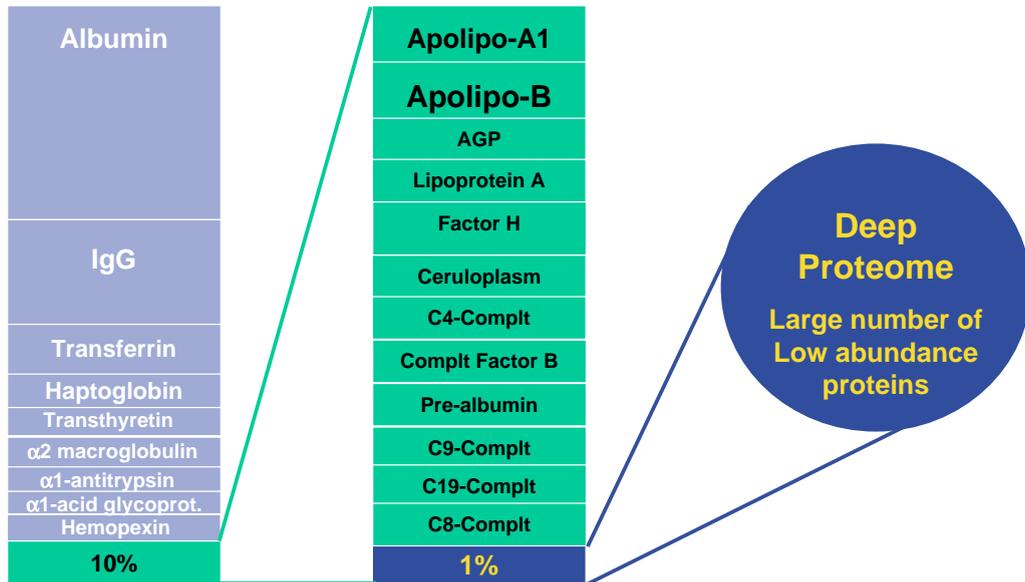


Fig. 5. The blood proteome: problems of dynamic range, disproportionate protein concentration and large number of species. The immuno-subtraction of the first set of nine most abundant proteins removes 90% of serum proteins, whereas removal of the next set of twelve species eliminates another 9%, thus leaving the remaining 1%, called the deep proteome.

drawbacks, as outlined in Fig. 6: the parasitic co-depletion of a number of proteins that might be physiologically bound to the proteins captured by immuno-subtraction or might somehow recognize some epitopes specific of the proteins to be harvested by a given set of antibodies. Perhaps, the approach described below might turn out to be the true big revolution for bringing to the limelight the “hidden proteome” and discovering several new biomarkers of importance in routine clinical chemistry analysis.

6. Reduction of dynamic range with ligand libraries

Additional problems in uncovering the “hidden serum proteome” are outlined in Fig. 7, although the Wenn diagram describing the “visible” region should be an ellipse elongated on the pI axis, rather than a circle, considering that, already in 1990, we described immobilized pH gradients extending from pH 2.5 to 12, thus able to resolve also strongly acidic and basic

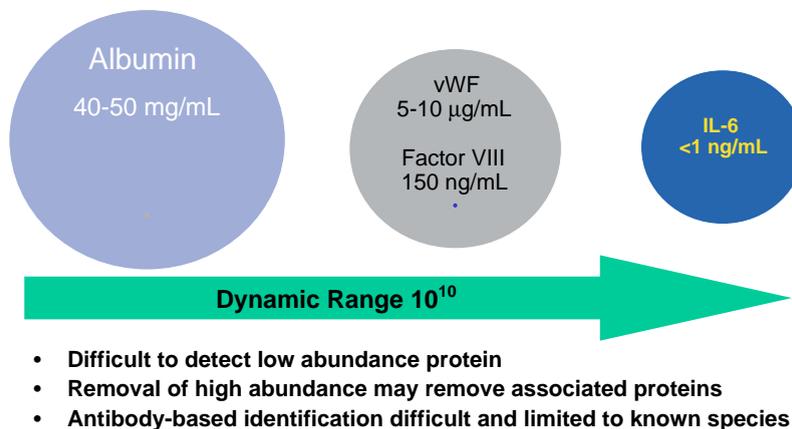


Fig. 6. Example of concentration difference of various proteins in human serum. vWF—von Willebrad factor; IL-6—interleukin 6.

- **Explored proteome may represent 20-30% of total**
- **Very low and high pI not discriminated by 2DE**
- **Hydrophobic species difficult to work out**
- **High Mr not visible by MS and 2DE**
- **Low Mr not visible by 2DE**
- **Large dynamic range not compatible with current methods**

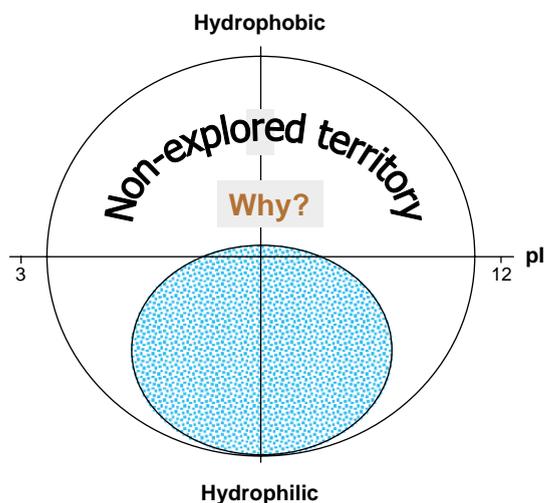


Fig. 7. Wenn graph on the explored vs. non-explored proteome as a function of pI values (horizontal axis) and hydrophobicity scale (vertical axis). The third dimension (not shown) would be an M_r scale.

species [45]. Nevertheless, the problems listed in Fig. 7 are real, and they further complicate the picture, considering also the drawbacks inherent to the immuno-subtraction technique described above. A novel approach, based on totally different principles, ought to be sought. One such a scheme could be, rather than removing the most abundant species, trying to capture equal quantities of each and all the species present in sera (and in general in any biological sample), thus “equalizing” their concentration or better reducing the concentration difference. By this process, one would cut the huge peaks and all crests of the most abundant species, while simultaneously greatly increasing the concentration of the rare ones and thus bringing them to the visibility threshold. This could be achieved by using solid-phase ligand libraries [46–48]. They are constituted of discrete beads, each of them carrying a relative large number of copies of the same ligand, said ligands being different from one bead to another. If the number of different ligands from the library is significantly larger than the supposed protein number in the initial mixture, statistically each protein may find its corresponding partner. The starting point of the technology is thus the synthesis of the ligands on which proteins are captured. Based on the pioneering work of Merrifield [49] on solid-phase synthesis, and using the “split, couple, recombine” method, libraries of potentially billions of different amino acid ligands can be created.

Briefly, the synthesis described few years ago [50,51] is performed in parallel-sequential chemical reactions: the first synthesis step is a batch of millions of microscopic resin beads divided into different reaction vessels and the first building block, e.g., a protected amino acid, is coupled to the resin. The beads are then mixed together and washed extensively, the amino group of the coupled amino acid is de-protected, and the beads are distributed randomly into a second set of reaction vessels and coupled with the next set of building blocks. This process is repeated until a ligand of the desired length is obtained.

Using just the 20 building blocks and making six reaction steps, a library of linear, hexamer ligands containing 20^6 , or 64×10^6 different members can be made. If unnatural amino acids including D-enantiomers are included in each position and the ligands are branched, linear, or circular, for all practical purposes, the potential ligand diversity of the library is unlimited. Thus, due to this tremendous ligand diversity, within the library there is theoretically a ligand for every protein, antibody, peptide, etc., present in the starting material. The first published work in this domain focused on a library of hexamers made by using 20 amino acids. However, the approach is extendable to any other mean to create libraries of discrete ligands. One condition for a properly working principle is to get ligands with dissociation constant values compatible with the respective protein concen-

- **Simultaneous dilution of HAP and concentration of LAP**
- **In a single and reproducible step.**
- **No species depletion is contemplated with this method**
- **“Equalization” exclusively achievable using bead libraries of a very large diversity.**
- **A controlled sample-to-bead ratio allows a given reduction in dynamic range of the proteome**

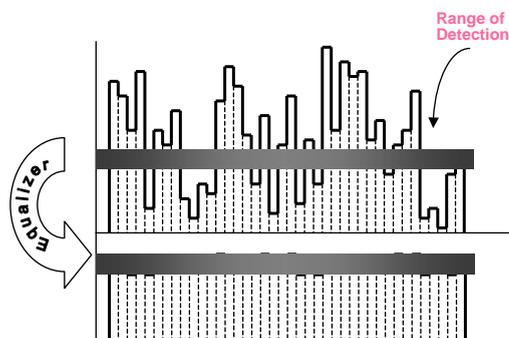


Fig. 8. Schematic representation of benefits resulting from the reduction of dynamic range using combinatorial ligands. (Upper panel) Situation of non-treated sample where a large portion of proteins escape the detection range. (Lower panel) The concentration difference reduced and consequently, significantly more proteins fall within the detection window. Abbreviations: HAP, high-abundance proteins; LAP, low-abundance proteins.

tration to be reversibly captured. Statistically, more than one ligand will be present in the library for a given protein and probably more than one protein may have an affinity for the same ligand.

As schematically represented in Fig. 8, when a complex protein mixture of a first variance, such a human serum, is incubated with this library, representatives of each component of the mixture will bind to these individual ligands. In large overloading conditions high-abundance proteins (HAP) saturate their specific affinity ligand and excess is removed during the washing step, while low-abundance proteins (LAP) continue to concentrate on their specific affinity ligands. After processing, representatives of all the original components can be eluted to produce a sample with a second variance where each represen-

tative is still present but in much different concentrations. Proteins of high abundance are significantly diluted while low-abundance species are concentrated. Just to give a representative image of this process, Fig. 9 shows two maps of human serum, a control one and a second, identical aliquot, sprinkled with a tiny amount of “equalizer beads.” Under identical loading conditions, it can be appreciated that, whereas in the control some 195 spots are visible, in the “equalized” serum sample a total of 487 spots are brought to the limelight. And this is obtained by using, as a starting material, only a small volume of serum! We expect that, when treating larger volumes, from 10 to 100 mL, many more detectable proteins should be visible by both mass spectrometry and electrophoretic techniques, perhaps just about all the missing serum pro-

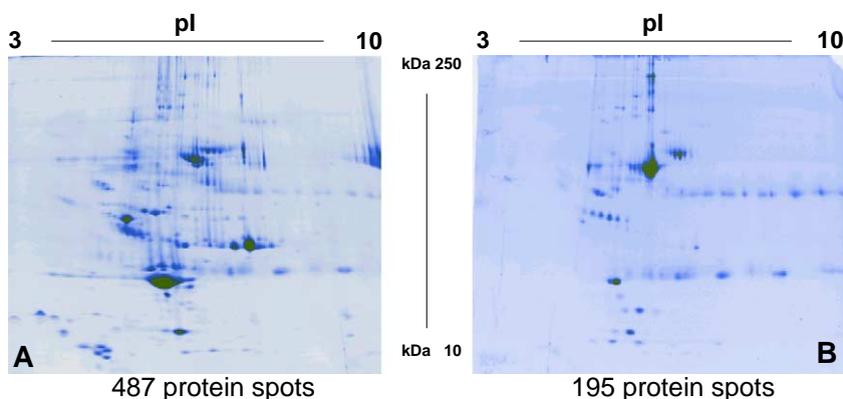


Fig. 9. Two-dimensional maps of control (right) and “equalizer-bead”-treated (left) human sera. Note that, whereas in the first case only 195 spots could be detected, in the second case, a total of 487 spots are visible. First dimension: IPG 3–10, non-linear; second dimension 8–18% T SDS-PAGE (Castagna et al., unpublished).

teome (Castagna, Boschetti and Righetti, manuscript in preparation).

7. Protein arrays: the proteomics of the future?

It would be unfair in this review not to comment on protein arrays, believed to represent the ultimate in massively parallel protein display, much like DNA microarrays are [52]. Not that one should rejoice and sing paeans, by all means. Protein arrays, presently, are more a dream than reality, although strong efforts are spent into this field by companies and scientists all over the world. For expression arrays, produced for detecting expression levels of individual analytes, one solution could be the selection of antibody probes, such as those selected from phage display libraries. In fact, such antibodies provide a combinatorial library approach to biology via amplification of selected clones of immunoglobulin short chain variable regions [53]. Accordingly, phage libraries could offer the potential for antibody production at the scale necessary for genomic-scale analysis of proteomes. As an alternative, one could select antibodies from synthetic peptide epitopes whose sequences had been derived from translated genomic sequences. It is, however, quite unlikely that the first generation of protein arrays will be able to recognize all the proteins expressed from higher eukaryote genomes.

8. The SELDI-MS approach

One of the latest technology platforms that have been developed for proteomics includes chip-based arrays associated with time-of-flight mass spectrometry. Such arrays have been elaborated for separating proteins based on surface chemistries (e.g., ion exchange, hydrophobic associations, metal chelate, mix-mode, and the like) or known protein ligands (e.g., antibodies), with subsequent analysis by mass spectrometry [54–56].

