

SCUOLA DI DOTTORATO DI RICERCA
IN BIOTECNOLOGIE INDUSTRIALI - AMBIENTALI
XVII CICLO

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

PhD Student DANIELA CECCONI

**COMPARATIVE PROTEOMICS ANALYSIS:
INNOVATIVE APPROACHES
AND BIOMEDICAL APPLICATIONS**

Course coordinator: Prof. HUGO MONACO

Supervisor: Prof. PIER GIORGIO RIGHETTI

ANNI ACCADEMICI 2002-2005

Ai miei genitori,

Il silenzio è una discussione portata avanti con altri mezzi.

ERNESTO CHE GUEVARA

RIASSUNTO:

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 delle collaborazioni con alcuni laboratori: in particolare si è collaborato con il laboratorio di Molecular Medicine del Centro Ricerche GlaxoSmithKline (GSK) di Verona, per la parte del progetto riguardante l'alchilazione differenziale con acrilamide deuterata, lo studio della dipendenza da nicotina e l'analisi proteomica dei cambiamenti molecolari indotti dall'antidepressivo fluoxetina. Si è collaborato inoltre con i Dipartimenti di Patologia e di Scienze Neurologiche e della Visione dell'Università di Verona per lo studio riguardante l'analisi degli eventi epigenetici coinvolti nell'adenocarcinoma pancreatico. La collaborazione con il Laboratorio di Mass Spectrometry del Centro Ricerche (GSK) di Verona ha permesso la parte di identificazione proteica delle analisi sopra-indicate; mentre lo sviluppo di nuovi metodi statistici 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 ed il Dipartimento di Chimica e Tecnologie Farmaceutiche e Alimentari dell'Università di Genova.

Gli obiettivi di questo progetto di dottorato sono stati:

- 1) lo sviluppo di metodi innovativi per l'analisi proteomica comparativa tramite elettroforesi bidimensionale (2DE);
- 2) l'applicazione dei metodi sviluppati e soprattutto dei metodi classici della proteomica comparativa, per l'analisi degli eventi epigenetici coinvolti nell'adenocarcinoma pancreatico, per lo studio della dipendenza da nicotina e degli eventi molecolari indotti dall'antidepressivo fluoxetina.

Per quanto riguarda il primo obiettivo del progetto di dottorato, si è proceduto con lo sviluppo di un nuovo metodo semplice ed economicamente competitivo, (rispetto alle tecnologie attualmente presenti in commercio, quali ICAT, MCAIT), per la determinazione delle quantità relative delle proteine separate tramite elettroforesi bidimensionali. Il metodo sviluppato si basa sull'*alchilazione differenziale* delle due miscele proteiche da confrontare tramite l'utilizzo di acrilammide deuterata e non (d_0 e d_3). Le due miscele, mescolate in rapporto 1:1, sono separate sulla stessa mappa 2D e le macchie proteiche vengono quindi analizzate e quantificate tramite spettrometria di massa. I risultati ottenuti per una miscela di proteine standards e per un campione reale (quale il siero di ratto) hanno dimostrato che il metodo sviluppato risulta essere uno strumento semplice ed efficace da poter

applicare per l'analisi proteomica comparativa. Il vantaggio offerto da questa tecnica rispetto ai metodi già esistenti, basati sulla separazione tramite LC/MS-MS, dipende dal fatto che vengono mantenute le caratteristiche positive di una separazione elettroforetica bidimensionale 2DE (cioè visualizzazione del profilo proteico generale e di eventuali modificazioni post-trascrizionali, nonché la possibilità di analizzare tramite la MS le proteine intere). Inoltre in uno studio condotto da Cahill *et al.* (*Rapid Commun. Mass Spectrom.*, 2003, 17, 1283-1290), successivamente alla pubblicazione del nostro metodo, è stato dimostrato che la diversa composizione isotopica delle acrilammidi da noi utilizzate (d_0 e d_3) non influenza la resa di alchilazione e nemmeno la corsa elettroforetica delle proteine. Risulta quindi prevedibile che il metodo di alchilazione differenziale messo a punto potrà trovare ampia applicazione nei laboratori di analisi proteomica differenziale: in uno studio eseguito da Turko e Murad (*The J. Biol. Chem.*, 2003, 278, 35844-35849), per esempio, il metodo è stato applicato per analizzare la variazione dell'espressione proteica in un modello di diabete ottenuto sperimentalmente con streptozotocina.

Sempre allo scopo di migliorare i metodi attualmente disponibili per l'analisi comparativa delle mappe bidimensionali si è proceduto, durante il progetto di dottorato, con l'implementazione di nuovi metodi statistici per il confronto e l'analisi dei profili proteici. Per ovviare all'inconveniente di scarsa riproducibilità delle mappe bidimensionali è stata applicata la *logica fuzzy* per la loro analisi. La logica fuzzy si basa sull'insieme fuzzy, cioè un insieme di oggetti nel quale non c'è un confine ben preciso o definito tra gli oggetti che vi appartengono e quelli che non vi appartengono. Il concetto chiave che sta alla base di tale definizione è quello di appartenenza: ad ogni elemento di un insieme è associato un valore che indica il grado di appartenenza di tale elemento all'insieme. Questo valore è compreso nell'intervallo $[0, 1]$, dove 0 e 1 indicano rispettivamente il nullo ed il massimo grado di appartenenza, mentre tutti i valori intermedi indicano delle appartenenze "parziali". Suddividendo la mappa bidimensionale in una griglia, si può indicare con 1 o 0 l'appartenenza o meno di una macchia proteica in una determinata zona della griglia. I risultati ottenuti tramite l'applicazione della logica fuzzy hanno dimostrato che è possibile suddividere correttamente 10 diverse mappe bidimensionali nei due gruppi sperimentali di appartenenza (campioni controllo e campioni trattati). Un ulteriore nuovo strumento statistico per la proteomica comparativa, sviluppato in questo progetto di dottorato, è l'*analisi dei componenti principali 3-way* (PCA). Tale strumento statistico è stato applicato allo studio di 10 diverse mappe bidimensionali di campioni di siero di ratto (5 di controllo e 5 trattati con nicotina). Il principale vantaggio offerto dall'analisi dei componenti principali 3-way (PCA) dipende dal fatto che, a differenza dei programmi di elaborazione attualmente disponibili

(quali PDQuest, Melanie, etc...), non viene richiesto l'intervento dell'operatore per l'allineamento ed il confronto delle mappe. I risultati ottenuti hanno dimostrato che l'analisi con PCA è un buon metodo per classificare correttamente le mappe 2D nei reali gruppi sperimentali e permette inoltre di individuare le zone delle mappe (e quindi le macchie proteiche) responsabili delle differenze tra i due gruppi in analisi.

Procedendo quindi con il secondo obiettivo del progetto di dottorato, alcuni dei metodi sviluppati (sopra-descritti) e soprattutto i metodi classici della proteomica comparativa, sono stati applicati per *l'analisi degli eventi epigenetici coinvolti nell'adenocarcinoma pancreatico*. I primi stadi nello sviluppo di un tumore coinvolgono molti cambiamenti molecolari, tra i quali le variazioni epigenetiche. Tali eventi, a differenza delle variazioni genetiche (come le mutazioni), possono essere soppressi dall'azione farmacologica di particolari sostanze antitumorali. Con lo scopo di studiare la deacetilazione degli istoni e la metilazione del DNA, quali principali eventi epigenetici coinvolti nell'adenocarcinoma pancreatico, abbiamo analizzato la variazione nell'espressione proteica di una linea cellulare (di adenocarcinoma pancreatico) trattata con tricostatina-A, (TSA) (inibitore della deacetilazione degli istoni) e, successivamente, con decitabina (DAC) (inibitore della metilazione del DNA). L'analisi proteomica differenziale classica, effettuata tramite confronto delle repliche delle mappe 2D con il software PDQuest, ha permesso di individuare 51 diverse macchie proteiche, che risultano avere una variazione nella loro espressione in seguito al trattamento delle cellule con tricostatina-A. I risultati ottenuti tramite l'analisi proteomica differenziale classica, abbinata all'identificazione proteica con MALDI-TOF, hanno permesso di migliorare la comprensione degli effetti del trattamento con il farmaco antitumorale TSA e dell'evento epigenetico connesso. Ad esempio, la nucleofosmina e la proteina tumorale trascrizionalmente regolata, (coinvolte rispettivamente nell'oncogenesi e nel processo di reversione del tumore), risultano essere sotto-esprese, mentre la proteina di morte cellulare programmata e la statmina, (coinvolte rispettivamente nell'apoptosi e nel blocco del ciclo cellulare), risultano essere sovra-esprese dopo trattamento con TSA, in perfetto accordo con gli effetti farmacologici di tale sostanza anti-tumorale.

Per verificare la veridicità dei risultati ottenuti dal confronto dei profili proteici effettuato con il PDQuest, si è proceduto con l'applicazione dell'analisi dei componenti principali (PCA). I dati ricavati da tale metodo statistico non solo erano in ottimo accordo con quelli derivati dall'analisi al PDQuest ma, essendo la PCA un metodo statistico robusto, hanno anche permesso di individuare

variazioni nell'espressione proteica inferiori al livello di soglia del 100%, tipico dell'analisi al PDQuest.

Si è proceduto quindi con l'analisi proteomica classica dei campioni cellulari trattati con decitabina, allo scopo di individuare gli effetti molecolari di tale trattamento anti-tumorale: un totale di 45 diverse macchie proteiche sono risultate essere diversamente espresse. Tra le 36 proteine identificate, tramite analisi MALDI-TOF, alcune sono note in letteratura come marcatori del carcinoma epatico, altre sono coinvolte nell'apoptosi, altre ancora sono chaperoni molecolari, mentre una proteina (la super ossido dismutasi) è un soppressore del tumore, la cui trascrizione viene riattivata in seguito al trattamento con DAC. I risultati ottenuti da tale analisi possono contribuire a migliorare la comprensione del meccanismo coinvolto nella risposta al trattamento con l'antitumorale decitabina.

Il progetto di dottorato si è sviluppato poi con *l'analisi dei cambiamenti proteomici nel siero di ratto dopo trattamento cronico con nicotina*. Il meccanismo d'azione della nicotina non è stato ancora del tutto chiarito né a livello del sistema nervoso centrale, né a livello periferico; soprattutto a livello periferico rimane ancora molto da chiarire. È noto ad esempio che la nicotina agisce sulla tachicardia, sulla vasocostrizione, sulla motilità intestinale, inoltre ha effetti aterogenici se somministrata in modo acuto ed agisce sul sistema immunitario. I risultati ottenuti dall'analisi proteomica dei campioni di siero di ratti, trattati con nicotina, confermano le proprietà anti-infiammatorie di tale sostanza ed un coinvolgimento nella risposta di fase acuta e nello stress ossidativo. Questi effetti osservati a livello periferico nel siero, possono derivare dall'interazione diretta della nicotina con i recettori specifici neuronali e periferici o direttamente dalle proprietà chimico-farmacologiche della nicotina stessa.

Si sono valutati infine gli *effetti molecolari del trattamento con l'antidepressivo fluoxetina* su colture primarie di neuroni corticali. Anche per tale sostanza il meccanismo d'azione rimane ancora da chiarire: è noto infatti che l'antidepressivo fluoxetina agisce come inibitore selettivo della ricaptazione della serotonina, tuttavia gli effetti farmacologici della sostanza sono evidenziabili solo dopo alcune settimane di trattamento, benché vi sia un innalzamento immediato dei livelli sinaptici di serotonina; è ipotizzabile quindi un meccanismo d'azione indipendente dalla trasmissione serotoninergica. Abbiamo utilizzato un approccio di analisi proteomica comparativa con lo scopo di migliorare la comprensione del meccanismo d'azione della fluoxetina ed i risultati ottenuti ci hanno permesso di stabilire che tale antidepressivo altera l'espressione di proteine coinvolte nel trasporto assonale, nell'assemblamento sinaptico delle vescicole e nella neuroprotezione. Tali risultati, in accordo con la letteratura già esistente, oltre a chiarire il meccanismo d'azione della fluoxetina, possono suggerire nuovi bersagli per lo sviluppo di strategie addizionali per il trattamento della depressione.

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 l'individuazione di marcatori molecolari di malattie nonché di bersagli per nuove strategie terapeutiche. L'analisi proteomica comparativa effettuata con 2DE, pur essendo una tecnica robusta, ormai consolidata, offre comunque la possibilità di introdurre ancora miglioramenti, riguardanti non solo la derivatizzazione differenziale dei campioni, o l'analisi statistica dei profili 2D, ma anche altri aspetti dell'analisi bidimensionale (dalla riproducibilità delle mappe 2D, fino alle banche dati proteiche).