The principle of separating proteins from crude mixtures on chip surfaces is very similar to a classical adsorption–desorption mechanisms used in column chromatography. The difference here is that the sorbent is a flat surface and the desorption–detection is

generated by a laser desorption ionization, concomitant to the use of an energy adsorption molecule in order to have proteins vaporized and ionized to enter a time-of-flight (TOF) mass spectrometry tube. A typical ProteinChip® Array and its use is schematically illustrated in Fig. 10. As per conventional column chromatography, the sample is first loaded on the functionalized surface of the chip in appropriate conditions of pH and ionic strength to capture the proteins of interest. Then a wash follows to eliminate non-retained species while captured proteins are ready for mass spectrometry analysis. A selection of captured proteins is typically accomplished by modifying either the ionic strength or the pH or both prior to mass spectrometry analysis. Contrary to chromatographic separations, where mobile phase elutropic strength is designed to elute all proteins for subsequent down-stream or off-line analysis, here only retained proteins are ultimately studied. During the final stage of sample preparation, a matrix solution is added to elute adsorbed proteins from the chip surface and entrain them in growing crystals. The crystals are irradiated by a focused pulse of laser light, which subsequently causes a phase transition, creating gaseous ions that are analyzed in a TOF MS. In this fashion, a mass spectrum can be generated, often indicating whether the protein of interest is adsorbed while additionally informing the presence of impurities. The lower the number of impurities, the higher the selectivity of the chip surfaces for the target protein in the conditions of exploitation. Extended applications are described in the literature [see [57] for a review]. With the use of selected functionalized surfaces, analyses of protein families are thus performed.

Promising applications of protein chip microarrays have been published on differential profiling [59] and high-throughput functional analysis [58]. Although the SELDI-MS approach is perceived by scientists as an analytical instrument, it should not be forgotten that the series of chips, underlying the capture of selected protein populations, are indeed a serial (or cascade) fractionation devices; i.e., a given protein population adsorbed onto a cation exchanger surface can be eluted and further purified onto an anion exchanger, to be subsequently manipulated onto a metal affinity surface, a reversed phase and so on, until obtaining the desired purity of the captured

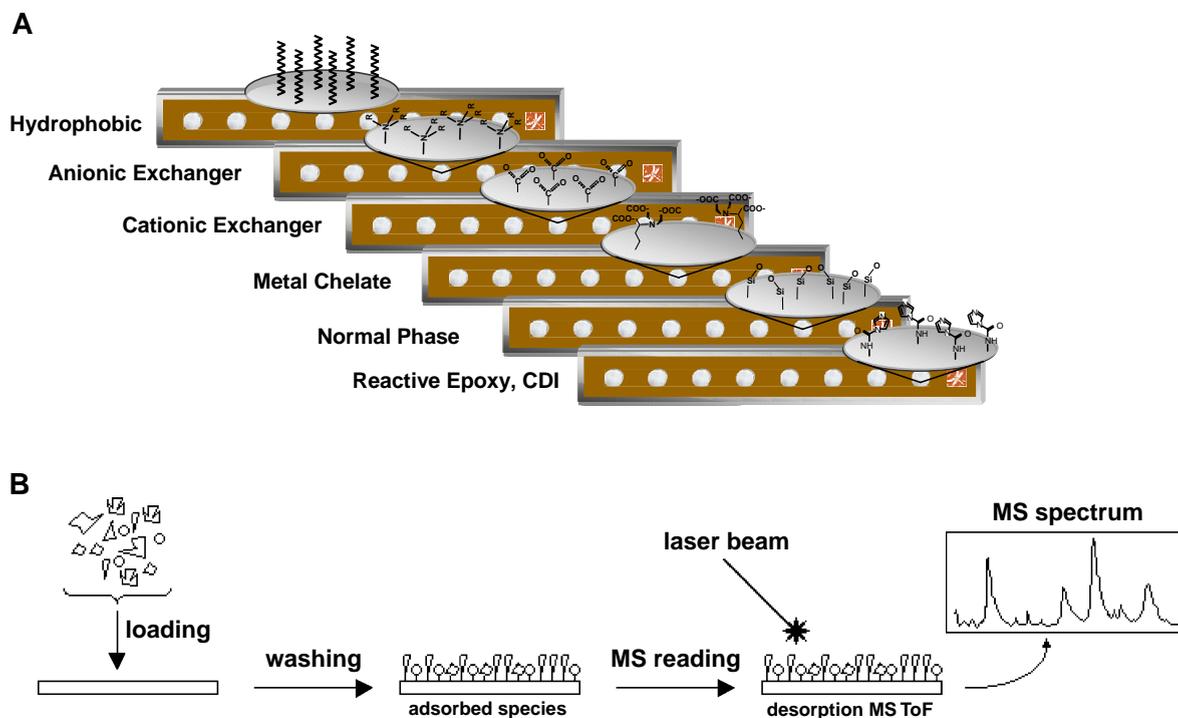


Fig. 10. Schematic representation of ProteinChip® Arrays with examples of chemical functions responsible for protein category capture (A). Panel (B) shows the mechanism of adsorption–desorption and ionization of proteins prior entering a mass spectrometer analyzer.

protein or the final capture of a biomarker (or a panel of biomarkers) of a given disease. Some examples of the results obtained with the SELDI approach can be found in, e.g., [60–68].

9. Concluding remarks

The panorama, in proteome analysis, did not look very bright up to recent times. As reviewed here, several published papers highlighted major limitations of available technologies for proteome investigations. Current approaches were qualified as incapable of having a whole vision of the proteome, even limited to structural aspects. For instance, strongly alkaline proteins seemed to be poorly represented using classical two-dimensional electrophoresis and highly hydrophobic proteins could not be properly solubilized and consequently not analyzed and/or identified. Electrophoresis-based methods taken alone (still the most commonly used to date) are neither appropriate

for polypeptides of masses lower than 5000 Da, nor effective for very high M_r species as well. Only mass spectrometry contributes significantly to the analysis of low sized polypeptides. To this panel, it is to be added that post-translational modifications and especially glycosylations are still part of the non-resolved dilemmas. In this situation, authors estimate that only about 20–30% of expressed proteins are detectable by standard methods to date. Yet, notwithstanding such limitations, the progress made has been outstanding, and patterns for tying up the expression of given proteins to some pathologies begin clearly to emerge. Take Table 1, for instance: even on such a limited survey as the one here presented, it would appear that the panel of proteins up- and down-regulated in the various diseases here listed, although shared by a number of them (as seen when reading the horizontal lines), appear nevertheless to be unique, as a whole, per each individual condition (as seen when reading the vertical columns). Could some of these proteins be targeted for drug treatment of some of these diseases,

Table 1
Panels of proteins associated with various diseases, as discovered by differential proteomic analysis

Disease association											
Protein name	Lung adenocarcinoma [69]	Mantle cell lymphoma [34]	Brain age [70]	Dementia [71,72]	Alcoholic brain [73]	Mammary epithelial differentiation [74]	Breast cancer [75]	Colon cancer [76,77]	Retinal degeneration [78]	Liver inflammation [79]	Aging fibroblasts [80]
Op18 (Stathmin)	✓	✓	✓								
Peroxiredoxin		✓	✓		✓	✓	✓	✓	✓		✓
ApoA1			✓							✓	
14–3–3 isoforms	✓			✓		✓			✓		
HSP70				✓	✓	✓	✓		✓		
GRP78	✓	✓			✓	✓					
Enolase				✓	✓		✓		✓		
Cofilin						✓		✓			✓
Annexin I						✓		✓			✓
Transthyretin	✓			✓						✓	
Protein disulfide isomerase	✓			✓		✓	✓	✓	✓		✓
GST		✓						✓			✓
α-1-antitrypsin		✓					✓			✓	
HSP27		✓				✓					✓
Superoxide dismutase				✓				✓			✓

in case they represent the early events associated with each pathological condition, or do they represent late manifestations of an already established pathology? Although we still do not have answers for that, it is clear that, as more and more data accumulate, a clear pattern will emerge and, for that, current proteomic analysis seems to be just the right tool. The principle of catching a panel of markers specific for a given disease, possibly at an early stage, is precisely the philosophy behind the SELDI-MS approach.

10. The strange ride of Morrowbie Jukes

In Anno Domini 2005, we will celebrate the thirtieth anniversary of two-dimensional mapping, as launched in 1975 by O'Farrell [81]. His system was already highly sophisticated since its inception: in fact, he was able to resolve and detect about 1100 different proteins from lysed *E. coli* cells on a single 2D map and suggested that the maximum resolution capability might have been as high as 5000 different proteins. O'Farrell had elaborated the technique to its utmost refinement, nullifying the attempts of the second and third wave of "discoverers" to get credit for it; even today 2D maps represent the most popular technique in proteome analysis. Already in the early eighties, the Andersons conceived their grandiose plan of the Human Protein Index [82] and elaborated extensively on the potential of 2D map analysis in the field of clinical chemistry. It is enough to peruse two special issues of *Clin. Chem.*, devoted to this topic, to find a wealth of reports on the use of 2D mapping in analysing a number of pathologies [83,84]. The large body of articles in these two issues amply demonstrates that already 20, 25 years ago, 2D map analysis was well under way and was producing impressive results. Such maps, although generated with the old technology of conventional IEF in soluble carrier ampholytes, had an outstanding resolution and generated spots of high quality, essentially devoid of horizontal and vertical streaking.

Yet, for all of its advanced properties and exquisite resolution, in all these years each one of us using 2D map analysis looked quite like the protagonist of the bone-chilling novel of Rudyard Kipling, "The Strange Ride of Morrowbie Jukes" [85]. In a delirium of fever and excitement, we all brandished our hog-spears and

went out under a great white moon in a mad gallop, shouting challenges to the camel-thorn bushes, chasing our pray. But we all seemed to miss our pray and ruin in the trap of the Living Dead. No way out from the horseshoe-shaped crater: the steeply-graded sand wall would prevent any climbing, the only open side, at the mouth of a river, being heavily guarded by troops ready to shoot at anybody attempting escape. At night, even in the absence of guards, quick sands on the river banks would also bar any escape route. There was a cryptic irony underlying the name of the protagonist: Morrowbie seems to stress the condition of someone with no future, and even the term Jukes had been taken from the fictitious name of a New York family struck by poverty and disease. Well, just as rescue came from without to poor Morrowbie, it seems that today, with the panoply of tools available in proteome analysis and the continuous refinements of the technique over the years, we will soon climb out of the deadly trap. Thus, the Morrow-bye-bye might soon be changed into Morrow-now or Morrow-today. A bountiful harvest seems to be hiding around the corner.

Acknowledgements

Supported in part by grants from the Associazione Italiana Ricerca sul Cancro (AIRC, Milano, Italy), by MURST (progetto coordinato 40%, 2003), by FIRB 2001 (grant No. RBNE01KJHT), by the European Community (proposal No. 12793, 2005), from Fondazione Cariverona 2002 "Neurodegenerative disorders" and from the Italian Ministry of Health in collaboration and with the contribution of the Istituto Superiore di Sanità ("Search for infection-related endogenous or exogenous factors in the cerebrospinal fluid of patients with transmissible spongiform encephalopathy" and "Study of pathogenic mechanisms in neurodegenerative disorders for the diagnosis and development of therapeutic approaches").

References

- [1] Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. *Science* 2001;291:1304–51.

- [2] Mewes HW, Albermann K, Bahr M, Frishman D, Gleissner A, Hani J, et al. Overview of the yeast genome. *Nature* 1997; 387:7–65.
- [3] Gygi SP, Rochon Y, Franza BR, Aebersold R. Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* 1999;19:1720–30.
- [4] Kreider BL. Post-translational modifications in proteins. *Annu Rep Med Chem* 2001;36:227–35.
- [5] Miklos GLG, Maleszka R. Protein functions and biological contexts. *Proteomics* 2001;1:169–78.
- [6] Graveley BR. Alternative splicing: increasing diversity in the proteomic world. *Trends Genet* 2001;17:100–7.
- [7] Honig B. Protein folding: from the Levinthal paradox to structure prediction. *J Mol Biol* 1999;293:283–93.
- [8] Righetti PG, Stoyanov AV, Zhukov MY. The proteome revisited: theory and practice of all relevant electrophoretic steps. Amsterdam: Elsevier; 2001. p. 1–395.
- [9] Righetti PG. Immobilized pH gradients: theory and methodology. Amsterdam: Elsevier; 1990. p. 1–397.
- [10] Tastet C, Charmont S, Chevillet M, Luche S, Rabilloud T. Structure-efficiency relationships of zwitterionic detergents as protein solubilizers in two-dimensional electrophoresis. *Proteomics* 2003;3:111–21.
- [11] Sarioglu H, Lottspeich F, Walk T, Jung G, Eckerskom C. Deamidation as a widespread phenomenon in two-dimensional polyacrylamide gel electrophoresis of human blood plasma proteins. *Electrophoresis* 2000;21:2209–18.
- [12] Kaufmann H, Bailey JE, Fussenegger M. Use of antibodies for detection of phosphorylated proteins separated by two-dimensional gel electrophoresis. *Proteomics* 2001;1:194–9.
- [13] Fivaz M, Vilbois F, Pasquali C, van der Goot FG. Analysis of glycosyl phosphatidylinositol-anchored proteins by two-dimensional gel electrophoresis. *Electrophoresis* 2000;21: 3351–6.
- [14] Hamdan H, Righetti PG. Assessment of protein expression by means of 2-D gel electrophoresis with and without mass spectrometry. *Mass Spectrom Rev* 2003;22:272–84.
- [15] Hamdan H, Righetti PG. Proteomics today, protein assessment and biomarkers using mass spectrometry, 2-D electrophoresis, and microarray technology. Hoboken: Wiley-VCH; 2005.
- [16] Gevaert K, Vanderkerckhove J. Protein identification methods in proteomics. *Electrophoresis* 2000;21:1145–54.
- [17] Jungblut P, Thiede B. Protein identification from 2-DE gels by MALDI mass spectrometry. *Mass Spectrom Rev* 1997;16:145–62.
- [18] Chalmers MJ, Gaskell SJ. Advances in mass spectrometry for proteome analysis. *Curr Opin Biotechnol* 2000;11:384–90.
- [19] Choudhary JS, Blackstock WP, Creasey DM, Cottrell JS. Matching peptide mass spectra to EST and genomic DNA databases. *Trends Biotechnol* 2001;19:S17–22.
- [20] Washburn MP, Wolters D, Yates III JR. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 2001;19:242–7.
- [21] Washburn MP, Yates III JR. In: Blackstock W, Mann M, editors. *Proteomics: a trend guide*. London: Elsevier; 2000. p. 27–30.
- [22] Liu H, Lin D, Yates III JR. Multidimensional separations for protein/peptide analysis in the post-genomic era. *BioTechniques* 2002;32:898–903.
- [23] Davis MT, Beierle J, Bures ET, McGinley MD, Mort J, Robinson JH, et al. Automated LC-LC-MS-MS platform using binary ion-exchange and gradient reversed-phase chromatography for improved proteomic analyses. *J Chromatogr B Analyt Technol Biomed Life Sci* 2001;752:281–91.
- [24] Chen J, Lee CS, Shen Y, Smith RD, Baehrecke EH. Integration of capillary isoelectric focusing with capillary reversed-phase liquid chromatography for two-dimensional proteomics separation. *Electrophoresis* 2002;23:3143–8.
- [25] Wang H, Kachman MT, Schwartz DR, Cho KR, Lubman DM. A protein molecular weight map of ES2 clear cell ovarian carcinoma cells using a two-dimensional liquid separation/mass mapping technique. *Electrophoresis* 2002; 23:3168–81.
- [26] Wilkins MR, Williams KL, Appel RD, Hochstrasser DF, editors. *Proteome research: new frontiers in functional genomics*. Berlin: Springer; 1997. 243 pp.
- [27] James P, editor. *Proteome research: mass spectrometry*. Berlin: Springer; 2001. 274 pp.
- [28] Simpson RJ, Dorow DS. Cancer proteomics: from signalling networks to tumor markers. *Trends Biotechnol* 2001;19: S40–8.
- [29] Martin DB, Nelson PS. From genomics to proteomics: techniques and applications in cancer research. *Trends Cell Biol* 2001;11:S60–5.
- [30] Baak JPA, Path FRC, Hermsen MAJA, Meijer G, Schmidt J, Janssen EAM. Genomics and proteomics in cancer. *Eur J Cancer* 2003;39:1199–215.
- [31] Ryu JW, Kim HJ, Lee YS, Myong NH, Hwang CH, Lee GS, et al. The proteomics approach to find biomarkers in gastric cancer. *J Korean Med Sci* 2003;18:505–9.
- [32] Alaiya AA, Franzén B, Auer G, Linder S. Cancer proteomics: from identification of novel markers to creation of artificial learning models for tumor classification. *Electrophoresis* 2000;21:1210–7.
- [33] Antonucci F, Chilosi M, Santacatterina M, Herbert B, Righetti PG. Proteomics and immunomapping of reactive lymph node and lymphoma. *Electrophoresis* 2002;23:356–62.
- [34] Antonucci F, Chilosi M, Parolini C, Hamdan M, Astner H, Righetti PG. Two-dimensional molecular profiling of mantle cell lymphoma. *Electrophoresis* 2003;24:2376–85.
- [35] Sobel A, Bouterin MC, Beretta L, Chneiweiss H, Doye V, Peyro-Saint-Paul H. Intracellular substrates for extracellular signaling. Characterization of a ubiquitous, neuron-enriched phosphoprotein (stathmin). *J Biol Chem* 1989;264:3765–72.
- [36] Sobel A. Stathmin: a relay phosphoprotein for multiple signal transduction? *Trends Biochem Sci* 1991;16:301–5.
- [37] Alli E, Bash-Babula J, Yang JM, Hait WN. Effect of stathmin on the sensitivity to antimicrotubule drugs in human breast cancer. *Cancer Res* 2002;62:6864–9.
- [38] Castagna A, Antonioli P, Aster H, Hamdan M, Righetti SC, Perego P, et al. A proteomic approach to cisplatin resistance in the cervix squamous cell carcinoma cell line A431. *Proteomics* 2004;4:3246–67.

- [39] Zanusso G, Righetti PG, Ferrari S, Terrin L, Farinazzo A, Cardone F, et al. Two-dimensional mapping of three phenotype-associated isoforms of the prion protein in sporadic Creutzfeldt–Jacob disease. *Electrophoresis* 2002;23:347–55.
- [40] Castagna A, Campostrini N, Farinazzo A, Zanusso G, Monaco S, Righetti PG. Comparative two-dimensional mapping of prion protein isoforms in human cerebrospinal fluid and central nervous system. *Electrophoresis* 2002;23:339–46.
- [41] Zanusso G, Ferrari S, Cardone F, Zampieri P, Gelati M, Fiorini M, et al. Detection of pathological prion protein in the olfactory epithelium of sporadic Creutzfeldt–Jakob disease subjects. *N Engl J Med* 2003;348:711–9.
- [42] Sanchez JC, Guillaume E, Lescuyer P, Allard L, Carrette O, Scherl A, et al. Cystatin C as a potential cerebrospinal fluid marker for the diagnosis of Creutzfeldt–Jakob disease. *Proteomics* 2004;4:2229–33.
- [43] Pieper R, Gatlin CL, Makusky AJ, Russo PS, Schatz CR, Miller SS, et al. The human serum proteome: display of nearly 3700 chromatographically separated protein spots on 2-D gels and identification of 325 distinct proteins. *Proteomics* 2003;3:1345–64.
- [44] Herbert B, Righetti PG. A turning point in proteome analysis: sample pre-fractionation via multicompartement electrolyzers with isoelectric membranes. *Electrophoresis* 2000;21:3639–48.
- [45] Sinha PK, Praus M, Kötting E, Gianazza E, Righetti PG. Two-dimensional maps in the most extended (pH 2.5–11) immobilized pH gradient interval. *J Biochem Biophys Methods* 1990;21:173–9.
- [46] Lomas L, Lathrop J, Thulasiraman V, Hammond D, Boschetti E. *Electrophoresis*, in press.
- [47] Gheorghiu L, Pathrop J, Lomas L, Thulasiraman V, Hammond D, Boschetti E. Sixth Siena Meeting, August 31–September 4, 2004, Poster No. 90.
- [48] Righetti PG, Castagna A, Antonioli P, Boschetti E. Pre-fractionation techniques in proteome analysis: the mining tools of the third millennium. *Electrophoresis* 2005;26:297–319.
- [49] Merrifield RB. Solid phase peptide synthesis: I. The synthesis of a tetrapeptide. *J Am Chem Soc* 1963;85:2149–54.
- [50] Furka A, Sebastyen F, Asgedom M, Dibo G. General method for rapid synthesis of multicomponent peptide mixtures. *Int J Pept Protein Res* 1991;37:487–93.
- [51] Lam KS, Salmon SE, Hersh EM, Hruby VJ, Kazmierski WM, Knapp RJ. A new type of synthetic peptide library for identifying ligand-binding activity. *Nature* 1991;354:82–4.
- [52] Cutler P. Protein arrays: the current state-of-the-art. *Proteomics* 2003;3:3–18.
- [53] Amstutz P, Forrer P, Zahnd C, Pluckthun A. In vitro display technologies: novel developments and applications. *Curr Opin Biotechnol* 2001;12:400–5.
- [54] Lueking AM, Horn H, Eickhoff H, Lehrqach H, Walter G. Protein microarrays for gene expression and antibody screening. *Anal Biochem* 1999;270:103–11.
- [55] Borrebaeck CA. Antibodies in diagnostics—from immunoassays to protein chips. *Immunol Today* 2000;21:379–82.
- [56] Borrebaeck CA, Ekstrom S, Molmberg Hager AC, Nilson J, Laurell T, Marko-Varga G. Protein chips based on recombinant antibody fragments: a highly sensitive approach as detected by mass spectrometry. *BioTechniques* 2001;30:1126–32.
- [57] Wiesner A. Detection of tumor markers with ProteinChip technology. *Curr Pharm Biotechnol* 2004;5:45–67.
- [58] Fung E.T., Thulasiraman V., Weinberger S.R., Dalmaso E.A.. Protein biochips for differential profiling. *Curr Opin Biotechnol* 2001;12:65–9.
- [59] MacBeath G., Schreiber S.L.. Printing proteins as microarrays for high-throughput function determination. *Science* 2000;289:1760–3.
- [60] Lin S, Tornatore P, King D, Orlando R, Weinberger SR. Limited acid hydrolysis as a means of fragmenting proteins isolated upon ProteinChip array surfaces. *Proteomics* 2001;1:1172–84.
- [61] Pawletz CP, Liotta LA, Petricoin III EF. New technologies for biomarker analysis of prostate cancer progression: laser capture microdissection and tissue proteomics. *Urology* 2001;57:160–3.
- [62] Boyle MD, Romer TG, Meeker AK, Sledjeski DD. Use of surface-enhanced laser desorption/ionization protein chip system to analyze streptococcal exotoxin B activity secreted by *Streptococcus pyogenes*. *J Microbiol Methods* 2001;46:87–97.
- [63] Thulasiraman V, McCutchens-Maloney SL, Motin VL, Garcia E. Detection and identification of virulence factors in *Yersinia pestis* using SELDI ProteinChip system. *BioTechniques* 2000;30:428–32.
- [64] Austen BM, Frears ER, Davies H. The use of SELDI protein chip arrays to monitor production of Alzheimer's beta-amyloid in transfected cells. *J Pept Sci* 2000;6:459–69.
- [65] Von Eggeling F, Davies H, Lomas L, Fiedler W, Junker K, Claussen U, et al. Tissue-specific microdissection coupled with ProteinChip array technologies: applications in cancer research. *BioTechniques* 2000;29:1066–70.
- [66] Merchant M, Weinberger SR. Recent advancements in surface-enhanced laser desorption/ionization-time of flight-mass spectrometry. *Electrophoresis* 2000;21:1164–7.
- [67] Weinberger S, Viner RI, Ho P. Tagless extraction-retentate chromatography: a new global protein digestion strategy for monitoring differential protein expression. *Electrophoresis* 2002;23:3182–92.
- [68] Zhang Z, Bast RC, Jinong YY, Sokoll LJ, Rai AJ, Rosenzweig JM, et al. Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. *Cancer Res* 2004;64:5882–90.
- [69] Chen G, Wang H, Gharib TGT, Huang CC, Thomas DG, Shedden KA, et al. Overexpression of oncoprotein 18 correlates with poor differentiation in lung adenocarcinomas. *Mol Cell Proteomics* 2003;2:107–16.
- [70] Chen W, Ji J, Xu X, He S, Ru B. Proteomic comparison between human young and old brains by 2D gel electrophoresis and identification of proteins. *Int J Dev Neurosci* 2003;21:209–16.

- [71] Schonberger SJ, Edgar PF, Kydd R, Faull RL, Cooper GJ. Proteomic analysis of the brain in Alzheimer's disease: molecular phenotype of a complex disease process. *Proteomics* 2001;1:1519–28.
- [72] Tilleman K, Van den Haute C, Geerts H, van Leuven F, Esmans EL, Moens L. Proteomic analysis of the neurodegeneration in the brain of tau transgenic mice. *Proteomics* 2002;2:656–65.
- [73] Witzmann FA, Li J, Strother WN, McBride WJ, Hunter L, Crabb DW, et al. Innate differences in protein expression in the nucleus accumbens and hippocampus of inbred alcohol-preferring and non-preferring rats. *Proteomics* 2003;3:1335–44.
- [74] Desrivieres S, Prinz T, Castro-Palomino Laria N, Meyer M, Boehm G, Bauer U, et al. Comparative proteomic analysis of proliferating and functionally differentiated mammary epithelial cells. *Mol Cell Proteomics* 2003;2:1039–54.
- [75] Somiari RI, Sullivan A, Russell S, Somiari S, Hu H, Jordan R, et al. High-throughput proteomic analysis of human infiltrating ductal carcinoma of the breast. *Proteomics* 2003;3:1863–73.
- [76] Seike M, Kondo T, Mori Y, Gemma A, Kudoh S, Sakamoto M, et al. Proteomic analysis of intestinal epithelial cells expressing stabilized beta-catenin. *Cancer Res* 2003;63:4641–7.
- [77] Stierum R, Gaspari M, Dommels Y, Quatas T, Pluk H, Jespersen S, et al. Proteome analysis reveals novel proteins associated with proliferation and differentiation of the colorectal cancer cell line Caco-2. *Biochim Biophys Acta* 2003;1650:73–91.
- [78] Cavusoglu N, Thierse D, Mohand-Said S, Chalmel F, Poch O, Van Dorsselaer A, et al. Differential proteomic analysis of the mouse retina: the induction of crystallin proteins by retinal degeneration in the rd1 mouse. *Mol Cell Proteomics* 2003;2:494–505.
- [79] He QY, Lau GK, Zhou Y, Yuen ST, Lin MC, Kung HF, et al. Serum biomarkers of hepatitis B virus infected liver inflammation: a proteomic study. *Proteomics* 2003;3:666–74.
- [80] Boraldi F, Bini L, Liberatori S, Armini A, Pallini V, Triozzo R, et al. Proteome analysis of dermal fibroblasts cultured in vitro from human healthy subjects of different ages. *Proteomics* 2003;3:917–29.
- [81] O'Farrell PH. High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* 1975;250:4007–21.
- [82] Anderson NG, Anderson NL. The human protein index. *Clin Chem* 1982;28:739–48.
- [83] Young DS, Anderson NG. (editors) Special issue: two dimensional electrophoresis. *Clin Chem* 1982;28:737–1092.
- [84] Anderson NL, Anderson NG. (editors) Special issue: two dimensional electrophoresis. *Clin Chem* 1984;30:1897–2108.
- [85] Kipling R. *The strange ride of Morrowbie Jukes*. London: 1885.