PREFACE:

Conventional protein analyses, based primarily on immunological methods, have been successfully used for studying proteins for many years. However, these methods can be used for characterizing only a small number of proteins in replicates of single experiments and are frequently limited by the availability of specific antibodies to proteins. *Proteomics* is an emerging area of science that attempts to study proteins on a massively parallel scale. The field of proteomics is relatively new as demonstrated by the fact that Marc Wilkins first introduced the term *proteome* (set of PROTEins expressed by the genOME) to the public as recently as in 1994. Quantitative and qualitative proteomic analyses can be performed using various gel-based and gel-free methods. In general, mass spectrometry is used to identify the proteins by either direct sequence analysis or pattern matching of peptide fragments. Using state-of-the-art proteomic techniques, a large number of proteins can be studied at once and analyses of thousands of proteins can be completed within a few months. The application of proteomic approaches is growing rapidly as demonstrated by the increasing number of published articles that use proteomic techniques. Proteomic technologies are thus applied to plants, micro organisms, food, animal and human samples analysis. Concerning this last field, proteomics is became a key technology in many biomedical studies, including molecular medicine, drug discovery and clinical diagnostic. All these different applications and the related problems to be solved, make proteomics a technology that may be improved.

The objectives of the present thesis work are:

- 1) the development of new tools for comparative proteomics analysis,
- 2) the biomedical application of classical proteomic analysis, and some of the new developed tools, for analyzing epigenetic events in pancreatic cancer, biochemical pathways involved in nicotine dependence and molecular events induced by the antidepressant fluoxetine treatment.

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 Molecular Medicine Unit at the Psychiatric Centre of Excellence in Drug Discovery of GlaxoSmithKline (GSK) S.p.A. in Verona, for the development of a new isotopically marked alkylating agent, the nicotine dependence analysis and the proteomic study of fluoxetine-induced molecular changes. The proteomic analysis of epigenetic events involved in pancreatic cancer was performed in collaboration with the Department of Pathology, Section of Anatomical Pathology, and the Department of Neurological and Visual Sciences, Section of Biochemistry, of the University of Verona. The

collaboration with the Mass Spectrometry laboratory of Computational, Analytical & Structural Sciences, at the GlaxoSmithKline in Verona, permitted the identification of proteins for all the analysis above mentioned; while the development of new statistical tools for comparative proteomics analysis it was possible thanks to the collaboration with Department of Environmental and Life Sciences, of the University of Eastern Piedmont, and the Department of Chemistry and Pharmaceutical and Alimentary Technologies of the University of Genoa.

The results thus obtained are here discussed and evaluated.

TABLE of CONTENTS

Riassunto		I
Preface		VII
Table of contents		IX
List of abbreviations and Symbols		XIII
CHAPTER 1: INTRODUCTION TO THE PROTEOME AND PROTEOMICS		
1.1	General introduction	1
1.2	Proteome analysis by 2DE gel electrophoresis and mass spectrometry	2
	1.2.1 Basic methodology	2
	1.2.2 Advantages of 2DE-based approaches	7
1.3	Quantitative proteome analysis	9
	1.3.1 Traditional approach (2DE/MS)	9
	1.3.2 The role of isotopes for quantitation	10
	1.3.2.1 <i>In vivo</i> labeling	11
	1.3.2.2 <i>In vitro</i> pre-digestion labeling	11
	1.3.2.3 <i>In vitro</i> post-digestion labeling	11
1.4	Proteomics in disease studies	13
	1.4.1 Proteomics in cancer research	14
	1.4.2 Proteomics in drug dependence and psychiatric disorders	15
1.5	Bibliography	17
CHAPTER 2: COMPARATIVE PROTEOMICS: A NEW ISOTOPICALLY MARKED ALKYLATING AGENT		
2.1	Introduction	19
2.2	Experimental procedures	20
	2.2.1 Labeling a standard proteins mixture	20
	2.2.2 Preparation and labeling of rat serum	20
	2.2.3 Two-dimensional electrophoresis	20
	2.2.4 Protein identification and quantitation	21
2.3	Results	22
	2.3.1 Accuracy of quantification	23
	2.3.2 Quantification of changes of proteins abundance in rat serum	26
2.4	Discussion	28
2.5	Concluding remarks	29
2.6	Bibliography	31

CHAPTER 3: NEW STATISTICAL TOOLS FOR COMPARATIVE PROTEOMICS (I): SPOT FUZZYFICATION AND MULTIDIMENSIONAL SCALING

3.1	Introduction	33
3.2	Theory	34
	3.2.1 Digitalisation	35
	3.2.2 Fuzzyfication	35
	3.2.3 Similarity index	39
	3.2.4 Multidimensional scaling	39
	3.2.5 Statistical analysis	40
3.3	Experimental procedures	40
	3.3.1 Simulated samples	40
	3.3.2 Real samples	42
3.4	Results and discussion	44
3.5	Concluding remarks	50
3.6	Bibliography	51

CHAPTER 4: NEW STATISTICAL TOOLS FOR COMPARATIVE PROTEOMICS (II): THREE WAY PRINCIPAL COMPONENT ANALYSIS

4.1	Introduction	53
4.2	Theory	54
	4.2.1 Digitalisation	54
	4.2.2 Data transformation	55
	4.2.3 Three-way principal component analysis	55
4.3	Experimental procedures	56
	4.3.1 Sample preparation	56
4.4	Results and discussion	58
	4.4.1 Data transformation	58
	4.4.2 Three-way PCA on the normalised dataset	59
	4.4.3 Difference analysis	61
4.5	Concluding remarks	63
4.6	Bibliography	64

CHAPTER 5: PANCREATIC ADENOCARCINOMA: PROTEOMICS ANALYSIS OF EPIGENETIC EVENTS

5.1	Pancreatic adenocarcinoma	65
5.2	Epigenetic regulation of genes	68
	5.2.1 Histone deacetylation	69
	5.2.2 DNA methylation	70
5.3	Histone deacetylation (I): proteomic profiling of pancreatic ductal carcinoma cell lines treated with trichostatin-A	72
	5.3.1 Introduction	72
	5.3.2 Experimental procedures	73
	5.3.2.1 Cells and growth conditions	73

	5.3.2.2	Cell proliferation assay	73
	5.3.2.3	Cell cycle analysis	73
	5.3.2.4	Apoptosis detection	73
	5.3.2.5	Two-dimensional gel electrophoresis	74
	5.3.2.6	Protein pattern analysis	74
	5.3.2.7	Protein identification by mass spectrometry	75
	5.3.3	Results	75
	5.3.4	Discussion	80
5.4		Histone deacetylation (II): Multivariate statistical analysis of proteomic profiling of pancreatic ductal carcinoma cell lines treated with trichostatin-A	83
	5.4.1	Introduction	83
	5.4.2	Theory	83
	5.4.2.1	Principal Component Analysis	83
	5.4.2.2	Cluster Analysis	84
	5.4.3	Experimental procedures	84
	5.4.3.1	Softwares	85
	5.4.3.2	Cell treatment with TSA	86
	5.4.3.3	Cell lysis	86
	5.4.3.4	Two-dimensional gel electrophoresis	86
	5.4.3.5	Protein identification by mass spectrometry	86
	5.4.4	Results and discussion	86
	5.4.4.1	Protein pattern analysis with the PDQuest software	86
	5.4.4.2	Principal Component Analysis	88
	5.4.4.3	Cluster Analysis	95
	5.4.4.4	Mass-spectrometry	96
	5.4.4.5	Biological significance of some interesting identified proteins	97
	5.4.5	Concluding remarks	98
5.5		DNA methylation: proteomic profiling of pancreatic ductal carcinoma cell lines treated with 5-aza-2'-deoxycytidine	99
	5.5.1	Introduction	99
	5.5.2	Experimental procedures	99
	5.5.2.1	Cells and growth conditions	99
	5.5.2.2	Cell proliferation assay	100
	5.5.2.3	Treatment of cells for proteomic analysis and cell lysis	100
	5.5.2.4	Two-dimensional gel electrophoresis	100
	5.5.2.5	Protein pattern analysis	101
	5.5.2.6	Protein identification by mass spectrometry	101
	5.5.3	Results	102
	5.5.4	Discussion	109
	5.5.4.1	Silenced proteins	110
	5.5.4.2	Down-regulated proteins	110
	5.5.4.3	Up-regulated proteins	112
	5.5.5	Concluding remarks	112
5.6		Bibliography	113

CHAPTER 6: DRUG DEPENDENCE: PROTEOMIC CHANGES IN RAT SERUM AFTER CHRONIC NICOTINE TREATMENT

6.1	Introduction	119
6.2	Experimental procedures	122
	6.2.1 Animals and nicotine treatment	122
	6.2.2 Serum preparation	123
	6.2.3 Serum 2-D electrophoresis	123
	6.2.4 Protein pattern differential analysis	124
	6.2.5 MALDI-TOF mass spectrometry	124
6.3	Results	126
	6.3.1 Differential proteomic analysis	126
	6.3.2 Identification of proteins	127
6.4	Discussion	129
6.5	Bibliography	133

CHAPTER 7: PSYCHIATRIC DISORDERS: PROTEOMIC ANALYSIS OF FLUOXETINE-INDUCED MOLECULAR CHANGES IN RAT CORTICAL NEURONAL CULTURES

7.1	Depression and antidepressants	135
7.2	Experimental procedures	137
	7.2.1 Rat cortical neuron primary culture	137
	7.2.2 Treatment with fluoxetine	138
	7.2.3 Immunocytochemistry	138
	7.2.4 Two-dimensional gel electrophoresis	139
	7.2.5 MALDI-TOF mass spectrometry	139
	7.2.6 Semi-quantitative Western blot analysis	140
7.3	Results and discussion	141
	7.3.1 Immunocytochemistry analysis	141
	7.3.2 Differential proteomic analysis	142
	7.3.3 Western blot analysis	145
7.4	Concluding remarks	147
7.5	Bibliography	148

Papers published during the PhD studies

Ringraziamenti

LIST of ABBREVIATIONS and SYMBOLS

Ach	acetylcholine
nAchRs	acetylcholine receptors
ACTH	adrenocorticotrophic hormone
ASF/SF2	alternative splicing factor/splicing factor 2
amu	atomic mass unit
POD	anti-DNA-peroxidase
CNS	central nervous system
CHAPS	(3[(3-cholamidopropyl)dimethylammonium]-1-propane-sulfonate)
CLP	coactosin-like protein
CRH	corticotropin-releasing hormone
CypA	cyclophilin A
DA	dopamine
DAC	5-aza-2'-deoxycytidine
ESI-Q-IT-MS	electrospray quadrupole ion-trap
ER	endoplasmic reticulum
ECL	enhanced chemiluminescence
CNS	central nervous system
CORT	glucocorticoids
fmol	femto-mole
GRP78	glucose-regulated protein 78
Glu	glutamate
HAT	histone acetyltransferases
HDAC	histone deacetylases
HPA	hypothalamic-pituitary-adrenal
IPG	immobilized pH gradient strips
IAA	iodoacetamide
IEF	isoelectric focusing
ICAT	isotope-coded affinity tag
LDT	lateral dorsal tegmentum
LC/MS-MS	liquid chromatography/tandem mass spectrometry
LOH	loss of heterozygosity
MS	mass spectrometry
MCAT	mass-coded abundance tagging
MALDI-ToF	matrix-assisted laser desorption/ionization time-of-flight
Mr	molecular mass
MDS	multidimensional scaling
NEM	N-ethylmaleimide
PanIN	pancreatic intraepithelial neoplasia
PPT	pedunculopontine nucleus
PPIA	peptidyl-prolyl cis-trans isomerase A
PMN	poly-morpho-nuclear
PVDF	polyvinylidene fluoride
PCA	principal component analysis

Rho GDI-2	Rho GDP-dissociation inhibitor
SSRIs	selective serotonin reuptake inhibitors
5-HT	serotonin
SDS-PAGE	sodium dodecil sulfate-polyacrylamide gel electrophoresis
SOD	superoxide dismutase
MS/MS	tandem mass spectrometry
TBP	tributylphosphine
Th	Thomson, unit of m/z
TSA	trichostatin-A
TBS	Tris-buffered saline
2DE	two-dimensional electrophoresis
2D-PAGE	two-dimensional poly-acrylamide gel electrophoresis
VPA	valproate

CHAPTER 1:

INTRODUCTION TO THE PROTEOME AND PROTEOMICS

1.1 GENERAL INTRODUCTION

Proteins are the final products manufactured in living cells according to the 'blueprint' contained in the genome. The proteome is defined as "the PROTEin complement expressed by a genOME", thus it is fusion word derived from two different terms [1, 2]. The proteome represents the array of proteins that are expressed in a biological compartment (cell, tissue, organ) at a particular time, under a particular set of conditions [3]. Because proteins are key structural and functional molecules, the molecular characterization of proteomes is necessary for a complete understanding of biological systems. Large-scale, comprehensive analysis of proteins is the objective of proteome science (proteomics). The scope of proteomics is broad; it encompasses:

- 1) the identification and quantification of proteins in cells, tissues and biological fluids;
- 2) the analysis of changes in protein expression in normal versus diseased cells;
- 3) the characterization of post-translational modifications;
- 4) the studies of protein-protein interactions;
- 5) other general applications.