Formation of Truncated Proteins and High-Molecular-Mass Aggregates upon Soft Illumination of Photosynthetic Proteins

Sara Rinalducci,[†] Natascia Camprostrini,[‡] Paolo Antonioli,[‡] Pier Giorgio Righetti,^{‡,||}
 Peter Roepstorff,[§] and Lello Zolla^{*,†}

Department of Environmental Sciences, Tuscia University, Viterbo, Italy, Department of Industrial Agricultural Biotechnologies, University of Verona, Verona, Italy, Department of Chemistry, Materials and Engineering Chemistry "Giulio Natta", Politecnico di Milano, Milano, Italy, and Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense, Denmark

Received July 26, 2005

Different spot profiles were observed in 2D gel electrophoresis of thylakoid membranes performed either under complete darkness or by leaving the sample for a short time to low visible light. In the latter case, a large number of new spots with lower molecular masses, ranging between 15 000 and 25 000 Da, were observed, and high-molecular-mass aggregates, seen as a smearing in the upper part of the gel, appeared in the region around 250 kDa. Identification of protein(s) contained in these new spots by MS/MS revealed that most of them are simply truncated proteins deriving from native ones, fragments, or aggregates. This resulted from the formation of extremely reactive oxygen species (ROS) that can derive by the exposure of chlorophyll binding proteins of photosynthetic apparatus to low-intensity light during laboratory manipulation of sample for electrophoresis runs.

Keywords: reactive oxygen species • thylakoid membranes • chlorophyll-binding proteins • protein degradation • protein aggregation • *N*-formylkynurenine

Introduction

The primary function of photosynthetic light-harvesting complexes is the absorption of light and the transfer of the excitation energy to the photochemical reaction centers. They are also essential for regulation and distribution of excitation energy within the photosynthetic apparatus and respond to both short- and long-term fluctuations in light intensity and quality.^{1–4} The antenna proteins carry out these regulatory functions through different mechanisms: phosphorylation, dissociation into monomers, migration from grana to stroma, binding of xanthophyll, and so forth (for a review, see ref 5). However, it is well-known that when plants are exposed to higher light intensities, they become prone to photoinhibition⁶ unless they activate mechanisms to decrease the light interception or dissipate the excess energy in a harmless way to avoid damage to the photosynthetic apparatus.⁷ In photosynthesis, excitation of pigments [chlorophylls (Chls) or carotenoids (Car)] leads efficiently to a free-energy stabilizing charge separation. This charge separation competes with several loss processes of the singlet excitations: fluorescence, internal conversion, and intersystem crossing. Intersystem crossing, which occurs with a quantum yield of about 4.5% in plants,⁸ leads to the formation of unwanted and potentially dangerous chlorophyll triplet excited states (³Chl). These states are so high in energy that

they can interact with oxygen (a triplet in its ground state) to form singlet oxygen. Indeed, the lifetime of ³Chl is substantially shortened in the presence of oxygen,⁹ and in the absence of a competitive compound, one possible decay pathway involves ³Chl–³O₂ interaction. This could result in the formation of extremely reactive singlet oxygen that can damage pigments and proteins of any chlorophyll protein,¹⁰ especially in regard to proteins of the photosynthetic apparatus. This might have fatal consequences, since due to their chemical nature, reactive oxygen species (ROS) can react with a large variety of biomolecules.

However, many authors reported light-dependent proteases as responsible of protein degradation for arresting electron flow rate during high illumination.^{11–14} In the particular case of PSII antenna proteins, it has been suggested that a proteolytic activity is involved in the degradation, but up to now, the identity of the protease and the location of the substrate recognition site for the regulatory protease remain unknown. Recently, the possible involvement of reactive oxygen species in the *in vitro* degradation of the LHCII proteins was investigated with isolated monomeric, trimeric, and BBY systems.¹⁵ We found that, in these isolated subcomplexes, active oxygen is involved in LHCII protein degradation by random cleavage, starting from the NH₂ terminal region and resulting in the complete destruction of the antenna proteins.¹⁵ However, at low light intensities or at short illumination exposures, only a few amino acids from the NH₂ terminal region are removed and a truncated protein is formed. Obviously, when a small part of the native protein is removed, it escapes SDS-PAGE detection, explaining why nobody has previously reported this

* Corresponding author: Prof. Dr. Lello Zolla, Tuscia University, Largo dell'Università snc, 01100 Viterbo, Italy. Phone, 0039 0761 357 100; fax, 0039 0761 357 630; e-mail, zolla@unitus.it.

[†] Tuscia University.

[‡] University of Verona.

[§] University of Southern Denmark.

^{||} Politecnico di Milano.

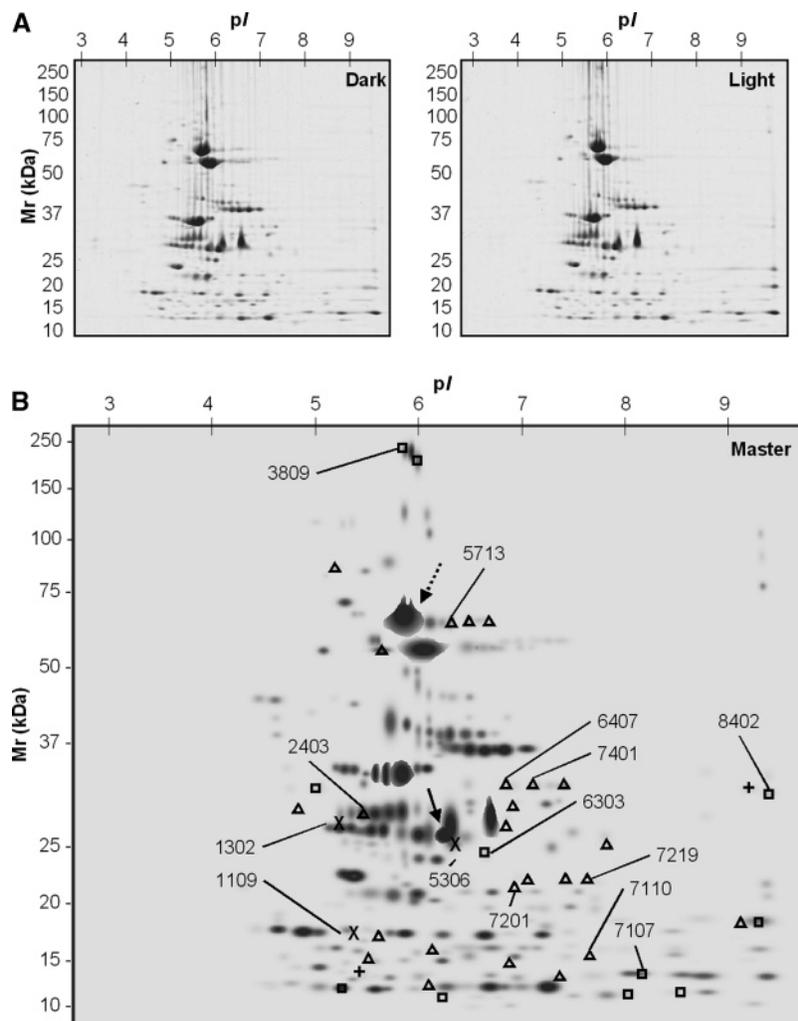


Figure 1. (A) 2D gel electrophoresis of spinach thylakoid membranes performed either under complete darkness (left) or by leaving the sample for 2 h under low visible light (right). (B) Master map of spinach thylakoid membranes created by comparison of samples left in complete darkness or under 2 h of light exposure. Analysis is performed by PDQuest software. Symbols refer to qualitative variation type: Δ , down-regulated; \square , up-regulated; \times , on (present only in the illuminated sample); $+$, off (present only in the control sample). PDQuest identity numbers are reported for spots identified by MS/MS.

occurrence. On the contrary, when 2D or more sophisticated methods are used, multiple spots may be observed.

Here, we report that 2D gel electrophoresis of spinach thylakoid membranes performed either under complete darkness or upon illumination of the sample with low-intensity visible light shows different spot profiles. In particular, a large number of new spots with lower molecular masses were detected, and their identification by MS/MS revealed that most of them are simply truncated proteins deriving from native ones where fragments were removed.

Materials and Methods

Chemicals and Materials. Tris, mineral oil, and DL-dithiothreitol (DTT), were purchased from Sigma-Aldrich Chemie GmbH (Steinheim, Germany). Glycerol, methanol, ethanol, and acetone were from Merck (Darmstadt, Germany). Acrylamide/Bis solution (40%), *N,N,N',N'*-tetramethylethylenediamine (TEMED), acrylamide (AA), ammonium persulfate (APS), the Protean IEF Cell, the GS-710 Densitometer, and the Versa Doc Scanner, as well as the linear Immobiline dry strips pH gradient 3–10 (17 cm long), were obtained from Bio-Rad Labs (Hercules,

CA); glycine, sodium dodecyl sulfate (SDS), urea, thiourea, tributylphosphine (TBP), and CHAPS were from Fluka. Bromophenol blue, carrier ampholytes, and agarose were purchased from Pharmacia Biotech (Uppsala, Sweden).

Isolation of the Thylakoid Membrane of Chloroplasts. For the separation of the thylakoid membrane of spinach (*Spinacia oleracea*), leaves, from a local market, were powdered in liquid nitrogen and subsequently homogenized in an ice-cold 20 mM tricine, pH 7.8, buffer containing 0.3 M sucrose and 5.0 mM magnesium chloride (B1 buffer). The whole procedure was performed while keeping leaves in complete dark. The homogenization was followed by filtration through one layer of Miracloth (Calbiochem, San Diego, CA) and centrifugation at 4 500g for 10 min at 4 °C. Pellets were suspended in B1 buffer and centrifuged again as above. These second pellets were resuspended in 20 mM tricine, pH 7.8, buffer containing 70 mM sucrose and 5.0 mM magnesium chloride (B2 buffer) and centrifuged at 10 000g for 10 min. The resulting pellets contain the thylakoid membrane. The concentration of chlorophyll was determined according to the method described by Porra et al.¹⁶

Table 1. Proteins Identified in 2D Electrophoresis Map of Spinach Thylakoid Membranes under Soft Light Exposure for 2 h^a

SSP	M_r (kDa)	pI	ratio (light/dark)	variation type	Peptides identified by MS/MS			score	NCBI accession number	protein [homologous organism]									
					m/z	charge state	start–end ^b				sequence								
3809	231.99	5.68	150.83	UP	799.71	(2+)	26–41	VVNTGTVLQVGDGIAR	488	gi 15988216	chain A, crystal structure of the chloroplast F1-ATPase								
					708.74	(2+)	95–107	IAQIPVSEAYLGR											
					422.89	(3+)	108–119	VINALAKPIDGR											
					633.68	(2+)	108–119	VINALAKPIDGR											
					475.11	(2+)	120–128	GEITASESR											
					408.12	(2+)	166–172	ELIIGDR											
					768.15	(2+)	203–216	ASSVAQVVTNFQER											
					659.54	(2+)	256–266	HTLHYDDLK											
					791.16	(1+)	457–462	YLVELR											
					811.12	(2+)	467–480	TNKPEFQEHSSTK											
					541.27	(3+)	467–480	TNKPEFQEHSSTK											
					626.24	(2+)	481–491	TFTEEAELK											
					848.17	(1+)	501–507	FLLQEQA											
					780.02	(2+)	3–17	INPTSDPGVSTLEK				479	gi 7636116	ATP synthase beta subunit					
					646.15	(2+)	40–50	MPNIYNALIVK											
604.61	(2+)	76–87	AVAMSATDGLTR																
590.04	(2+)	135–145	SAPFTQLDTK																
523.13	(2+)	155–163	VVDLLAPYR																
488.07	(2+)	168–178	IGLFGGAGVGK																
5713	62.70	6.21	0.45	DOWN	759.09	(2+)	218–231	ESGVINEQNAESK	388	gi 114527	ATP synthase alpha chain								
					744.57	(2+)	249–261	VGLTALTMAEYFR											
					717.13	(2+)	278–291	FVQAGSEVSALLGR											
					800.34	(2+)	26–41	VVNTGTVLQVGDGIAR											
					708.90	(2+)	95–107	IAQIPVSEAYLGR											
					633.90	(2+)	108–119	VINALAKPIDGR											
					754.03	(3+)	142–162	SVYEPLQTGLIADAMIPVGR ^c											
					768.31	(2+)	203–216	ASSVAQVVTNFQER											
					659.38	(2+)	255–266	HTLHYDDLK											
					770.30	(2+)	285–297	EAYPGDVFYLHSR											
					811.36	(2+)	467–480	TNKPEFQEHSSTK											
					626.41	(2+)	481–491	TFTEEAELK											
					6407	29.32	6.92	0.58				DOWN	617.77	(2+)	2–13	VKGPGLYSDIGK	281	gi 1256259	voltage-dependent anion channel protein
													464.16	(3+)	2–14	VKGPGLYSDIGKR			
													581.85	(2+)	4–14	GPGLYSDIGKR			
654.43	(2+)	48–59	KGELFLADVSTK																
436.69	(3+)	48–59	KGELFLADVSTK																
590.82	(2+)	49–59	GELFLADVSTK																
824.85	(2+)	249–264	SIVTISGEVDTSAIEK																
7401	29.36	7.16	0.66	DOWN					617.65	(2+)	2–13		VKGPGLYSDIGK	447	gi 1256259	voltage-dependent anion channel protein			
463.96					(3+)	2–14	VKGPGLYSDIGKR												
503.67					(2+)	4–13	GPGLYSDIGK												
581.65					(2+)	4–14	GPGLYSDIGKR												
654.25					(2+)	48–59	KGELFLADVSTK												
436.52					(3+)	48–59	KGELFLADVSTK												
590.15					(2+)	49–59	GELFLADVSTK												
744.66					(3+)	71–91	VDTNSNLF ^d TTITVDEPAPGLK												
1116.81	(2+)	71–91	VDTNSNLF ^d TTITVDEPAPGLK																
824.85	(2+)	249–264	SIVTISGEVDTSAIEK																
8402	28.40	9.52	1.63	UP	609.87	(2+)	89–98	DLQMVNLTLR ^c	108	gi 7716458	prohibitin [<i>Zea mays</i>]								
					584.90	(2+)	123–133	VLPSIGNEVLK											
2403	25.80	5.18	0.51	UP	740.34	(2+)	9–21	TVQSSSPWYGPDGR	116	gi 47168897	chain J, crystal structure of spinach major light-harvesting complex at 2.72 Å resolution								
					853.93	(2+)	9–23	TVQSSSPWYGPDVRK											
					959.48	(2+)	71–87	WAMLGALGCVPELLAR											
					492.14	(2+)	92–99	FGEAVWFK											
					1045.87	(3+)	204–2	GPLENLADHLADPVNNNAWNFATNFVPGK											
1302	24.02	4.92	364.00	ON	740.34	(2+)	9–21	TVQSSSPWYGPDGR	82	gi 47168897	chain J, crystal structure of spinach major light-harvesting complex at 2.72 Å resolution								
					853.93	(2+)	9–23	TVQSSSPWYGPDVRK											
					492.16	(2+)	92–99	FGEAVWFK											
					1045.78	(3+)	204–232	GPLENLADHLADPVNNNAWNFATNFVPGK											
					5306	22.23	6.26	1100.06				ON	768.67	(2+)	15–27	NTEFMPYNGDGFK ^c	336	gi 21265	23 kDa OEC protein
401.91	(3+)	39–48	EKEFPQQVLR																
945.45	(1+)	41–48	EFPGQVLR																
795.67	(3+)	49–69	YEDNFDATSNLSVLVQPTDKK																
908.72	(3+)	97–124	TDSEGGFDSGVASANVLESSTPVDGK																
622.53	(2+)	125–134	QYYSITVLTR ^d																
613.80	(2+)	125–134	QYYSITVLTR																
599.80	(3+)	135–152	TADGDEGGKHQVIAATVK																
483.93	(2+)	144–152	HQVIAATVK																
680.36	(2+)	174–186	KFVESATSSFSVA																
616.39	(2+)	175–186	FVESATSSFSVA																