The goals of proteomics research include: clarification of molecular mechanisms that govern cellular processes; characterization of complex protein networks and their perturbations; discovery of biomarker proteins for detection and diagnosis of diseases; and, identification of targets for the design of drug treatments. The expansion of proteomics was driven by technology. Although the concept of global protein analysis as a complete inventory of human proteins was proposed 20 years ago [4], proteomic research was made possible in the mid-1990s only because of concurrent developments in three areas:

1. Two-dimensional electrophoresis (2DE) has evolved into a robust method to rapidly separate the many proteins contained in a proteome.
2. Mass spectrometry (MS) methods were developed for ready and accurate analysis of 2DE separated proteins with a high degree of sensitivity and specificity.

3. Large-scale genome research has produced a constantly increasing number of sequences that were catalogued in several databases, which could be accessed over the Internet, and search engines and other bioinformatics tools were developed to interrogate these databases.

2DE, MS and bioinformatic tools are the key components of an approach that has been termed “the classical proteomic methodology”.

1.2 PROTEOME ANALYSIS BY 2DE GEL ELECTROPHORESIS AND MASS SPECTROMETRY

1.2.1 BASIC METHODOLOGY

The basic 2DE-based proteomic methodology includes several steps (Fig. 1):

- solubilization of proteins from the sample (e.g. tissue);
- separation of the proteins by 2DE;
- digitization of 2D gels and computer-assisted analysis of protein spot patterns;
- determination of specific attributes of the proteins of interest by MS;
- searching of databases with these attributes for identifying the proteins.

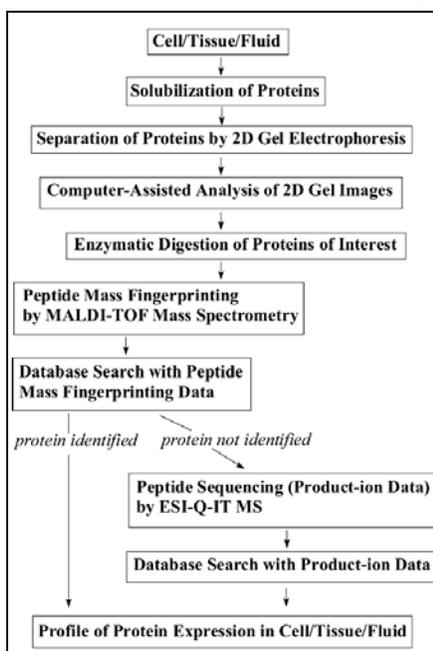


Figure 1. Steps involved in 2DE-based proteomics analysis. (Abbreviations: MALDI-ToF-MS, matrix-assisted laser desorption/ionization time-of-flight MS; ESI-Q-IT-MS, electrospray quadrupole ion-trap MS).

Treatment of samples for 2DE involves cell lysis and solubilization of proteins. Sample preparation is commonly carried out in a solution containing chaotropes, detergents, reducing agents, carrier ampholytes and, depending on the sample, protease inhibitors [5, 6]. An aliquot of the sample solution is subjected to 2DE. 2DE is a powerful separation technique, which allows simultaneous resolution of thousands of proteins. The high-resolution capability of 2DE stems from the fact that the first and second dimensions are based on two independent protein characteristics. The first dimension of 2DE is isoelectric focusing (IEF), during which the proteins are separated based on their charge. In the second dimension, the proteins are separated orthogonally by SDS-PAGE according to their molecular mass (M_r). 2DE was first introduced in the early 1970s [7]. However, its widespread application was hampered by experimental drawbacks and by the lack of techniques capable of analyzing 2DE-separated proteins. The experimental difficulties were overcome by the introduction of immobilized pH gradient for IEF, that are now commercially available as gel strips. Using these standardized gels, it is now possible to separate higher loads of proteins, sufficient for further characterization, and to generate highly-reproducible 2D maps. After separation, proteins in 2D gels are visualized by staining, commonly with a Coomassie Blue dye, Sypro Ruby fluorescent stain, or with a modified silver stain that is compatible with subsequent MS analysis [8]. The 2D gels are 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 gels 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 2DE can be identified based on unique attributes that are measured by MS. These attributes are determined from analysis of peptides generated by proteolytic digestion of the protein of interest. The most commonly used enzyme for protein digestion is trypsin, which cleaves the protein at the C-terminal side of lysine and arginine. If needed, proteases with other specificities can also be employed.

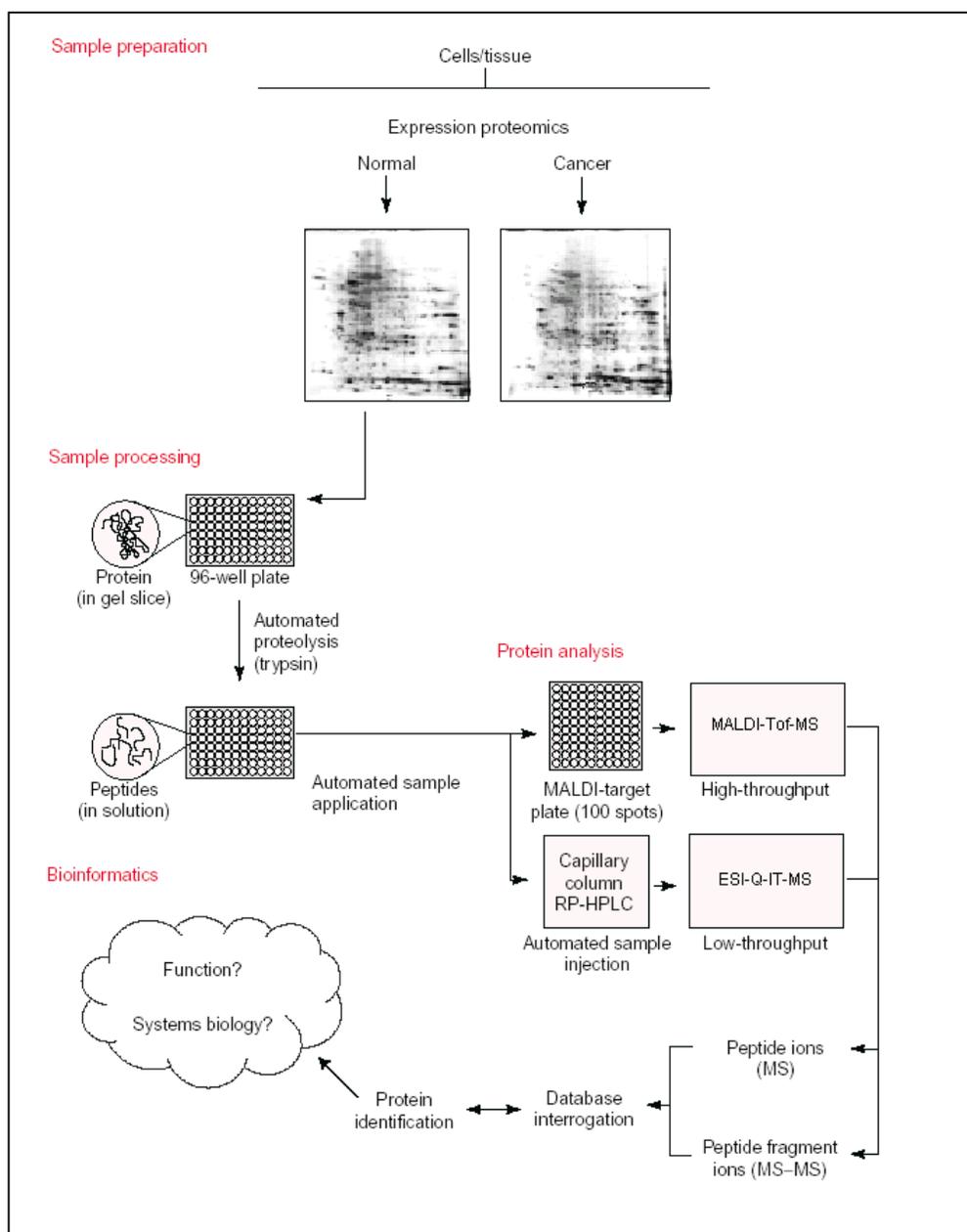


Figure 2. General strategy for proteome analysis by 2DE, MS, and database searching.

Two specific protein attributes can be obtained by MS analyses of proteolytic digests. The first protein attribute is the so-called peptide-mass fingerprint which involves determination of the masses of all peptides in the digest. The second attribute includes fragmentation of selected peptides inside the mass spectrometer into series of sequence-diagnostic product ions. From these product ions, a portion of the amino acid sequence of the peptide (a 'sequence tag') can be deduced;

alternatively, uninterpreted product-ion spectra can be used directly for protein identification. MS instrumentation techniques that have played a key role in proteomics and in the analysis of peptides and proteins in general are: matrix-assisted laser desorption/ionization time-of-flight MS (MALDI-ToF-MS) and electrospray ionization-quadrupole ion trap MS (ESI-Q-IT-MS).

MALDI-ToF-MS and ESI-Q-IT-MS are based on different physicochemical principles and have different characteristics [9]; therefore, they can yield complementary analytical information. These two methods provide the needed sensitivity and specificity for proteomics research. MALDI-ToF-MS is commonly used for peptide-mass fingerprinting. The method is user-friendly, fast (3-5 min per analysis), sensitive (fmol peptide levels), and it measures peptide masses with accuracy better than 50 ppm. However, MALDI-ToF-MS is impractical for the sequencing of peptides in proteolytic digests, unless the peptides are derivatized prior to analysis [10].

To obtain peptide-sequence data (product-ion spectra), ESI-Q-IT-MS is usually employed; typically, peptide mixtures are first separated by liquid chromatography, which is coupled on-line to the ESI-Q-IT mass spectrometer. This method requires about 1h per analysis, but good-quality product-ion data are a very specific attribute for protein identification. Many proteomics laboratories have adopted a two-tiered protein identification strategy (Fig. 2), where MALDI-ToF-MS is used first to obtain peptide-mass fingerprinting data, and a database search is performed with these data. If the protein is not unambiguously identified, then the more time-consuming ESI-Q-IT-MS analysis is carried out to generate product-ion data. Peptide-mass fingerprints, product-ion data or peptide-sequence tags are used to search a protein sequence database to identify the protein of interest (Fig. 3) [11].