Table 1. (Continued)

SSP	M_r (kDa)	pI	ratio (light/dark)	variation type	Peptides identified by MS/MS				score	NCBI accession number	protein [homologous organism]
					m/z	charge state	start–end ^b	sequence			
6303	21.42	6.62	2.63	UP	601.78	(2+)	39–48	EKEFPQQLR	250	gi 21265	23 kDa OEC protein
					795.63	(3+)	49–69	YEDNFDATSNLSVLVQPTDKK			
					622.35	(2+)	125–134	QYYISITLVR			
					599.81	(3+)	135–152	TADGDEGGKQVIAATVK			
					680.31	(2+)	174–186	KFVESATSSFSVA			
			616.25	(2+)	175–186	FVESATSSFSVA					
7219	18.67	7.73	0.44	DOWN	847.94	(3+)	20–44	GFTPPELDPNTSPIFAGSTGGLLR	78	gi 19855891	Photosystem I reaction center subunit II (PsaD)
					842.39	(2+)	61–75	EQIFEMPTGGAAIMR ^c			
					443.31	(2+)	146–153	NVSPIEVK			
7201	17.87	7.05	0.62	DOWN	1407.11	(1+)	15–27	VYFDISIGNPVGK	84	gi 32130452	(segment) peptidyl-prolyl <i>cis</i> – <i>trans</i> isomerase TLP20, chloroplast (PPIase) (Rotamase)
					704.56	(2+)	15–27	VYFDISIGNPVGK			
1109	14.27	5.04	234.05	ON	799.89	(2+)	26–41	VVNTGTVLQVGDGIAR	88	gi 67825	H ⁺ -transporting two-sector ATPase (EC 3.6.3.14) alpha chain
					708.83	(2+)	95–107	IAQIPVSEAYLGR			
7110	13.06	7.67	0.59	DOWN	1670.88	(1+)	69–83	IGNNEITLVNDAER	88	gi 114606	ATP synthase epsilon chain
					1155.70	(1+)	111–120	QKIEANLALR			
7107	12.16	8.16	1.52	UP	476.92	(3+)	6–19	IKVDKPLGIGGGMK ^c	80	gi 81471	photosystem II 10K protein precursor
					720.91	(2+)	8–21	VDKPLGIGGGMKLR			

^a In the first column, the number of the identified spots is reported (see Figure 1B); the second and third columns, respectively, show the apparent molecular mass and the isoelectric points, whereas the quantitative and qualitative variations were reported in the fourth and fifth columns. MS/MS data with m/z values, ion charge states, start–end positions, and amino acid sequences of identified peptides were listed in the sixth, seventh, eighth, and ninth columns, respectively. In the following column, there is the relative score. Finally, the NCBI accession number and the protein identification are indicated in the 11th and 12th columns, respectively. If the sequence(s) did not match a protein from spinach, a homologous organism is displayed. ^b Start-end positions of identified peptides were calculated against complete amino acid sequence of the mature protein. ^c Peptide with oxidation of methionine (M). ^d Peptide with Pyro-glu modification of glutamine (Q).

Protein Extraction and Sample Preparation for 2D Electrophoresis. Thylakoid membranes were first resuspended in B2 buffer, supplemented with protease inhibitors, and incubated in dark and in light for 2 h (samples used in ESR and 2D PAGE analyses) and 6 h for 2D electrophoresis. The three samples were then subjected to pigment removal and protein extraction according to Hippler et al.,¹⁷ with minor modifications as follows: after every methanol addition, the tubes were vigorously mixed for 10 min and every centrifugation was at 20 000g for 10 min after 1 h incubation at –20 °C.

The obtained pellets were finally resuspended in a solution containing 7 M urea, 2 M thiourea, 3% (w/v) CHAPS, 1% (v/v) carrier ampholytes (pH 3–10), 40 mM Tris, 5 mM TBP (as reducing agent), and 10 mM AA (as alkylating agent).^{18,19} After incubation at room temperature for 1 h (alkylation was blocked with 10 mM DTT), the samples were centrifuged at 20 000g for 15 min and the supernatants were recovered and stored at –20 °C until use.

2D PAGE Analysis. Five replicate gels were run for each one of the two samples kept in dark or exposed to light for 2 or 6 h. The first dimension run was performed on strips (17 cm length, 0.5 mm thickness) with a linear pH gradient from pH 3 to pH 10.²⁰ The IPG strips were rehydrated with 4 mg/mL of sample containing traces of bromophenol blue for monitoring the electrophoretic run. The total product time × voltage applied was 75 000 V/h for each strip, and the temperature was set at 20 °C. Each strip was then equilibrated with a SDS denaturing solution containing 6 M urea, 2% SDS, 20% glycerol, and 0.375 M Tris-HCl (pH 8.8), and the second dimension run was performed on 8–20% T linear gradient polyacrylamide gels (20 cm × 20 cm). At the end of the separations, the gels were fixed with a solution containing 40% ethanol and 10% acetic acid for 30 min and stained with the fluorescent dye Sypro ruby. Gel images were digitized and analyzed with the PDQuest 7.3

software (Bio-Rad, Hercules, CA). To compare the gels involved in the study, spot quantities were normalized against the total optical density in the valid spots of each gel. The two groups of gels exposed to light or kept in dark for 2–6 h were compared either at a qualitative or at a quantitative level, with a minimum variation threshold of 1.5. Statistic significance of each spot was evaluated by Student's *t*-test, with a significance level of 0.05.

In-Gel Digestion. Spots were carefully cut out from 2D Sypro ruby-stained gels and subjected to in-gel trypsin digestion according to Shevchenko et al.²¹ with minor modifications. The gel pieces were swollen in a digestion buffer containing 50 mM NH₄HCO₃ and 12.5 ng/μL of trypsin (modified porcine trypsin, sequencing grade, Promega, Madison, WI) in an ice bath. After 30 min, the supernatant was removed and discarded, 20 μL of 50 mM NH₄HCO₃ was added to the gel pieces, and digestion was allowed to proceed at 37 °C overnight. The supernatant containing tryptic peptides was dried by vacuum centrifugation. Prior to mass spectrometric analysis, the peptide mixtures were redissolved in 10 μL of 5% FA (formic acid).

Peptide Sequencing by Nano-RP-HPLC–ESI–MS/MS. Peptide mixtures were separated by using an LC-Packings (a Dionex company, Italy) autosampler/nano-HPLC. A sample volume of 8 μL was loaded by the autosampler and concentrated on a commercial trap column (Zorbax 300 SB–C18, 5 μm, 300 μm i.d. × 5 mm, Agilent Technologies) at a flow rate of 10 μL/min. Separation was performed using a homemade capillary column (Zorbax 300 SB–C18, 3.5 μm, 75 μm i.d. × 10 cm, Agilent Technologies) and a flow rate of 200 nL/min. HPLC solvents contained 0.1% FA and either 2% ACN (solvent A) or 98% ACN (solvent B). The column was pre-equilibrated with 100% solvent A. Elution was performed by multisegment gradient with a first step from 0 to 5% solvent B in 5 min, a second one from 5 to 12% solvent B in 2 min, a third one from

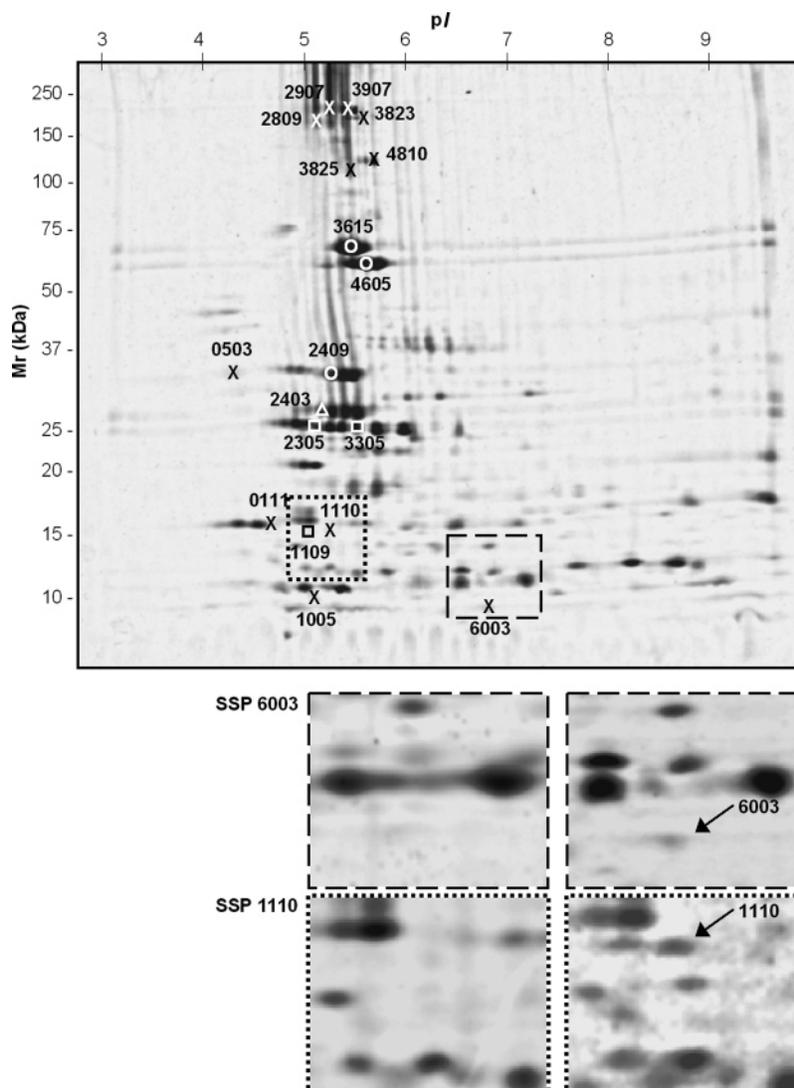


Figure 2. 2D electrophoresis of spinach thylakoid membranes left under 6 h of visible light. Symbols are as in Figure 1 (in addition, \circ indicates native form of some proteins). Black symbols are changed into white where required by the background staining. PDQuest identity numbers are reported for spots identified by MS/MS. The two lower panels show the focused images of the spots SSP 1110 and 6003, corresponding to CP47 and PsaB protein, respectively.

12 to 50% solvent B in 30 min, and a last segment from 50 to 100% solvent B in 8 min.

Peptides were eluted directly into an ion trap Esquire 3000 plus (Bruker-Daltonik, Germany). Capillary voltage was 1.5–2 kV, and a dry gas flow rate of 3 L/min was used with a temperature of 230 °C. The scan range used was from 400 to 1800 m/z . Protein identification was performed by searching in the National Center for Biotechnology Information nonredundant database (NCBI nr) using the Mascot program (<http://www.matrixscience.com>). The following parameters were adopted for database searches: complete carbamidomethylation of cysteines and partial oxidation of methionines, peptide Mass Tolerance ± 1.2 Da, Fragment Mass Tolerance ± 0.9 Da, missed cleavages 2. For positive identification, the score of the result of $[-10 \text{ Log}(P)]$ had to be over the significance threshold level ($P < 0.05$).