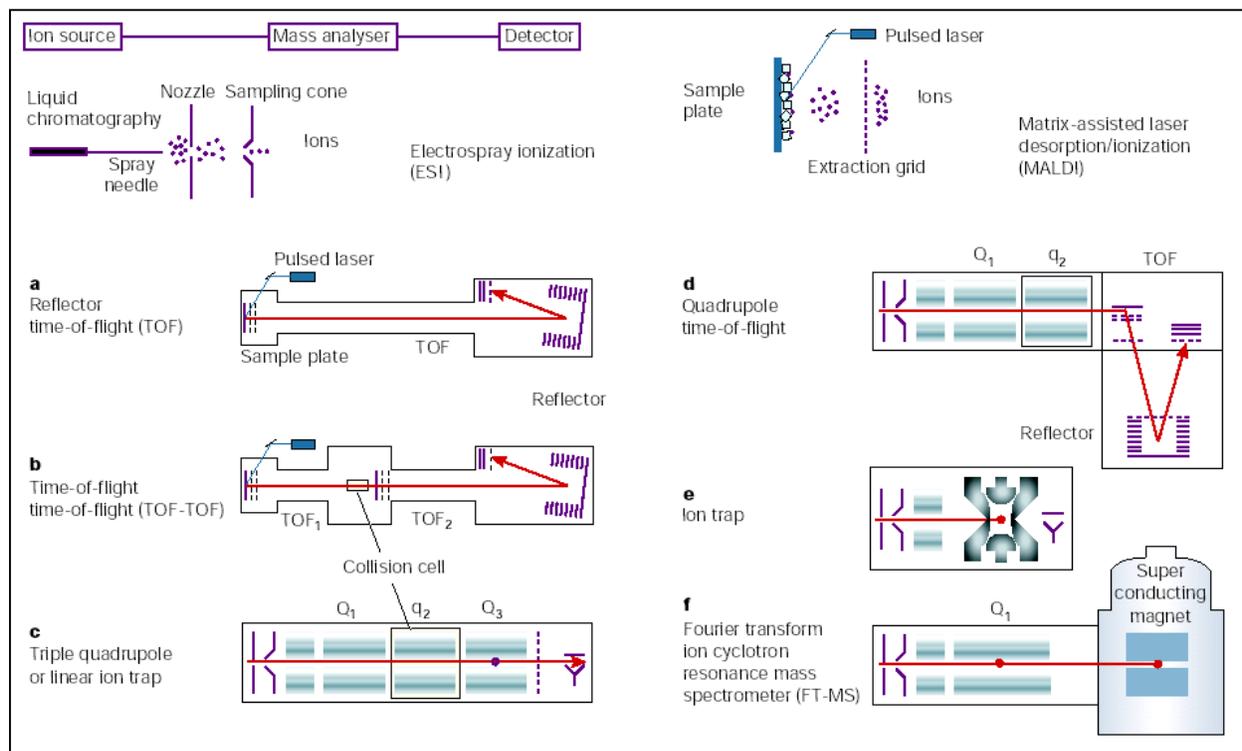


Figure 3. Mass spectrometers used in proteome research. The left and right upper panels depict the ionization and sample introduction process in electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) respectively. The different instrumental configurations (a–f) are shown with their typical ion source. **a**, In reflector time-of-flight (TOF) instruments, the ions are accelerated to high kinetic energy and are separated along a flight tube as a result of their different velocities. The ions are turned around in a reflector, which compensates for slight differences in kinetic energy, and then impinge on a detector that amplifies and counts arriving ions. **b**, The TOF-TOF instrument incorporates a collision cell between two TOF sections. Ions of one mass-to-charge (m/z) ratio are selected in the first TOF section, fragmented in the collision cell, and the masses of the fragments are separated in the second TOF section. **c**, Quadrupole mass spectrometers select by time-varying electric fields between four rods, which permit a stable trajectory only for ions of a particular desired m/z . Again, ions of a particular m/z are selected in a first section (Q_1), fragmented in a collision cell (q_2), and the fragments separated in Q_3 . In the linear ion trap, ions are captured in a quadrupole section, depicted by the red dot in Q_3 . They are then excited via resonant electric field and the fragments are scanned out, creating the tandem mass spectrum. **d**, The quadrupole TOF instrument combines the front part of a triple quadrupole instrument with a reflector TOF section for measuring the mass of the ions. **e**, The (three-dimensional) ion trap captures the ions as in the case of the linear ion trap, fragments ions of a particular m/z , and then scans out the fragments to generate the tandem mass spectrum. **f**, The FT-MS instrument also traps the ions, but does so with the help of strong magnetic fields. The figure shows the combination of FT-MS with the linear ion trap for efficient isolation, fragmentation and fragment detection in the FT-MS section.

The identification is made by comparing the experimentally-generated data with theoretical data calculated for each database entry. The rationale is to retrieve proteins that would produce the same set of data if digested and analyzed in the same manner as the protein under study. Usually, a list of

candidate proteins that most closely match the input data is generated by the search, and the candidate proteins are ranked using various scoring algorithms. Constraints can be included to limit the search to a specific subset of database entries, e.g., proteins from a particular species. Several protein sequence databases are available in the public domain. An excellent annotated database is the SWISSPROT database that is maintained by The Swiss Institute of Bioinformatics and The European Bioinformatics Institute [12, 13]. The main advantages of the SWISSPROT database are low redundancy and a high degree of annotation. The August 2002 release (Release 40.25) of the database contains 112,657 protein entries. Database search programs are often included in commercial software packages that are provided with 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 MultiIdent programs at the ExPASy Molecular Biology server [14], MS-Fit and MS-Tag at the Protein Prospector server [15], or MASCOT at the Matrix Science server [16]. These websites also provide additional proteomics software tools, technical information, and links to other resources.

An important set of web-based resources are reference databases of proteins identified by 2DE-based proteomics. In recent years, such reference databases have been established for various tissues, cell lines, body fluids and other biological systems. A reference database usually contains one or more 2D-gel images and textual information about identified proteins that can be retrieved by clicking on a particular spot on the gel images. Guidelines have been proposed in an effort to standardize the building of 2DE reference databases [17]. An index of 2DE databases that are available to the scientific community can be accessed at the ExPASy proteomics server [14].

1.2.2 ADVANTAGES OF 2DE-BASED APPROACHES

The 2DE-based approach has several characteristics that are currently unmatched by other proteomics methodologies:

1. The information content of the data obtained by the 2DE-based approach is high because a number of specific protein attributes can be determined. Thousands of proteins can be resolved and visualized simultaneously on a single 2D gel; for each protein, the isoelectric point, M_r , and the relative quantity can be measured. With MS, each protein can be characterized via a unique peptide mass fingerprint and/or amino-acid sequence tag.

2. High-resolution capabilities of 2DE allow the separation and detection of post-translationally modified proteins. In many instances, such proteins can be readily located in 2D gels because they appear as distinctive horizontal or vertical clusters of spots. In addition, modified proteins can be revealed by MS analysis, when multiple spots of the same protein are identified.

3. Individual steps of the proteome analysis (2DE, imaging, MS, database searching) can be separated in space and time. For example, 2DE and computer-assisted analyses of spots patterns can be performed in an investigator's laboratory, and protein spots of interest can be analyzed in a service facility at a later date. It is also important to point out that 2D gels can serve as high-capacity 'fraction collectors' for the purification and long term filing of proteins. On a single, dried 2D gel, thousands of proteins can be stored at room temperature, in a space equal to one notebook page. In this manner, proteins from precious sources, such as rare tumor-tissue specimens can be preserved for extended periods of time. Proteins stored within dried gels for months, even years, can be identified by MS.

4. In terms of equipment and personnel resources, the 2DE-based technology is well suited for research conducted in an academic setting. Most scientists engaged in biological research are familiar with one-dimensional gel electrophoresis; 2DE, while more complex and labor-intensive, is a natural extension of their expertise. In addition, 2DE equipment is relatively inexpensive and can therefore be supported by individual project grants. Access to other essential components, such as mass spectrometers and bioinformatic resources, can be obtained through shared-instrumentation and/or fee-for-service facilities, which are in place at many academic institutions. Thus, many investigators from various scientific disciplines can incorporate proteomics into their research programs. By contrast, alternative proteomics methodologies [18,19] rely almost exclusively on cutting-edge, high-cost MS instrumentation.

1.3 QUANTITATIVE PROTEOME ANALYSIS

1.3.1 TRADITIONAL APPROACH (2DE/MS)

The investigation of the structure, function, and control of biologic systems and processes defines a significant part of biologic and medical research [20]. Traditionally, such investigations have been essentially reductionist in nature. A system (biochemically, pharmacologically, or genetically dissected) was reconstructed from the knowledge gained from detailed analysis of individual components. Although this approach has been generally very successful, it has limited ability to establish functional connections between processes and pathways or functional modules that are concurrently active in the same cell. Furthermore, the success of this approach depends on the availability of specific assays that indicate the function of a specific component. The genome projects have yielded the complete genomic sequence for a number of species, including humans. They have also led to the emergence of an array of technologies for the systematic collection of biomolecular data and have catalyzed a new approach to study biology that is somewhat interchangeably referred to as discovery science, comprehensive biology, or systems biology [21]. The essence of this approach is the expectation that the global and, if applicable, quantitative analysis of the components that constitute a biologic system under different perturbed conditions will provide useful information that describe the state and, potentially, the mechanism of operation of the system, even in the absence of prior hypotheses [22, 23]. Several methods, including serial analysis of gene expression [24], oligonucleotide and cDNA arrays [25, 26], and large-scale sequencing of expressed sequence tags have been developed to systematically measure gene expression at the mRNA level. The discovery of post-transcriptional mechanisms that control rates of synthesis and half-life of proteins suggests that the mRNA level of a particular gene might not accurately reflect the amount of the corresponding protein expressed in the same cell, a notion that was recently verified experimentally [27–29]. Therefore, the direct measurement of protein expression is also essential for the genome wide analysis of biologic processes and systems. The global analysis of gene expression at the protein level is also termed “proteomics.” The traditional method for quantitative proteome analysis combines protein separation by high-resolution two-dimensional isoelectric focusing (IEF)/SDS-PAGE (2DE) with mass spectrometric (MS) or tandem mass spectrometric (MS/MS) identification of selected protein spots detected in the 2DE gels by use of specific protein stains.

1.3.2 THE ROLE OF ISOTOPES FOR QUANTITATION

In quantitative proteomics, the main sources of error are variations in the sample preparation procedures (e.g. protein extraction, enrichment, fractionation) and/or variations in the analysis [30]. These errors can be significantly reduced using internal standards. For this reason, stable isotopes have been used since the early 1990s in peptide analysis [31]. However these early studies were limited to the quantitation of one or a few analytes and stable isotopes have only recently been used for quantitative proteomics. Normally, one sample is labeled with a heavy reagent and a second sample is labeled with a light reagent. The two samples are then mixed and analysed by MS. The ratio between the two isotopic distributions (one for the light reagent and one for the heavy reagent) can then be determined from the mass spectra and used to calculate the relative protein quantities. Several labeling approaches in combination with MS have been developed to do quantitative profiling. These methods can be mainly divided into three classes: *in vivo* labeling, *in vitro* pre-digestion labeling, and post-digestion labeling (Fig. 4).

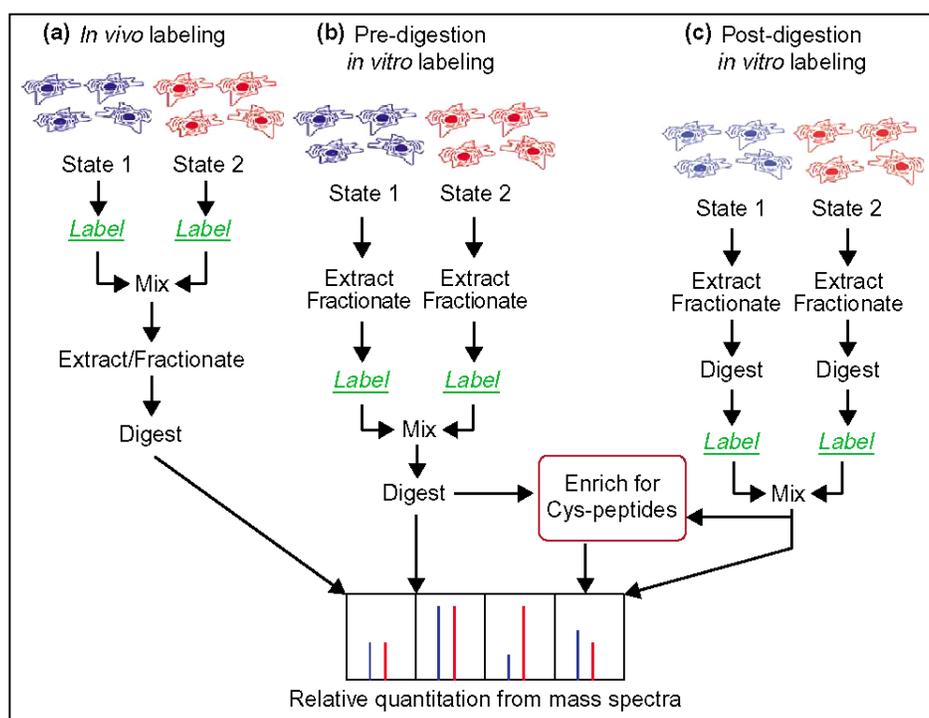


Figure 4. Schematic representation of quantitative proteomics procedures using MS. Depending on the point in the process where the label is introduced, most procedures can be classified as (a) *in vivo* labeling, (b) *in vitro* pre-digestion labeling or (c) *in vitro* post-digestion labeling. In the procedure that uses ^{18}O , the label is introduced during the digestion. However the two samples are mixed after digestion, thus it could be categorized as post-digestion labeling.