MALDI-TOF/TOF Measurements. Peptides were desalted and concentrated according to Gobom et al.²² Homemade 5 mm nanocolumns were packed with POROS R2 chromatographic resin (PerSeptive Biosystems, Framingham, MA) in a constricted GELoader tip (Eppendorf, Hamburg, Germany). A

syringe was used to force liquid through the columns by applying a gentle air pressure. The columns were equilibrated with 20 μL of 5% FA (formic acid), and the analyte solutions were added. After washing with 20 μL of 5% FA, the bound peptides were eluted directly onto the MALDI target with 0.5 μL of CHCA (α -cyano-4-hydroxycinnamic acid) solution (5 $\mu\text{g}/\mu\text{L}$ in ACN, 0.1% TFA, 70:30 v/v). Samples were analyzed on the Applied Biosystems 4700 Proteomics Analyzer MALDI-TOF/TOF (Applied Biosystems, Framingham, MA). Positively charged ions were analyzed in reflectron mode, and the collision gas used for fragmentation was atmospheric air. The MoverZ software (Proteometrics, Winnipeg, Canada) and the Data Explorer (Applied Biosystems) were used for spectra analysis. The spectra were calibrated using trypsin autolysis products (m/z 842.504 and 2211.109) as internal standards. In a few cases, this was not possible, and external calibration was performed using peaks from tryptic digest of β -lactoglobulin. Protein identification was performed by searching in the NCBI database using the Mascot program (<http://www.matrixscience.com>). The following parameters were considered for database searches: all entries, peptide tolerance at 70 ppm, MS/MS

Table 2. Proteins Identified in 2D Electrophoresis Map of Spinach Thylakoid Membranes under Soft Light Exposure for 6 h

SSP	apparent M_r (kDa)	pI	Peptides identified by MS/MS			score	NCBI accession number	protein [homologous organism]	
			m/z	charge state	start–end ^a				sequence
2907	400.0	5.4	983.50	(1+)	123–130	FGEAVWFK	97	gi 50839002	chlorophyll <i>a/b</i> binding protein [<i>Oryza sativa</i>]
			2408.11	(1+)	186–208	VYPGGAFDPLGLADDPDTFAELK			
3907	401.4	5.5	1036.52	(1+)	15–22	ERIEGYNR	462	gi 67825	H ⁺ -transporting two-sector ATPase alpha chain
			1955.01	(1+)	23–41	EKVVNTGTVLQVGDGIAR			
			1598.89	(1+)	26–41	VVNTGTVLQVGDGIAR			
			1416.78	(1+)	95–107	IAQIPVSEAYLGR			
			1266.75	(1+)	108–119	VINALAKPIDGR			
			2197.28	(1+)	108–128	VINALAKPIDGRGEITASESR			
			1535.78	(1+)	203–216	ASSVAQVVTNFQER			
			1317.71	(1+)	256–266	HTLIYDDLSK			
			1553.73	(1+)	285–297	EAYPGDVFYLHSR			
			771.51	(1+)	417–422	LRELLK			
			920.56	(1+)	456–462	KYLVELR			
			792.45	(1+)	457–462	YLVELR			
1621.84	(1+)	467–480	TNKPEFQEIISSTK						
2809	300.0	5.3	817.49	(1+)	1–7	DIITSIR	139	gi 33327884	cytochrome b-559 alpha subunit
			954.57	(1+)	50–58	QGIIPLITGR			
			1485.70	(1+)	59–70	FDSLEQLDEFGR			
3823	344.0	5.6	983.50	(1+)	123–130	FGEAVWFK	133	gi 33327771	ATP synthase beta subunit
			2408.19	(1+)	186–208	VYPGGAFDPLGLADDPDTFAELK			
4810	148.5	5.8	1955.02	(1+)	110–127	IFNVLGEPVDNLGPVDTR	484	gi 33327771	ATP synthase beta subunit
			1949.99	(1+)	262–277	DVNEQDVLFFIDNIFR			
			1433.80	(1+)	278–291	FVQAGSEVSALLGR			
3825	132.1	5.5	1735.03	(1+)	23–39	IAQIIGPVLDVAFPPGK	362	gi 67825	H ⁺ -transporting two-sector ATPase alpha chain
			1291.72	(1+)	40–50	MPNIYNALIVK ^b			
			1955.03	(1+)	110–127	IFNVLGEPVDNLGPVDTR			
			1007.58	(1+)	146–154	LSIFETGIK			
			1201.72	(1+)	155–164	VVDLLAPYRR			
			975.56	(1+)	168–178	IGLFGGAGVGK			
			1328.69	(1+)	192–205	AHGGVSVFVGGER			
			1617.81	(1+)	232–246	VALVYQMNPPGAR ^b			
			1433.79	(1+)	278–291	FVQAGSEVSALLGR			
			1955.01	(1+)	23–41	EKVVNTGTVLQVGDGIAR			
3615	64.4	5.5	1598.88	(1+)	26–41	VVNTGTVLQVGDGIAR	532	gi 67825	H ⁺ -transporting two-sector ATPase alpha chain
			1416.77	(1+)	95–107	IAQIPVSEAYLGR			
			1266.73	(1+)	108–119	VINALAKPIDGR			
			2197.34	(1+)	108–128	VINALAKPIDGRGEITASESR			
			1535.77	(1+)	203–216	ASSVAQVVTNFQER			
			1317.69	(1+)	256–266	HTLIYDDLSK			
			1553.73	(1+)	285–297	EAYPGDVFYLHSR			
			2113.08	(1+)	463–480	TVVKTNKPEFQEIISSTK			
			799.72	(2+)	26–41	VVNTGTVLQVGDGIAR			
			708.80	(2+)	95–107	IAQIPVSEAYLGR			
4605	54.8	5.7	634.19	(2+)	108–119	VINALAKPIDGR	836	gi 33327771	ATP synthase beta subunit
			423.05	(3+)	108–119	VINALAKPIDGR			
			768.32	(2+)	203–216	ASSVAQVVTNFQER			
			659.16	(2+)	256–266	HTLIYDDLSK			
			776.94	(2+)	285–297	EAYPGDVFYLHSR			
			811.26	(2+)	467–480	TNKPEFQEIISSTK			
			626.25	(2+)	481–491	TFTEAEALLK			
			847.95	(1+)	501–507	FLLQEQA			
			779.69	(2+)	3–17	INPTSDPGVSTLEK			
			867.53	(2+)	23–39	IAQIIGPVLDVAFPPGK			
			578.94	(3+)	23–39	IAQIIGPVLDVAFPPGK			
			652.30	(3+)	110–127	IFNVLGEPVDNLGPVDTR			
			977.75	(2+)	110–127	IFNVLGEPVDNLGPVDTR			
			589.76	(2+)	135–145	SAPAFQQLDTK			
504.26	(2+)	146–154	LSIFETGIK						
523.25	(2+)	155–163	VVDLLAPYR						
1044.88	(1+)	155–163	VVDLLAPYR						
401.44	(3+)	155–164	VVDLLAPYRR						
488.27	(2+)	168–178	IGLFGGAGVGK						
974.98	(1+)	168–178	IGLFGGAGVGK						
744.28	(2+)	179–191	TVLIMELINNIK ^b						
443.69	(3+)	192–205	AHGGVSVFVGGER						
664.71	(2+)	192–205	AHGGVSVFVGGER						
759.01	(2+)	218–231	ESGVINEQNIK ^b						
744.04	(2+)	249–261	VGLTALTMAEYFR ^b						
717.25	(2+)	278–291	FVQAGSEVSALLGR						
722.09	(2+)	379–390	IVGEEHYEIAQR						

Table 2. (Continued)

SSP	apparent M_r (kDa)	pI	Peptides identified by MS/MS				score	NCBI accession number	protein [homologous organism]			
			m/z	charge state	start–end ^a	sequence						
0503	31.0	4.5	2294.11	(1+)	186–206	FEEKDGDIDYAAVTVQLPGGER	100	gi 21283	OEC 33kD precursor protein			
			1760.95	(1+)	190–206	DGIDYAAVTVQLPGGER						
			964.62	(1+)	207–214	VPFLFTIK						
			2284.13	(1+)	215–235	QLVASGKPEFSGDFLVPSYR						
2409	30.4	5.3	808.76	(3+)	105–128	GTGTANQCPTVEGGVDSFAFKPGK	344	gi 21283	OEC 33kD precursor protein			
			1150.18	(2+)	165–185	LYTLDEIEGPFVSSDGTVK						
			587.44	(3+)	190–206	DGIDYAAVTVQLPGGER						
			880.63	(2+)	190–206	DGIDYAAVTVQLPGGER						
			482.79	(2+)	207–214	VPFLFTIK						
			761.93	(3+)	215–235	QLVASGKPEFSGDFLVPSYR						
			781.70	(2+)	246–262	GGSTGYDNAVALPAGGR						
			668.04	(2+)	321–331	IEGVWYAQLEQ						
2403	27.0	5.3	740.35	(2+)	9–21	TVQSSSPWYGPDR	149	gi 47168897	chain J, crystal structure of spinach major light-harvesting complex at 2.72 Å resolution			
			1310.64	(3+)	24–60	YLGPFSGESPSYLTGEFFPGDYGWDTAGLSADPETFAK						
			492.16	(2+)	92–99	FGEAVWFK						
			1045.77	(3+)	204–232	GPLENLADHLADPVNNNAWNFATNFVPGK						
			711.35	(2+)	9–21	NVSSGSPWYGPDR				125	gi 556367	light-harvesting chlorophyll <i>a/b</i> -binding protein [<i>Prunus persica</i>]
			492.16	(2+)	92–99	FGEAVWFK						
2305	23.9	5.1	492.16	(2+)	92–99	FGEAVWFK	109	gi 115780	chlorophyll <i>a/b</i> -binding protein, chloroplast precursor (LHCII type I)			
			1367.78	(1+)	61–70	NRELEVIHCR						
			1176.27	(3+)	178–212	IAGGPLGEVVDPLYPGGSFDPLGLADDPFAFAELK						
			1045.68	(3+)	204–232	GPLENLADHLADPVNNNAWNFATNFVPGK						
3305	23.1	5.5	1673.94	(1+)	65–81	YAMLGAVGAIAPILGK	125	gi 20794	Type III chlorophyll <i>a/b</i> -binding protein [<i>Pinus sylvestris</i>]			
			1689.94	(1+)	65–81	YAMLGAVGAIAPILGK ^b						
			997.55	(1+)	142–149	QYFLGLEK						
			980.52	(1+)	142–149	QYFLGLEK ^c						
0111	14.9	4.7	729.93	(3+)	237–255	FGQEEETYNIIVAAHGYFGR	123	gi 61230125	Photosystem II D1 protein			
1109	14.2	5.1	799.87	(2+)	26–41	VVNTGTVLQVGDGIAR	141	gi 67825	H ⁺ -transporting two-sector ATPase alpha chain			
			708.83	(2+)	95–107	IAQIPVSEAYLGR						
			1266.77	(1+)	108–119	VINALAKPIDGR						
1110	14.3	5.2	656.28	(3+)	287–304	RVSAGLAENQSFSEAWSK	327	gi 19855072	CP47			
			905.86	(2+)	288–304	VSAGLAENQSFSEAWSK						
			743.32	(2+)	309–321	LAFYDYIGNNPAK						
			723.97	(3+)	327–347	AGSMDNNGDGIAGVWLGHPFR						
			1074.77	(3+)	390–419	YSVEQVGVTVVEFYGGELNGVSYSDPATVKK						
			770.67	(3+)	477–497	DVFAGIDPDLVQVEFGAFQK						
1005	8.1	5.1	954.57	(1+)	50–58	QGIPLITGR	85	gi 33327884	cytochrome b-559 alpha subunit			
			1485.69	(1+)	59–70	FDSLEQLDEFNR						
6003	7.5	6.9	1320.66	(1+)	8–19	FSQGLAQDPTTR	72	gi 131150	Photosystem I P700 chlorophyll A apoprotein A2 (PsaB)			
			830.71	(3+)	21–41	IWFGIATAHDFESHDDITEER						

^a Start–end positions of identified peptides were calculated against complete amino acid sequence of the mature protein. ^b Peptide with oxidation of methionine (M). ^c Peptide with Pyro-glu modification of glutamine (Q).

tolerance at 0.5 Da, carbamidomethylation of cysteine (fixed modification), and methionine oxidation (variable modifications). For positive identification, the score of the result of [–10 Log(*P*)] had to be over the significance threshold level (*P* < 0.05).

Electron Spin Resonance (ESR) Spectroscopy. Spinach thylakoid membranes diluted in the B2 buffer to 0.1 mg chlorophyll/mL were exposed to a low light intensity for 2 h at room temperature, and 80 mM DMPO (5,5-dimethyl-1-pyrroline *N*-oxide) was added. ESR spectra were measured with a Bruker ESP300 spectrometer by using 10-mW power at 9.79 GHz. Spectra were recorded using 1.0-G modulation and 100-G scanning in 21 s.

Results and Discussion

To verify if illumination of the photosynthetic apparatus by low-intensity light, which occurs during laboratory manipula-

tions, may induce small degradation of protein components, spinach thylakoid membranes, once extracted in darkness, were prepared for electrophoretic runs either in complete darkness or under normal laboratory light conditions (<100 μmol m⁻² s⁻¹) for 2–6 h.