1.3.2.1 *In vivo* labeling

The *in vivo* labeling method was first described by Oda *et al.* [32]. In this procedure, yeast cells were grown in two separated media, one of which contained heavy isotopes (in this case ^{15}N). The two yeast cultures were combined, the proteins extracted, fractionated and then separated by gel electrophoresis. Finally, the proteins of interest were digested with trypsin before MS analysis and the relative quantities determined from the isotopic distribution ratios.

In vivo labeling proved to be an effective way for performing quantitative proteomic experiments. However, this approach cannot be applied to tissues or body fluids and is limited to cells that can be grown in culture in “controlled media”. However, in *in vivo* labeling procedures, the internal standards are introduced early in the process, thus obviating the variations caused by sample preparation and giving higher accuracy to the quantitation.

1.3.2.2 *In vitro* pre-digestion labeling

Gygi *et al.* [33] developed a quantitative proteome analysis using a class of reagents termed isotope-coded affinity tags (ICATs) and electrospray ionization MS. For comparing the protein profiles of yeast grown in two different conditions (galactose and ethanol as carbon source), they synthesized biotinylated iodoacetamide derivatives in a heavy form (deuterated) and in a light form, and used them to label the cysteines of the two protein extracts before combining and proteolyzing them with trypsin. The major innovation of this approach was that an affinity tag (biotin) was used to purify cysteine-containing peptides (Fig. 4), reducing the complexity of a peptide mixture by about a factor of 10 as a result, several proteins that usually can't be observed in an approach like 2DE could be identified and quantified.

Several studies have further proved the utility of these reagents in quantitative proteomics [34-36]; however when evaluating the ICAT approach, the high cost and the several chromatographic steps involved before the MS analysis should also be considered.

1.3.2.3 *In vitro* post-digestion labeling

In several procedures, isotopic labels have been introduced after proteolysis. For example, Mirgorodskaya *et al.* [37] carried out the enzymatic digest in the presence of ^{18}O -water or regular water. The sample digested with ^{18}O -water incorporates ^{18}O , generating an isotopic label that was used for relative quantitation. Another simple procedure for incorporating a label is to esterify the

carboxyl groups of the tryptic peptides with deuterated methanol (CD_3OH) [38]. This sample can then be mixed and compared with another sample that has been esterified with regular methanol. One of the minor disadvantages of this approach is that the harsh esterification condition requirement (hydrochloric acid solution) might produce partial deamidation of asparagine and glutamine residues, increasing the sample heterogeneity. In addition, this procedure should be carried out in anhydrous conditions, which can be difficult to obtain.

In a strategy named mass-coded abundance tagging (MCAT), the ϵ -amino group of lysine was also labeled [39]. Only one sample is modified with the reagent (*O*-methylisourea) and compared with the unmodified sample for determining the relative quantities. Although this procedure is simple and inexpensive, several issues related to the difference in the chemical-physical characteristics between the labeled and unlabeled peptides markedly reduce the accuracy of the quantitation.

1.4 PROTEOMICS IN DISEASE STUDIES

In combination with genomics, proteomics can provide a holistic understanding of the biology underlying disease processes. Information at the level of the proteome is critical for understanding the function of specific cell types and their role in health and disease. Mammalian systems are much more complex than can be deciphered by their genes alone. Expression analysis directly at the protein level is necessary to unravel the critical changes that occur as part of disease pathogenesis. This is because proteins are often expressed at concentrations and forms that cannot be predicted from mRNA analysis. Proteomics also provides an avenue to understand the interaction between the functional pathways of a cell and its environmental milieu, independent of any changes at the RNA level.

Quantitative proteomics strives to investigate the changes in protein expression in different states, such as in healthy and pathological tissue or at different stages of the disease. This enables the identification of state- and stage-specific proteins (Fig. 5).

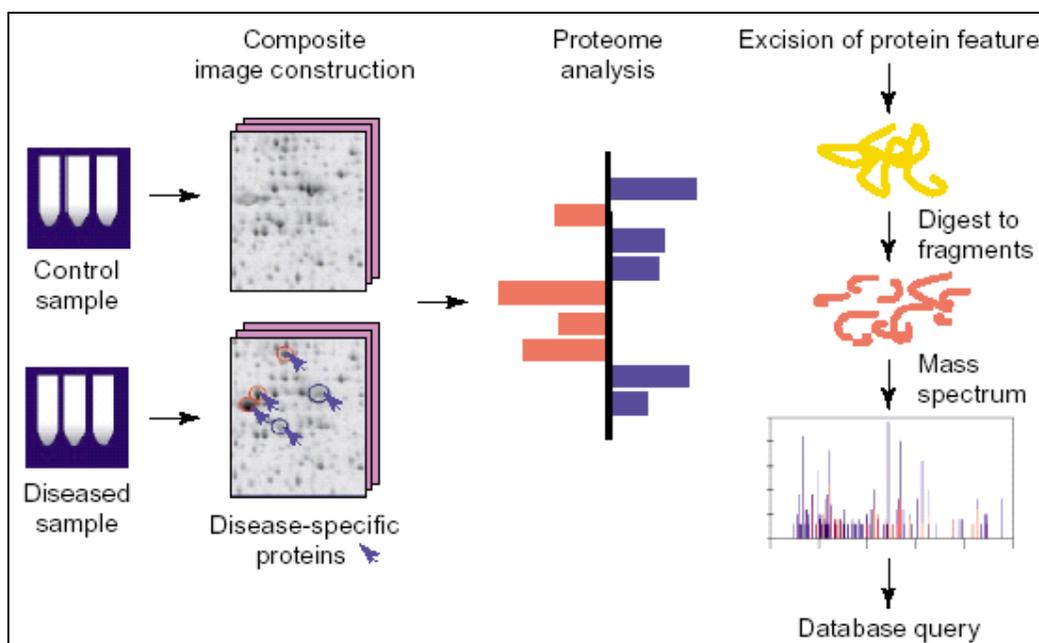


Figure 5. Scheme of a proteomic disease study. Protein samples from control and pathological samples are run on two-dimensional gels. Multiple images are generated and the resulting information is assembled into a proteomic database. The blue arrows indicate proteins that are specifically altered in disease samples. Protein profiles can be compared using appropriate software. The columns represent proteins that are increased (red) or decreased (blue) on the gels. Protein features of interest are identified by excision from the gel, followed by protease digestion and MS analysis.

1.4.1 PROTEOMICS IN CANCER RESEARCH

Cancer is a multi-faceted disease that presents many challenges to clinicians and cancer researchers searching for more-effective ways to combat its often devastating effects [40]. Among the central challenges of this disease, are the identification of markers for improved diagnosis and classification of tumors, and the definition of targets for more-effective therapeutic measures. Although some cancer-related genes have been mainly identified by mutational analysis, at present, tumor classification is a complex process based primarily on site and histological examination. However, tumors with a similar histological appearance can follow significantly different clinical courses and show different responses to therapy [41]. Given the wide diversity of tumors, even those derived from the same tissue, additional methods of classification are urgently required. Furthermore, it is now clear that the genetic make-up of both the tumor and the individual patient can influence the outcome of a given treatment. Therefore, to be most effective, future treatments will need to be tailored not only for the specific tumor type but also, in some cases, for the individual as well.

Proteomics approaches to tumor marker identification hold the promise of identifying specific protein modifications in tumor tissues to assist in individualizing treatments for certain cancers. With the completion of the draft sequence of the human genome [42, 43], there is a great deal of interest in the use of functional genomics, especially gene expression profiling techniques such as DNA microarrays and proteomics, to identify cancer-associated genes and their protein products. These two complementary technologies permit the analysis of thousands of genes or proteins simultaneously, and have the potential to identify markers for early detection, classification and prognosis of tumors, as well as pinpointing targets for improved treatment outcomes.

Global protein profiles can be produced for normal compared with tumor cells in a given tissue, or for cells before and after treatment with a specific drug. Currently, this is the most widely used model of proteomics and is largely dependent upon 2D gel electrophoresis (2DE) for visualization of protein profiles. Expression proteomics is the protein equivalent of DNA microarray analysis that define global patterns of RNA expression under various conditions. Like DNA microarrays, it has the advantage of being non-prejudicial and could define unexpected ways in which known proteins regulate cellular responses.

1.4.2 PROTEOMICS IN DRUG DEPENDENCE AND PSYCHIATRIC DISORDERS

Drug addiction, like all other psychiatric disorders, is diagnosed today solely on the basis of the behavioral abnormalities that patients exhibit [44]. For example, addiction can be defined as compulsive drug seeking and taking despite adverse consequences or as loss of control over drug use. However, there is no objective diagnostic information we can offer individuals concerning their risk for addiction in general, let alone addiction for a specific substance, nor can we offer patients informed advice concerning their risk for relapse. Moreover, current treatments for drug addiction are inadequate for most individuals.

Recent advances in proteomics and functional genomics can be expected to dramatically improve psychiatric practice overall and the treatment of addictive disorders in particular. Two major areas of advances in this evolving field of “psychogenomics” are seen: (1) identification of genes that confer risk for an addiction and (2) identification of genes and proteins that contribute to the regulation of reward, motivation, and cognition under normal circumstances and to abnormalities in these behaviors that characterize an addicted state. There is now considerable optimism that these advances will lead to objective diagnostic tests, improved treatments, and eventually preventive measures and cures.

Epidemiological studies have indicated that drug addiction is a highly heritable disorder. Approximately 40–60% of the risk for alcohol, cocaine, or opiate addiction appears to be genetic [45]. Data are not yet available for nicotine or other substances, although anecdotal information suggests similar degrees of heritability. Despite this genetic basis, however, efforts to identify specific genes involved in drug addiction have not to date been successful. The difficulty in finding such genes is comparable with the difficulty in finding genes for other common conditions (e.g., hypertension, congestive heart failure, and asthma). One possibility is that these diseases are caused by a relatively large number of genes, such that it is extremely difficult to identify the individual genes involved, each of which is only responsible for a small percentage of the overall risk. Another possibility is that the tools for fine genome-wide scans in large numbers of individuals have only recently become available.

Knowledge of genes that confer risk for addiction could be used to select the optimal treatment program for an individual addict. For example, in the depression field, some antidepressants are serotonin-reuptake inhibitors, whereas others inhibit norepinephrine reuptake. There is a major

effort today to identify genetic factors that can be used to predict whether a person with depression would respond better to one or the other. Such pharmacogenomic studies are in their early stages, and would appear premature for addiction, because treatments for drug addiction, and our knowledge of the underlying genetic factors, are still limited.

Once addiction vulnerability genes or genes that predict pharmacological responses are discovered, the next step will be to place these genes in mice to enable studies of the underlying molecular and neural mechanisms that link the genes to abnormal behavior.

As genomic and proteomic efforts succeed in identifying genes and proteins involved in normal behavior and in addiction, several tangible benefits will result. The most obvious is the identification of novel targets for psychotherapeutic medications. All but a few currently used psychiatric medications act on neuro-transmitter receptors or transporter proteins. Yet these proteins represent a minuscule fraction of all neuronal proteins, and it is likely that among this remaining array of proteins are viable drug targets, including targets for truly effective anti-addiction medications. The challenge is to find them.

Similarly, as we define the detailed etiology and pathophysiology of addiction, it should be possible to develop medications that intervene in the addiction process.

Advances in our understanding of the genetics and neurobiology of drug addiction will have dramatic implications for diagnosis and prevention as well.

The fields of genomics and proteomics provide tools of unprecedented power for identifying genes and proteins that control complex behavior under normal and pathological conditions. Eventually, these discoveries can be exploited for clinical applications as diverse as improved treatments, diagnostic tests, and ultimately disease prevention and cure.

1.5 BIBLIOGRAPHY

- [1] Wilkins, M.R., Pasquali, C., Appel, R.D., Ou, K., Golaz, O., Sanchez, J.C., Yan, J.X., Gooley, A.A., Hughes, G., Humphery-Smith, I., Williams, K.L. and Hochstrasser, D.F., *Bio/Technology* 1996, **14**: 61-65.
- [2] Pennington, S.R., Wilkins, M.R., Hochstrasser, D.F., Dunn, M.J., *Trends Cell. Biol.* 1997, **7**: 168-173.
- [3] Beranova-Giorgianni S., *Trends Anal. Chem.* 2003, **22**: 273-281.
- [4] Anderson, N.G., Matheson, A., Anderson, N.L., *Proteomics* 2001, **1**: 3-12.
- [5] Herbert, B., *Electrophoresis* 1999, **20**: 660.
- [6] Rabilloud, T., Chevillet, M. in: T. Rabilloud (Editor), *Proteome Research: Two-Dimensional Gel Electrophoresis and Identification Methods*, Springer, Berlin, Germany, 2000, p. 9.
- [7] O'Farrell, P.H., *J. Biol. Chem.* 1975, **250**: 4007.
- [8] Shevchenko, A., Wilm, M., Vorm, O., Mann, M., *Anal. Chem.* 1996, **68**: 850.
- [9] Kinter, M., Sherman, N.E., *Protein Sequencing and Identification Using Tandem Mass Spectrometry*, Wiley, New York, USA, 2000, p. 29.
- [10] Keough, T., Youngquist, R.S., Lacey, M.P., *Proc. Natl. Acad. Sci. U.S.A.* 1999, **96**: 7131.
- [11] Yates III, J.R., *Electrophoresis* 1998, **19**: 893.
- [12] Bairoch, A., in: M.R. Wilkins, K.L. Williams, R.D. Appel, D.F. Hochstrasser (Editors), *Proteome Research: New Frontiers in Functional Genomics*, Springer, Berlin, Germany, 1997, p. 95.
- [13] Bairoch, A., Apweiler, R., *Nucleic Acids Res.* 2000, **28**: 45.
- [14] <http://www.expasy.org>
- [15] <http://prospector.ucsf.edu>
- [16] <http://www.matrixscience.com>
- [17] Appel, R.D., Bairoch, A., Sanchez, J.-C., Vargas, J.R., Golaz, O., Pasquali, C., Hochstrasser, D.F., *Electrophoresis* 1996, **17**: 540.
- [18] Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., Aebersold, R., *Nat. Biotechnol.* 1999, **17**: 994.
- [19] Smith, R.D., Pasa-Tolic, L., Lipton, M.S., Jensen, P.K., Anderson, G.A., Shen, Y., Conrads, T.P., Udseth, H.R., Harkewicz, R., Belov, M.E., Masselon, C., Veenstra, T.D., *Electrophoresis* 2001, **22**: 1652.
- [20] Aebersold, R., *Review Quant. Proteome Technol.* 2003, **187**: 315-320.

- [21] Aebersold, R., Hood, L.E., Watts, J.D., *Nat. Biotechnol.* 2000, **18**: 359.
- [22] DeRisi, J., Penland, L., Brown, P.O., *et al.*, *Nat. Genet.* 1996, **14**:457-60.
- [23] Ideker, T., Thorsson, V., Ranish, J.A., *et al.*, *Science* 2001, **292**: 929-34.
- [24] Velculescu, V.E., Zhang, L., Vogelstein, B., Kinzler K.W., *Science* 1995, **270**: 484-7.
- [25] Shalon, D, Smith, SJ, Brown, PO., *Genome Res* 1996, **6**: 639-45.
- [26] Chee, M.R., Yang, E., Hubbell, A., *et al.*, *Science* 1996, **274**: 610-4.
- [27] Anderson, L., Seilhamer, J., *Electrophoresis* 1997, **18**: 533-7.
- [28] Gygi, S.P., Rochon, Y., Franza, B.R., Aebersold, R., *Mol. Cell. Biol.* 1999, **19**: 1720-30.
- [29] Griffin, T.J., Gygi, S.P., Ideker, T., *et al.*, *Mol. Cell. Proteomics* 2002, **1**: 323–33.
- [30] Sechi, S., Yoshiya, O., *Curr. Opin. in Chem. Biol.* 2003, **7**: 70-77.
- [31] Desiderio, D.M., *Life Sci.* 1992, **51**: 169-176.
- [32] Oda, Y., Huang, K., Cross, F.R., Cowburn, D., Chaqit, B.T., *Proc Natl Acad USA* 1999, **96**: 6591-6596.
- [33] Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., Aebersold, R., *Nat. Biotechnol.* 1999, **17**: 994-999.
- [34] Griffin, T.J., Han, D.K., Gygi, S.P., Rist, B., Lee, H., Aebersold, R., Parker, K.C., *J. Am. Soc. Mass. Spectrom.* 2001, **12**: 1238-1246.
- [35] Han, D.K., Eng, J., Zhou, H., Aebersold, R., *Nat. Biotechnol.* 2001, **19**: 946-951.
- [36] Smolka, M., Zhou, H., Aebersold, R., *Mol. Cell. Proteomics* 2002, **1**: 19-29.
- [37] Mirgorodskaya, O.A., Kozmin, Y.P., Titov, M.I. Korner, R., Sonksen, C.P., Roepstorff, P., *Rapid Commun. Mass Spectrom.* 2000, **14**: 1226-1232.
- [38] Goodlett, D.R., Keller, A., Watts, J.D., Newitt, R., Yi, E.C., Purvine, S., Eng, J.K., von Haller, P., Aebersold, R., Kolker, E., *Rapid Commun. Mass Spectrom.* 2000, **15**: 1214-1221.
- [39] Cagney, G., Emili, A., *Nat Biotechnol* 2002, **20**: 163-170.
- [40] Simpson, R.J. and Dorow, D.S., *TRENDS Biotech.* 2001, **19**: 40-48.
- [41] Wasinger, V.C., Cordwell, S.J., Cerpa-Poljak, A., Yan, J.X., Gooley, A.A., Wilkins, M.R., *et al.* *Electrophoresis* 1995, **16**: 1090-4.
- [42] Banks, R.E., Dunn, M.J., Hochstrasser, D.F., Sanchez, J-C, Blackstock, W., Pappin, D.J., *Lancet* 2000, **356**: 1749-56.
- [43] Anderson, N.L., Matheson, A.D., Steiner, S., *Curr. Opin. Biotechnol.* 2000, **11**: 408-12.
- [44] Nestler, E.J., *J. Neurosci.* 2001, **21**: 8324-8327.
- [45] Nestler, E.J., *Nat. Genet.* 2000, **26**: 277-281.

CHAPTER 2

COMPARATIVE PROTEOMICS: A NEW ISOTOPICALLY MARKED ALKYLATING AGENT

2.1 INTRODUCTION

Often one of the important aspects in proteomics studies is the need to measure the relative amount of proteins [1]. This is essential for studying the effect of an agent on a biological system or for the comparison of two different biological states. The most traditional approach for quantifying proteins in a proteomic experiments is to measure the ratios of the protein spot intensities in two independent gels that are stained by Coomassie or fluorescent stain. More recently, many other alternative procedures have been introduced to quantify proteins. These include the use of ICAT reagents [2-4], the use of differential labeling of cells in culture [5, 6], the use of a H₂¹⁸O water in the digest [7, 8] and the use of deuterated methanol for peptide esterification [9]. All of these methodologies have been shown to be useful in some cases. However, none of them works all the time and they have their pros and cons. There is still a need for simple and inexpensive alternatives for quantifying proteins. The approach proposed here is a simple method for determining relative quantities of proteins isolated by gel electrophoresis. The method is based on the differential labelling of the mixtures by use of a commercially available acrylamide and deuterium-labelled [2,3,3'-d₃]-acrylamide to alkylate proteins prior to 2-D gel electrophoresis. The tryptic digests of the separated proteins were subjected to reflector MALDI-TOF analysis and the relative peak heights of cysteine-containing peptides were used to quantify their precursor proteins. This approach was tested for the relative quantification of proteins within an artificial mixture of standard proteins and for proteins observed in the 2-D map of rat serum. A good correlation was found between the measured ratios derived from MALDI-TOF data and those theoretically calculated prior to 2-D analysis via known mixing ratios of the two alkylating reagents. It is worth noting that the use of deuterium labelled acrylamide was used by Sechi and Chait [10] to improve the information content of MS experiments. The present data demonstrate that the alkylation efficiency is not influenced by the isotopic composition of the acrylamide, which makes the present approach fairly simple and

fairly reliable for the relative quantification of proteins derived from two different mixtures and treated with the light and heavy forms of this commonly used alkylating agent.

2.2 EXPERIMENTAL PROCEDURES

2.2.1 LABELING A STANDARD PROTEINS MIXTURE

The standard proteins used were: chicken egg lysozyme (accession no. P00698 in Swiss-Prot database), human α_1 acid glycoprotein (accession no. P02763 in Swiss-Prot database) and bovine apo-transferrin (accession no. Q29443 in Swiss-Prot database) purchased from Sigma-Aldrich (St. Louis, MO; USA). Proteins were dissolved (50 μ L) in 40 mM Tris, pH 8.5 and fully unfolded and reduced by treatment with 5 mM tributyl phosphine, 2 M thiourea and 7 M urea for 90 minutes at room temperature. The resulting reaction mixture was divided into two parts, the first was alkylated/labeled with 100 mM d_0 -acrylamide, while the second was alkylated/labeled with 100 mM d_3 -acrylamide. After 5 h, the two fractions were mixed in five different ratios (90/10* - 70/30* - 50/50* - 30/70* - 10/90*; * = d_3 -acrylamide alkylated fraction) and subjected to overnight dialysis at 4°C to stop the alkylation and to remove excess reagents and salts.

2.2.2 PREPARATION AND LABELING OF RAT SERUM

One hundred μ l of rat serum containing 6 mg total protein were added to 40 mM Tris buffer, pH 8.5, containing 5 mM TBP, 1.5 M thiourea, and 6 M urea and left for 2 hours at room temperature. The resulting mixture was divided in two parts, the first was alkylated with 50 μ mol d_0 -acrylamide, while the second fraction was alkylated with 50 μ mol d_3 -acrylamide. The resulting two fractions were then mixed in the ratios 50/50* and 30/70*, dialysed overnight and subjected to 2-D gel analysis.

2.2.3 TWO-DIMENSIONAL ELECTROPHORESIS

Seven cm long, pH 3-10 immobilized pH gradient strips (IPG, BioRad laboratories, Hercules, CA, USA) were rehydrated for 4 h with 150 μ L of 2-D solubilizing solution containing 450 μ g protein. Isoelectric focusing was conducted at 20°C for 25000 Vh at a maximum of 5000 V using a Protean IEF cell (BioRad). For the second dimension, the IPG strips were equilibrated in a solution of 6 M

urea, 2% SDS, 20% glycerol, Tris-HCl pH 8.8, then were laid on a 7-20% T gradient SDS-PAGE. Gels were run at 10°C and 5 mA/gel for 1 h, 10 mA/gel for 1 h, 20 mA/gel for 2 h. Proteins were detected by colloidal Coomassie blue, and destained in 5% acetic acid.

2.2.4 PROTEIN IDENTIFICATION AND QUANTITATION

Protein identification and quantitation were performed by the Department of Computational, Analytical and Structural Sciences staff of GlaxoSmithKline Discovery Research.

The gel bands of interest were excised with a razor blade, placed in eppendorf tubes, and destained by washing twice with 50% 5 mM Tris/50% acetonitrile solution. The gel pieces were dehydrated by addition of acetonitrile. Excess solvent was removed, followed by a 20 min. drying by a SpeedVac centrifuge apparatus. The gels were rehydrated by adding 15-30 μ L of a solution containing 2.5 mM Tris and modified trypsin from Promega. The rehydrated gels were kept for 4 h at 37°C. Peptides were extracted by adding 30 μ L of a solution containing 50% acetonitrile and 50% water with 1% of formic acid. The extracted peptides were loaded onto the target plate by mixing 1 μ L of each solution with the same volume of a matrix solution (10 mg/mL α -cyano-4-hydroxycinnamic acid in 50% ethanol and 50% acetonitrile) and left to dry at room temperature. The MS analyses were performed using a TofSpec-2E MALDI-TOF instrument (Micromass, Manchester, UK), equipped with a pulsed nitrogen laser (337 nm, pulse width 4 ns) and operated in reflectron mode with an accelerating voltage of 20 kV. Identification of the standard proteins was performed by submitting its known sequences to the PeptideMass tool and manually comparing the predicted peptide masses of each standard protein to the measured masses in the mass spectrum. For identification of the proteins from rat serum, the measured mass of the tryptic peptides were searched against *rat* entries from Swiss-Prot using the MS-Fit program (University of California, San Francisco; prospector.ucsf.edu/). The observed pI and Mr of the identified spots in the 2DE map were compared with the theoretical pI and Mr to confirm identifications. For quantifying the relative abundance of a protein, the peak height of the monoisotopic peak of the light acrylamide-labeled peptide was divided by the peak height of the monoisotopic peak of the heavy [2,3,3'-d₃]-acrylamide-labeled form of the peptide. When more than one cysteine-containing peptide was detected for a particular protein, the abundance ratio was calculated from the peak pair with the highest signal intensity. In cases in which the shape of the isotope distributions of the light and heavy [2,3,3'-d₃]-acrylamide-labeled peptides differed (*i.e.* asymmetric shape of the clusters) we

suspected the presence of contaminating peaks and did not use such peak pairs for quantification. Instead, the peak pair with the next highest signal intensity was used.