Figure 1A compares 2D maps of thylakoids kept in dark and under light exposition for 2 h. The two maps, apparently similar, show in illuminated thylakoids some new spots prevalently localized in the low-molecular-mass region and high-molecular-mass aggregates, as a smearing in the region round about 250 kDa. To have statistically significant results (Student's *t*-test, *p* < 0.05), five replicate gels for each sample were performed and a master map was created (Figure 1B). The PDQuest analysis allowed the study of proteins that were significantly increased or decreased after thylakoid illumination. Only the protein spots selected by both the statistic test and the quali-quantitative analysis were considered for further

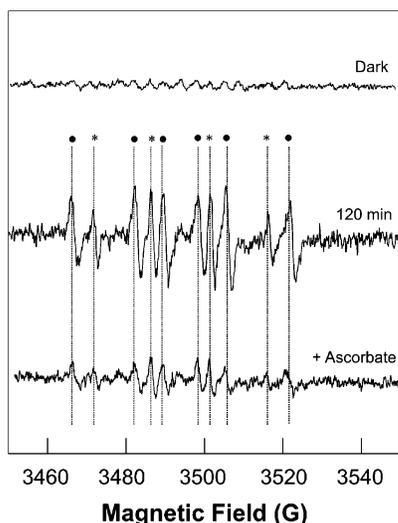


Figure 3. ESR spectra of spin-trapped radicals generated under illumination of spinach thylakoid membranes for 2 h in the presence of 80 mM DMPO. Circles indicate line components belonging to DMPO-R \cdot (hyperfine splittings $a^N = 14.7$ G, $a^H = 22.6$ G); asterisks mark ESR spectra of DMPO-HO \cdot (hyperfine splittings $a^N = a^H = 14.7$ G). Lower panel shows ESR spectra of thylakoids illuminated for 2 h with addition of ascorbate.

investigations by MS. Since light exposure was short-lasting and not really intense, we decided to set a low quantity variation threshold (1.5) for the light-exposed proteins as compared to the control ones. Such an analysis identified 41 varied protein spots after soft light exposure, concentrated in the low-molecular-mass region (below 35 kDa) of the maps and evenly distributed along the pH gradient (Figure 1B). Out of the 41 spots, the major part appeared to be increased (11) or decreased (25) after light treatment. All of the varied spots detected in the molecular mass region ranging from 35 up to 100 kDa looked decreased, while at very high-molecular masses (above 150 kDa), some spots appeared to be clearly increased after light exposure. Light-modulated spots of membrane proteins were cut out from the SDS-PAGE, digested with trypsin, and analyzed either by MALDI-TOF or by internal peptide sequencing, by using mainly RP-HPLC-ESI-MS/MS and MALDI-TOF/TOF.

Table 1 summarizes the MS/MS analyses of some representative spots collected in the low-, medium-, and high-molecular-mass areas of the gel; qualitative and quantitative variations are also listed.

The new spot SSP 5306 resulted from our analysis to be the OEC23 protein, which in the control (pointed out by an arrow in Figure 1B) is in a region of higher molecular masses. This indicated that this new spot contains a truncated OEC23 which has lost some peptides. A damaged form of the OEC23 protein has been previously described by Henmi et al.²³ By analysis of the sequence coverage obtained (see Table 1), it was possible to note that all of the identified peptides were localized at the C-terminal of the protein, suggesting a possible removal of some amino acids at the N-terminal portion. Similarly, spot SSP 1109 represents a fragment of the ATPase α -subunit whose intact form was identified in the spot indicated by a dotted arrow in Figure 1B, whereas spot SSP 1302 contains a truncated light-harvesting Lhcb1 protein having a lower molecular mass than the native one present in the spot 2403. The appearance of proteins showing lower molecular masses, as a consequence

of lacking a small fragment, agrees with what was previously revealed by HPLC-ESI-MS.¹⁵ It was found that the primary cleavages took place in the hydrophilic portion of the NH₂ region, and these were apparent only at low light intensity and short illumination times; otherwise, the protein would be completely destroyed. In the spot 3809, recovered at high-molecular masses (M_r of 232 kDa), our analysis showed the presence of the two ATPase subunits (having a molecular mass about 50 kDa). The presence of ATPase at this high-molecular masses can only be explained by the formation of aggregates, which are absent in the control, considering that supramolecular complexes cannot be present in sample treated with SDS. Finally, spots of decreased stain intensity can be attributed to a decreased amount of protein due to its degradation or participation to aggregation.

Thus, it can be concluded that short exposition of thylakoid membranes to low-intensity light induces some protein modifications which may be better-investigated by treating a sample for longer illumination times. With this aim in mind, thylakoid membranes were exposed to lab light for 6 h and then analyzed by 2D electrophoresis. Figure 2 shows the 2D gels obtained. The new spots revealed by PDQuest were identified by a cross and by identity numbers (SSP). Additional fragments and aggregates were observed. Table 2 lists the examined spots, along with their apparent molecular mass, pI, database accession number, and protein description; the identified peptides by MS/MS are also reported.

At high-molecular masses, our analysis revealed the Lhcb1 light-harvesting protein in the spot 2907, as an aggregation product. Native Lhcb1 has an apparent M_r of 27 kDa and was found in the spot 2403.

Regarding the new fragments, it can be seen that, upon prolonged illumination time, beside the presence of truncated forms, such as ATPase α -chain (SSP 1109), and Lhcb1 (SSP 2305), previously revealed at short time, other new fragments appeared: one deriving from CP47 protein (spot SSP 1110); one from PsaB (spot SSP 6003), the large subunits of PSI reaction center; one from D1 (spot SSP 0111); and one from cyt b_{559} (spot SSP 1005). The lower panel of Figure 2 displays the focused images of spots corresponding to CP47 and PsaB proteins, chosen as representative examples of a light-induced fragment removal from the N-terminal or C-terminal of native protein, respectively (see Table 2). In agreement with what was observed here, illumination of purified PSI from spinach or cucumber leaves showed the degradation of the *psaA/psaB* gene product.²⁴ The size of degradation products, estimated from electrophoretic mobility of each band, was of 51 and 45 kDa, even by low-light irradiances ($60\text{--}90 \mu\text{mol m}^{-2} \text{s}^{-1}$),²⁵ as may have occurred in our case.

Finally, our investigation detected proteins representing OEC33 (Oxygen Evolving Complex) in the new spot SSP 0503, having an apparent molecular mass close to that of native protein (SSP 2409) but having slight different pI values. This suggests that some amino acids were modified inducing a more acid isoelectric point for the entire protein without any degradation of the polypeptide chain. This agrees with what was reported previously by Henmi et al.²³ Moreover, it has been documented that Cys, Met, Trp, His, and Tyr are highly susceptible to oxidation,²⁶ and at least two of these can be charged under physiological conditions (Cys as an anion, and His in a protonated form). Oxidation of either or both of these could result in an overall change in charge. Furthermore, OEC33

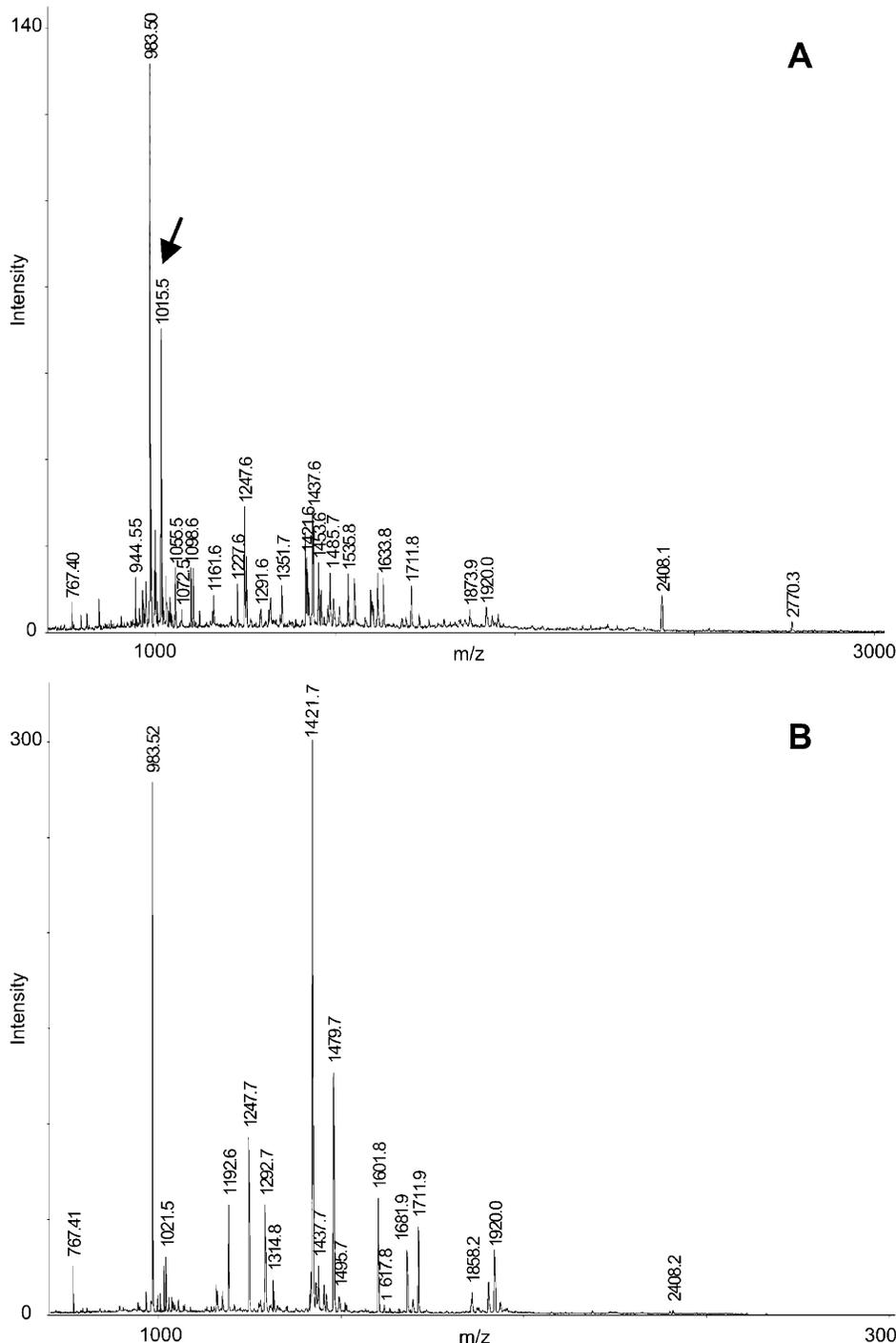


Figure 4. MALDI-TOF spectra of spots SSP 2907 (panel A) and 2403 (panel B) corresponding to Lhcb1 protein in aggregated or native status, respectively. The arrow indicates the new tryptic peptide subjected to MS/MS analysis.

protein contains many Lys and Arg residues, and the former are often consumed in reactions with various photooxidation products, so this might result in a decrease in overall positive charge.

Decreased intensity of staining with Coomassie blue and decreased mobility of protein bands in SDS-PAGE are phenomena usually observed when proteins are exposed to active oxygen species.²⁷ The smearing and mobility shift of intrinsic protein bands, aggregate formation, and also protein fragmentation could be considered to be caused by active oxygen species generated during illumination. To verify this hypothesis,

three small volumes of thylakoid membrane suspension were incubated for 2 h in complete darkness or in low-intensity light, either in the presence or in the absence of ascorbate, a radical scavenger. At the end of the incubation, the spin trap DMPO (5,5-dimethyl-1-pyrroline *N*-oxide) was added at a final concentration of 80 mM, and ESR spectra were measured with a Bruker ESP300 spectrometer. Figure 3 shows the ESR spectra of the three samples. DMPO is an efficient spin trap for hydroxyl (HO^\bullet) and alkyl radicals (R^\bullet), and the different stable adducts can be easily distinguished due to the characteristic four-line (DMPO-HO^\bullet) and a six-line (DMPO-R^\bullet) spectra. After

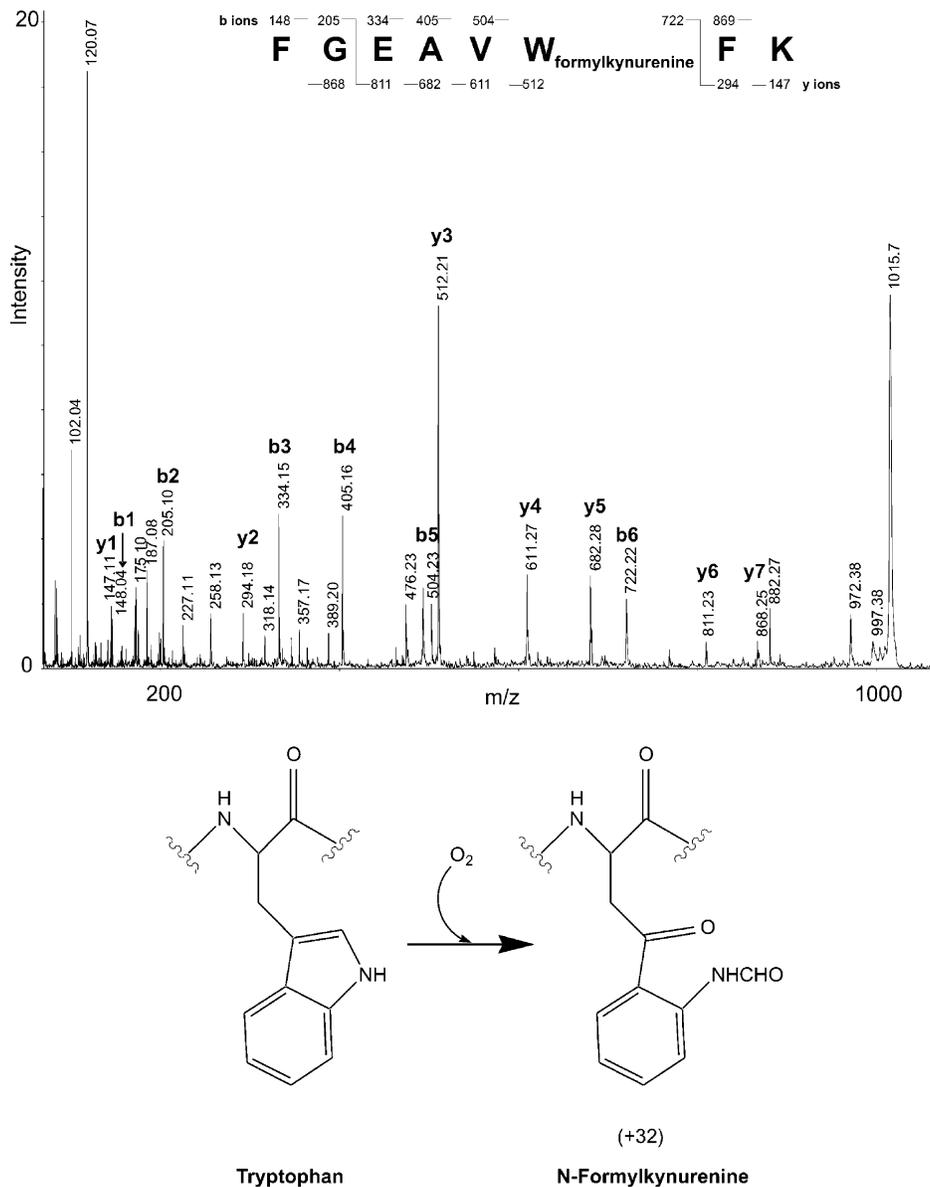


Figure 5. MALDI-TOF/TOF tandem mass spectrum of the mono-charged ion at m/z 1015.5. The masses of the predicted fragment ions are shown at the top. The lower drawings display the chemical structure of tryptophan and *N*-formylkynurenine residues.