2.3 RESULTS

The method for quantitative protein profiling described in this work is schematically illustrated in Fig. 1. It is based on the separation of proteins alkylated/labeled with acrylamide (d_0/d_3) by 2DE and their identification and quantification by mass spectrometry. The proteins contained in two separate samples are first alkylated with the isotopically normal (non-deuterated) or heavy forms (deuterated), respectively, of acrylamide. The two samples are then combined and concurrently separated by 2DE in the same gel, and the separated proteins are detected by an MS compatible staining protocol. The proteins migrating to specific spots are enzymatically digested in the gel matrix, and the resulting peptides are extracted and analyzed by mass spectrometry. Protein identification is achieved by MALDI-TOF analysis and sequence data base searching. It should be pointed out that the analysis of the intact protein is an optional step which may be used as an early screening for possible gel-induced and/or post-translational modifications. The ratio of abundance of the protein in the spot analyzed is determined by the ratio of signal intensities for the isotopically normal and heavy forms of a specific, tagged peptide.

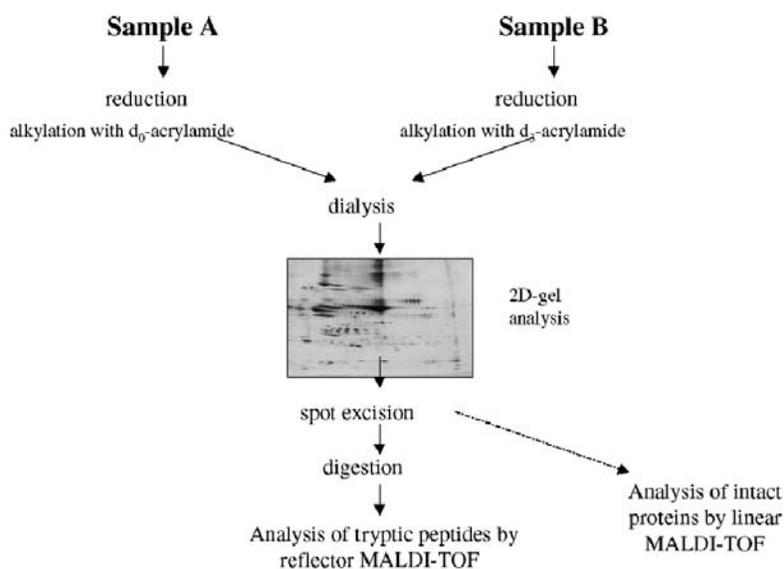


Figure 1. Schematic illustration of the proposed method for quantitative protein profiling.

2.3.1 ACCURACY OF QUANTIFICATION

For evaluating the accuracy of quantification of the method we prepared and analyzed a mixture of three standard proteins, lysozyme, α_1 acid glycoprotein, and apo-transferrin. The mixture was divided into two parts, one sample was alkylated with the isotopically normal and the other with the heavy form of the acrylamide, the samples were combined in five different ratios (90/10* – 70/30* – 50/50* – 30/70* – 10/90*) and separated by 2DE which yielded five different 2D maps (Fig. 2).

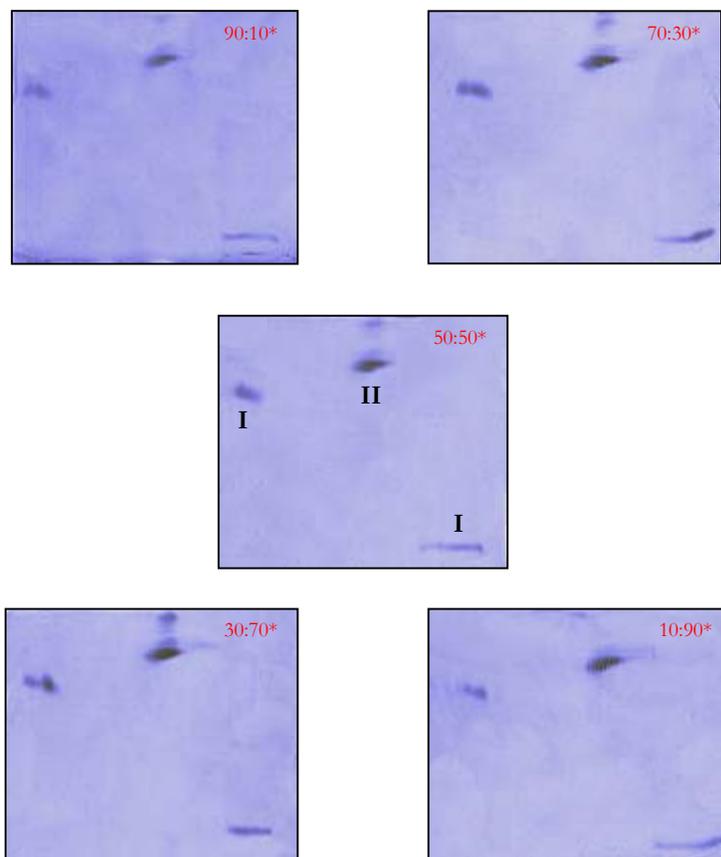


Figure 2. Five different 2D maps obtained by d_0 and d_3 alkylated samples combined in five different ratios: (1) 90/10*, (2) 70/30*, (3) 50/50*, (4) 30/70*, (5) 10/90*. The three standards proteins correspond to: (I) lysozyme, (II) α_1 acid glycoprotein, (III) apo-transferrin.

Protein spots were detected by colloidal Coomassie blue stain and in-gel digested with trypsin, and the resulting peptides were identified and quantified by mass spectrometry. For each protein, the predicted masses of the labeled, cysteine-containing peptides were also calculated and used to identify the corresponding peaks in the mass spectra. Fig. 3 shows the peptide mass spectrum of

apo-transferrin and expansions of the peak areas containing the signals for cysteine-containing peptides.

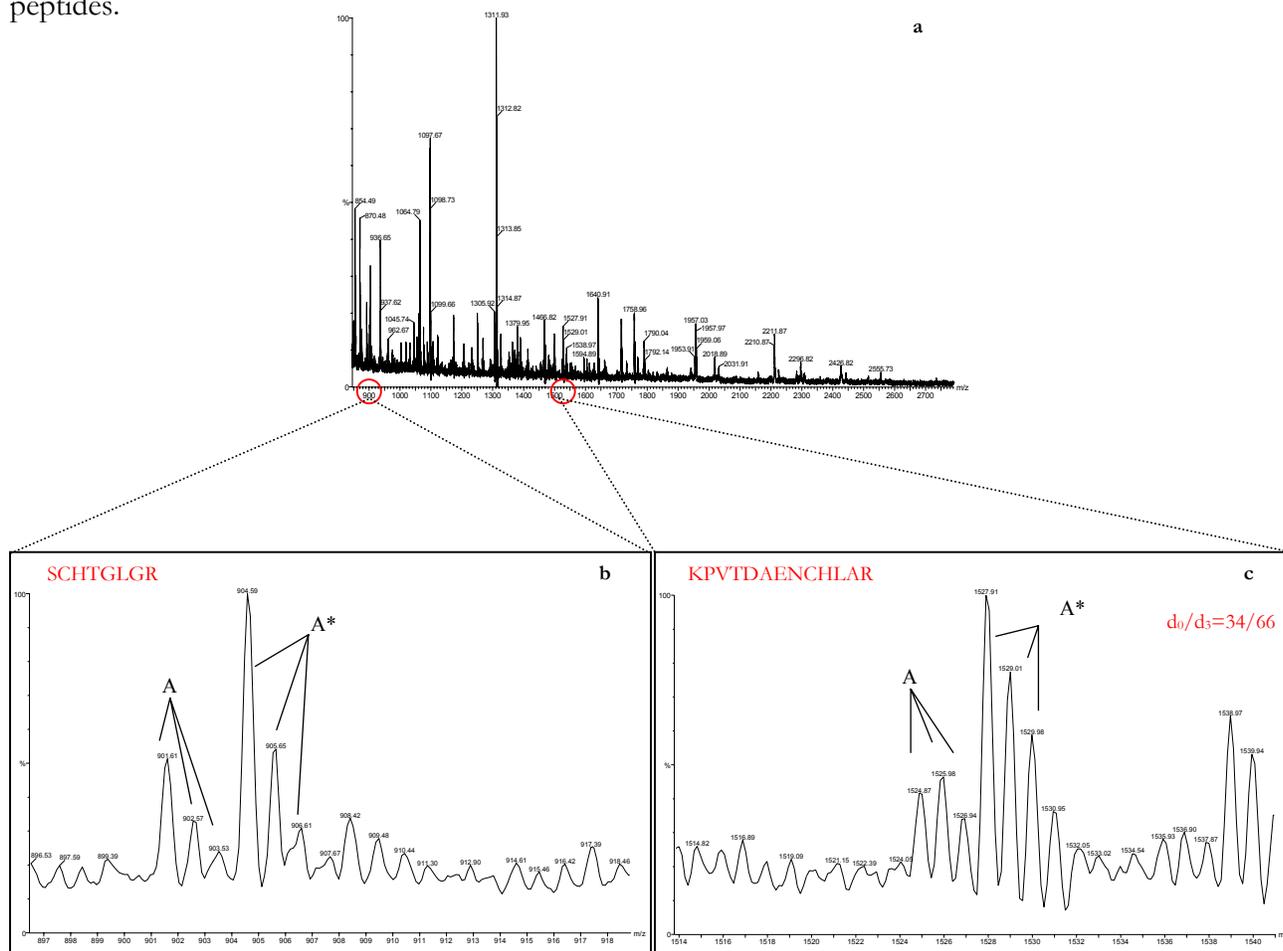


Figure 3. Measurement of the relative abundances of proteins labeled with isotopically normal and heavy acrylamide. (a) Reflector MALDI mass spectrum of an *in-situ* digest of apo-transferrin taken from a 2-D map prepared using a mixture of standard proteins alkylated with d_0 -acrylamide and d_3 -acrylamide in the ratio 30/70*. (b) and (c) are two short intervals taken from (a) pertaining to the sequences SCHTGLGR and KPVTDAENCLAR.

Fig. 3(a) gives the MALDI spectrum of the entire digest, while (b, c) give two short intervals taken from the same spectrum. These intervals refer to the indicated peptide sequences each of which has a single cysteine residue. In each of the two spectra there are two isotopic distributions marked A and A* where a difference of 3 Th (Thomson, unit of m/z) in the m/z of the corresponding peaks within the two distributions is clearly evident. This observation can be easily justified by the following considerations: It is well documented¹⁶⁻¹⁸ that the free $-SH$ group in the side chain of cysteine is the favourite site for interaction with acrylamide and other commonly used alkylating

agents. The two isotopic distributions in (b) and (c) refer to two sister peptides derived from two protein fractions, the one treated with d_0 - and the other one with d_3 -acrylamide. The same spectra show that the relative peak heights associated with (A) is $\sim 34\%$ compared to $\sim 66\%$ associated with the corresponding peaks in distribution (A*), which is in reasonable agreement with the labelling ratio (30:70*) effected prior to the electrophoretic separation. Similar spectra were used to construct table 1, which compares the measured and calculated protein ratios marked with d_0 - and d_3 -acrylamide.

These results demonstrate the possibility of using d_0 - and d_3 - acrylamide alkylation for accurate quantification of proteins separated by 2DE.

Protein	Expected ratio	Observed ratio	Identified Sequences
	d_0/d_3	d_0/d_3	
Apo-transferrin	30/70	33/67	DGTRKPVTDAENCHLAR
		34/66	KPVTDAENCHLAR
		34/66	KNYELLCGDNTR
		35/65	SCHTGGR
	50/50	47/53	DGTRKPVTDAENCHLAR
		49.5/50.5	KPVTDAENCHLAR
		47/53	SCHTGGR
$\alpha 1$ -acid glycoprotein	30/70	33/67	CEP LLEKQHEK
	50/50	53/47	CEP LLEKQHEK

Table 1. Composition of each of the standard protein mixtures used for the experiment shown in Fig. 2. The expected quantitative ratios and the quantitative ratios that were experimentally determined are listed in columns two and three.

2.3.2 QUANTIFICATION OF CHANGES OF PROTEINS ABUNDANCE IN RAT SERUM

To assess the ability of the method in identifying proteins in complex mixtures and to quantify changes in their abundance, we investigated changes in the protein profile on rat serum, which yielded the 2-D map in Fig.4.