2 h of light exposure, the presence of both these adducts can be clearly observed in the samples, whereas, when the aspecific scavenger ascorbate is added, the signal intensity is greatly reduced. In agreement with these observations, the smearing and mobility shift of protein bands were greatly suppressed by active oxygen species scavengers such as *n*-propyl gallate and histidine (data not shown). Thus, soft illumination of thylakoid membranes induces the formation of active oxygen radicals. However, to verify that protein fragmentation or aggregation is really related to active oxygen radicals attacks of native protein, we searched in light-induced protein products for the presence of typical post-translational modifications caused by ROS. In Figure 4, the MALDI spectrum of spots nos. 2907 and 2403 is, respectively, shown. A new LHCII tryptic peptide at m/z 1015.5 was found in the spot 2907 corresponding to an aggregation product. From MALDI molecular mass analysis, this peptide was predicted to be FGEAVWFK (m/z 983.50), with a post-translational modification. The measured shift of 32 Da could correspond to an addition of O_2 , and to

verify this hypothesis, the fragmentation pattern was analyzed. Figure 5 shows the MALDI-TOF/TOF tandem mass spectrum of the mono-charged ion at m/z 1015.5. The y_1 (m/z 147) and y_2 (m/z 294) fragment ions from this peptide were identical to those measured in the MS/MS spectrum of the unmodified peptide (data not shown), whereas ions from y_3 (m/z 512) to y_7 (m/z 868) were shifted by 32 Da higher than the masses of the original peptide fragments. This evidence indicates that the post-translational modification is located in the portion of the peptide containing the initial sequence FGEAVW. A comparison of the b ion masses between unmodified and oxidized peptide shows that only b_6 (m/z 722) and b_7 (m/z 869) are increased by 32 Da upon illumination. The observed m/z increments are consistent with the replacement of tryptophan residue at position 6 with *N*-formylkynurenine (NFK). Mascot search resulted in the identification of this post-translationally modified peptide with an ion score of 53. The lower part of Figure 5 shows the chemical structure of tryptophan and *N*-formylkynurenine. That the aromatic amino acid residues of proteins are prime

targets for oxidation by various forms of ROS is well-documented. In particular, phenylalanine residues are oxidized to *ortho*- and *meta*-tyrosine derivatives;^{28,29} tyrosine residues are converted to the 3,4-dihydroxy (dopa) derivative,^{28,30,31} and also to bi-tyrosine cross-linked derivatives.^{29,32,33} Finally, it has been reported that tryptophan residues are converted to either the 2-, 4-, 5-, 6-, or 7- hydroxy derivatives, and also to *N*-formylkynurenine and kynurenine.^{34,35} Recently, Anderson and co-workers described oxidation of a critical conserved tryptophan residue in luminal loop E of the chloroplast photosystem II protein CP43, providing the first example of selective modification *in vivo*.³⁶ The same authors reported evidence for a ROS-induced post-translational modification (aspartyl aldehyde) in CP47 protein too.³⁷

Thus, it is evident that several photosynthetic proteins had been attacked by ROS produced from excited chlorophylls,³⁸ through an initial oxidation of amino acid residues with consequent protein fragmentation and/or aggregation phenomena.

In conclusion, since findings in spinach were consistent with other species (e.g., barley and tomato, data not shown), data reported here should alert researchers performing proteomic studies on complexes containing chlorophyll-binding proteins that they should work in the dark until complete removal of chlorophylls and should take into account the possible presence of truncated proteins in data interpretation. In previous papers,^{39–41} in fact, researchers using monodimensional SDS-PAGE observed more than one band of antenna proteins, which were identified by antibody and interpreted as isoforms of Lhcb1 or Lhcb2 antenna. From the data reported here, if only few amino acids are removed from the NH₂ terminal region upon illumination, the epitope is still conserved and, obviously, the truncated protein is still recognized by the antibody. Thus, it might have been erroneously identified as a new protein or a post-translationally modified protein. On the contrary, considering that the protein cleavage induced by ROS takes place randomly,¹⁵ when amino acids belonging to the epitope are removed, only by MS/MS is it possible to recognize that some apparently new proteins are only truncated forms, fragments, or aggregates of native proteins.

Abbreviations: ACN, acetonitrile; BBY, photosystem II membranes; DM, *n*-dodecyl- β -D-maltoside; ESI-MS/MS, electrospray ionization-tandem mass spectrometry; ESR, electron spin resonance; FA, formic acid; FtsH, an ATP-dependent metalloprotease; IEF, isoelectric focusing; LHC, light-harvesting complex; MALDI, matrix-assisted laser desorption ionization; PAGE, polyacrylamide gel electrophoresis; PS, Photosystem; ROS, reactive oxygen species; RP-HPLC, reversed-phase-high performance liquid chromatography; SDS, sodium dodecyl sulfate; TOF, time-of-flight.

Acknowledgment. We wish to thank Corrado Ciambella, Gian Maria D'Amici, and Anna Maria Timperio from Tuscia University and all in the Protein Research Group from University of Southern Denmark for their precious help. This work is supported by CIB (Consorzio Interuniversitario per le Biotecnologie), PRIN 2005 (MURST), FIRB 2001 (No. RBNE01KJHT), PRIN 2003 (MURST, Rome), Fondazione Cassa di Risparmio di Verona (Bando 2004), and the Danish Research Agency for funds to the Danish Biotechnology Instrument Center.

References

- Anderson, J. M.; Andersson, B. *Trends Biochem. Sci.* **1988**, *13*, 351–355.
- Melis, A. *Biochim. Biophys. Acta* **1991**, *1058*, 87–106.
- Melis, A. *Adv. Photosynth.* **1996**, *4*, 523–538.
- Zolla, L.; Rinalducci, S.; Timperio, A. M. In *Recent Research Developments in Bioenergetics*; Pandalay, S. G., Ed.; Transworld Research Network: Trivandrum, IN, 2000; Vol. 1, pp 9–32.
- Bassi, R.; Sandonà, D.; Croce, R. *Physiol. Plant.* **1997**, *100*, 769–779.
- Andersson, B.; Aro, E.-M. *Physiol. Plant.* **1997**, *100*, 780–793.
- Björkman, O.; Demmig-Adams, B. In *Ecological Studies*; Schulze, E. D., Caldwell, M., Eds.; Springer: New York, 1994; Vol. 100, pp 17–47.
- Sieffermann-Harms, D. *Physiol. Plant.* **1987**, *69*, 561–568.
- Durrant, J. R.; Giorgi, L. B.; Barber, J.; Klug, D. R.; Porter, G. *Biochim. Biophys. Acta* **1990**, *1017*, 167–175.
- Krieger-Liszskay, A. *J. Exp. Bot.* **2005**, *56*, 337–346.
- Shanklin, J.; DeWitt, N. D.; Flanagan, J. M. *Plant Cell* **1995**, *7*, 1713–1722.
- Itzhaki, H.; Naveh, L.; Lindahl, M.; Cook, M.; Adam, Z. *J. Biol. Chem.* **1998**, *273*, 7094–7098.
- Lensch, M.; Herrmann, R. G.; Sokolenko, A. *J. Biol. Chem.* **2001**, *276*, 33645–33651.
- Lindahl, M.; Spetea, C.; Hundal, T.; Oppenheim, A. B.; Adam, Z.; Andersson, B. *Plant Cell* **2000**, *12*, 419–431.
- Zolla, L.; Rinalducci, S. *Biochemistry* **2002**, *41*, 14391–14402.
- Porra, R. J.; Thompson, W. A.; Kriedemann, P. E. *Biochim. Biophys. Acta* **1989**, *975*, 384–394.
- Hippler, M.; Klein, J.; Fink, A.; Allinger, T.; Hoerth, P. *Plant J.* **2001**, *28*, 595–606.
- Herbert, B.; Galvani, M.; Hamdan, M.; Olivieri, E.; MacCarthy, J.; Pedersen, S.; Righetti, P. G. *Electrophoresis* **2001**, *22*, 2046–2057.
- Righetti, P. G.; Stoyanov, A. V.; Zhukov, M. Y. *The Proteome Revisited: Theory and Practice of All Relevant Electrophoretic Steps*; Elsevier: Amsterdam, The Netherlands, 2001; pp 1–395.
- Righetti, P. G. *Immobilized pH Gradients: Theory and Methodology*; Elsevier: Amsterdam, The Netherlands, 1990.
- Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. *Anal. Chem.* **1996**, *68*, 850–858.
- Gobom, J.; Nordhoff, E.; Mirgorodskaya, E.; Ekman, R.; Roepstorff, P. *J. Mass Spectrom.* **1999**, *34*, 105–116.
- Henmi, T.; Miyao, M.; Yamamoto, Y. *Plant Cell Physiol.* **2004**, *45*, 243–250.
- Sonoike, K.; Terashima, I. *Planta* **1994**, *194*, 287–293.
- Sonoike, K. *Plant Sci.* **1996**, *115*, 157–164.
- Davies, M. J. *Biochem. Biophys. Res. Commun.* **2003**, *305*, 761–770.
- Davies, K. J. A. *J. Biol. Chem.* **1987**, *262*, 9895–9901.
- Maskos, Z.; Rush, J. D.; Koppenol, W. H. *Arch. Biochem. Biophys.* **1992**, *296*, 521–529.
- Wells-Knecht, M. C.; Huggins, T. G.; Dyer, S. R.; Thorpe, S. R.; Baynes, J. W. *J. Biol. Chem.* **1993**, *269*, 12348–12353.
- Davies, K. J. A.; Delsignore, M. E.; Lin, S. W. *J. Biol. Chem.* **1987**, *262*, 9902–9907.
- Dean, R. T.; Geiseg, S.; Davies, M. J. *Trends Biochem. Sci.* **1993**, *18*, 437–441.
- Huggins, T. G.; Wells-Knecht, M. C.; Detorie, N. A.; Baynes, J. W.; Thorpe, S. R. *J. Biol. Chem.* **1993**, *268*, 12341–12347.
- Heinecke, J. W.; Li, W.; Daehnke, H. L., III; Goldstein, J. A. *J. Biol. Chem.* **1993**, *268*, 4069–4077.
- Winchester, R. V.; Lynn, K. R. *Int. J. Radiat. Biol.* **1970**, *17*, 541–549.
- Kikugawa, K.; Kato, T.; Okamoto, Y. *Free Radical Biol. Med.* **1994**, *16*, 373–382.
- Anderson, L. B.; Maderia, M.; Ouellette, A. J.; Putnam-Evans, C.; Higgins, L.; Krick, T.; MacCoss, M. J.; Lim, H.; Yates, J. R., III; Barry, B. A. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 14676–14681.
- Anderson, L. B.; Ouellette, A. J.; Eaton-Rye, J.; Maderia, M.; MacCoss, M. J.; Yates, J. R., III; Barry, B. A. *J. Am. Chem. Soc.* **2004**, *126*, 8399–8405.
- Rinalducci, S.; Pedersen, J. Z.; Zolla, L. *Biochim. Biophys. Acta* **2004**, *1608*, 63–73.
- Sigrist, M.; Staehelin, L. A. *Biochim. Biophys. Acta* **1992**, *1098*, 191–200.
- Sigrist, M.; Staehelin, L. A. *Plant Physiol.* **1994**, *104*, 135–145.
- Plumley, F. G.; Schmidt, G. W. *Plant Cell* **1995**, *7*, 689–704.

PR0502368

Ringraziamenti

Desidero ringraziare.....

il Prof. Pier Giorgio Righetti, per avermi seguito durante questo progetto di dottorato, per i suoi preziosi consigli e per il suo contagioso entusiasmo,

il Dottor Mirco Ponzoni, il Dottor Fabio Pastorino e il Dottor Danilo Marimpietri, del Laboratorio di Oncologia dell'Ospedale Gaslini di Genova, per le stimolanti collaborazioni,

il Prof. Emilio Marengo, la Dottoressa Elisa Robotti, il Prof. Francesco Dondi e la Prof.ssa MariaChiara Pietrogrande per il prezioso contributo dato a questo progetto di dottorato,

"la banda dei massisti" il cui contributo è stato di fondamentale importanza (Dottor Mahmoud Amdan, Dottor Hubert Aster, Dottor Juri Rappsilber, Dott.ssa Liliana B. Arces, Dottor Gian Maria Fimia, Dottor Carmine Mancone)

Un grazie particolare a.....

Annalisa, Daniela, Alessandra, Francesca, Rita, Paolo, Alberto, Giacomo per avermi dimostrato una sincera amicizia,

Laura, Erna, Barbara, Francesca, Chiara, Mayte.....memorie del laboratorio,

gli amici della compagnia, per essermi stati vicini rendendomi la vita più felice,

la mia nonna Edvige, che ha sempre una parola giusta per tutti,

Damiano, che mi sostiene sempre con amore e pazienza,

la mia famiglia, punto di riferimento della mia vita, che non smette mai di sostenermi ed il cui entusiasmo ed energia mi sono sempre stati d'esempio.