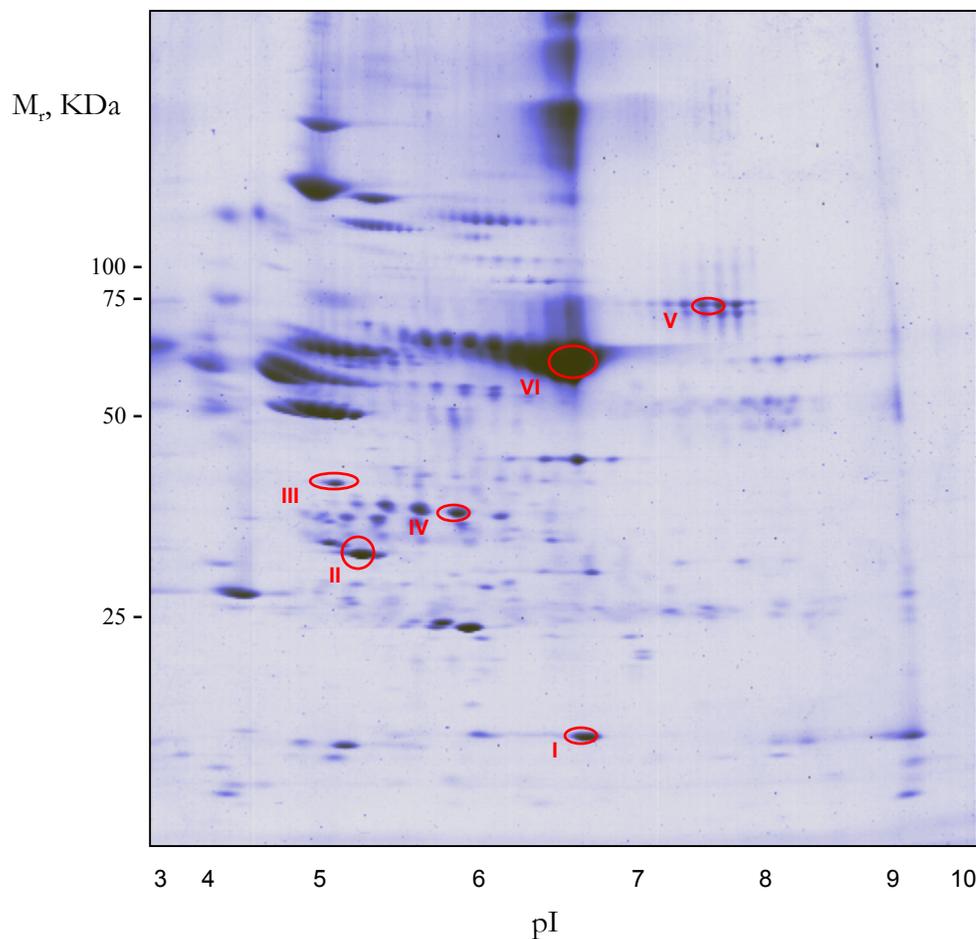


Figure 4. 2D maps in the 3-10 IPG interval of rat serum composed of two fractions, the first (30%) was alkylated with d_0 -acrylamide, while the second (70%) was alkylated with d_3 -acrylamide. The circled spots refer to: transthyretin(I), apo-lipoprotein(II), apo-lipoprotein A4(III), α_1 -macroglobulin(IV), sero-transferrin(V) and albumin(VI).

The alkylation was achieved using the ratio 30/70* of d_0/d_3 -acrylamide; a number of separated proteins were digested *in situ* and subjected to reflector MALDI-TOF analysis. A representative

spectrum pertaining to albumin is given in Fig. 5(a), while short scan intervals taken from the same spectrum are presented in Figs. 5 (b, c). The observed isotopic distributions in both spectra are attributed to the indicated sequences, each of which contains a single cysteine residue. Furthermore, the more intense peaks in both spectra are displaced by 3 Th from their weaker counterparts. Such displacement coincides with a single alkylation channel associated with the d_3 -acrylamide. Considering the relative peak heights in both isotopic distributions we have calculated a ratio of 34/66*, which is in reasonable agreement with the labelling ratio 30/70* prior to 2-D separation.

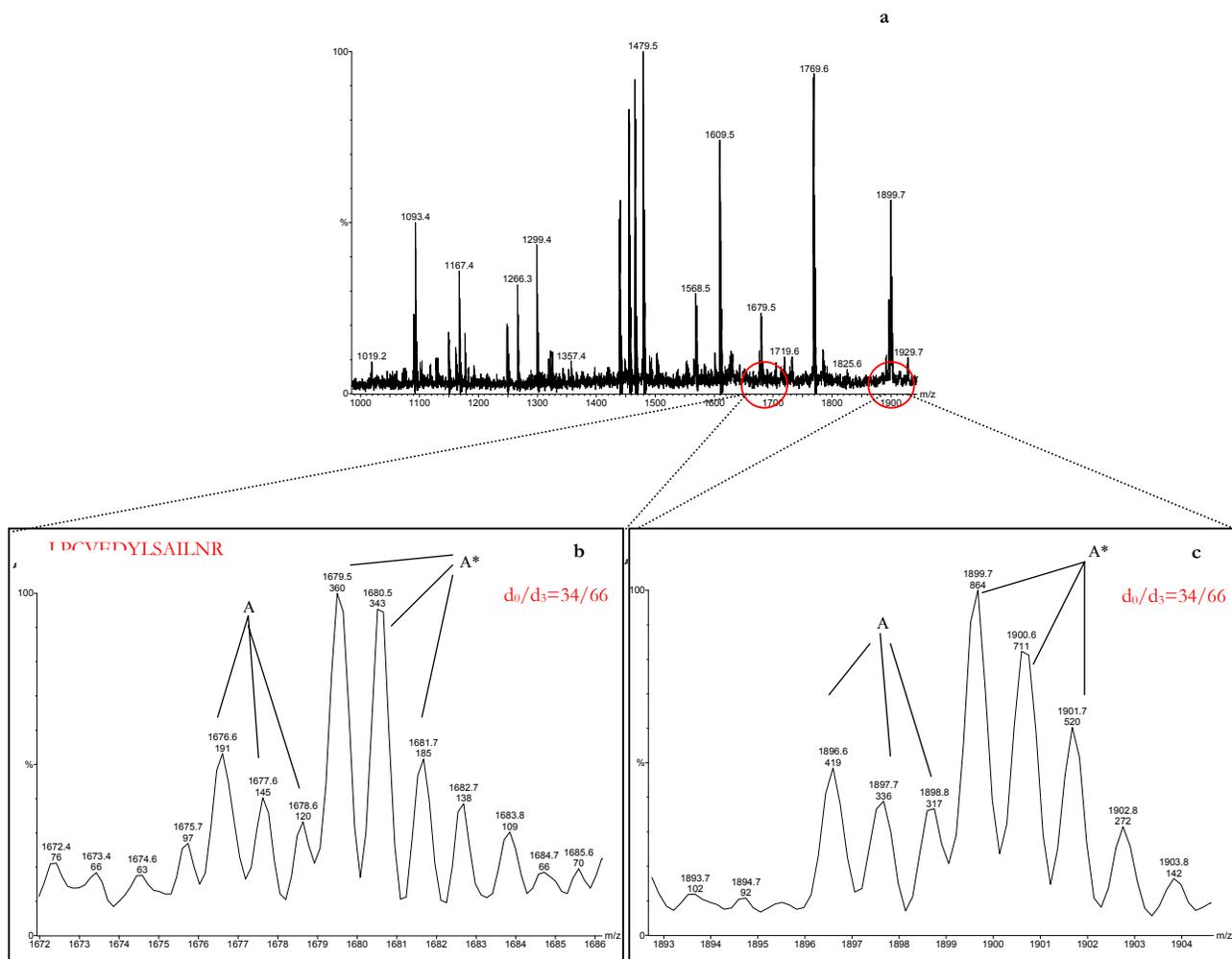


Figure 5. (a) Reflector MALDI mass spectrum of an *in-situ* digest of albumin (spot VI) in Fig. 4. (b) and (c) are two short intervals taken from spectrum in (a) pertaining to the sequences LPCVEDYLSAILNR and RPCFSALTVDETYVPK used for the relative quantification.

2.3 DISCUSSION

Cys residues are found in a wide variety of storage and structural proteins, enzymes, hormones, receptors, microbial toxins [11]. Disulphide bonds associated with Cys residues are structural units essential in most proteins and are largely responsible for the conformations and tertiary structures that proteins assume both in solution and in the solid state. Because -SH groups are among the most reactive functionalities in chemical and biochemical systems (e.g., they undergo sulphydryl-disulphide exchange, oxidation-reduction, nucleophilic addition, displacement conjugation reactions [12]) it is only natural that a variety of methods have been described for stabilizing and/or protecting them, in order to facilitate protein analysis. Among the most popular methods reported have been alkylation with iodoacetamide (IAA) and N-ethylmaleimide (NEM). The use of IAA is not ideal because it is difficult to drive the reaction to completion [13]; moreover, the reaction is not specific for -SH groups, since it has been reported to modify also α -NH₂ groups of N-terminal chains and ϵ -NH₂ groups of Lys. Modification with NEM can result in a multiplicity of labelled peptides owing to the formation of diastereoisomers and hydrolysis of the succinimide ring of the Cys adducts [14].

In the light of the above considerations, it is worth critically evaluating the results here presented. The data in Figs. 3 and 5 give an initial indication that isotopic labelling of the acrylamide alkylating agent together with MALDI-TOF mass spectrometry can be considered a valid approach for relative quantification of gel-separated proteins. The difference of 3Th in the measured m/z values of sister peptides was enough to distinguish between them; however, there is no reason why other alkylating agents with a higher number of deuterium atoms cannot be used in a similar fashion. Having said this, it is fair to look closely at this approach compared to existing strategies for identifying the strengths and the weaknesses of each of such approaches. Both the ICAT and the present approaches assume specific and complete alkylation of all free -SH groups within the investigated sequence(s), an assumption which is not fully supported by recent work [15-18]. These authors have demonstrated that alkylation times as long as 6 h were not sufficient to produce more than 80% alkylation; such efficiency is likely to be less in the case of the ICAT which has to accommodate a biotin group and an ethylene glycol linker group. Not all proteins of interest contain cysteine, some proteomes will have lower total protein coverage than the 92% commonly cited for yeast¹². Some cysteine residues may contain post-translational modifications prior to isolation making them unavailable for alkylation with cysteine-specific ICAT or acrylamide. To address the drawback of *in vitro* labelling of non-cysteine-containing peptides for protein expression analysis, a method

complementary to the use of ICAT reagents has been described by Goodlett et al [19]. The method uses differential stable isotope esterification of carboxylic acids in peptides to provide an internal standard for relative quantification of peptides in a mixture regardless of their cysteine content. Basically, protein digests derived from two different biological states are esterified with either d_0 - or d_3 -methanol; the two samples are combined and analysed by LC/MS-MS.

The MCAT strategy can not be considered an alternative to its ICAT counterpart, it is more adequate to describe the two strategies as complementary approaches to protein quantification. There are differences in details, yet the outcome of either approach gives a partial answer to the question of protein quantification in complex mixtures. This statement can be supported by the following considerations: both strategies are based on the modification of a specific amino acid, Cys in the case of ICAT and C-terminal Lys in the case of MCAT, to allow discrimination between two sister peptides derived from two different biological samples. In the first approach the alkylation of Cys results in 8 Da difference, while in the second case guanidination of Lys results in 42 Da increase. The use of MCAT alongside ICAT would overcome two main drawbacks of the latter approach, mainly analysis of Cys-free proteins and the detection of peptides which happen to contain post-translational modified Cys.

2.4 CONCLUDING REMARKS

Significant advances in protein identification have been accomplished over the last ten years through the use of mass spectrometry, and large-scale or even proteome-wide protein identification is now a common practice. The descriptive information obtained from such protein cataloguing projects can be significantly enhanced if the quantity and changes thereof can be determined precisely for each protein in a sample. The strategy described in the present method provides a simple and effective tool by which accurate protein quantitation can be achieved in proteome-wide experiments. The method also allows the determination of the absolute 2DE and mass spectrometry, two techniques established in most proteomics laboratories. Moreover Cahill *et al.* [20] validated our approach (d_0 - and d_3 - acrylamide-based) firstly by measuring the yield of proteins alkylated with AA, and secondly by using differential radioactive labels (^{125}I and ^{131}I) to quantitatively establish that non-comigration in 2D-PAGE is negligible.

The present method is not an alternative to existing strategies which rely on LC/MS-MS; however, it offers a number of advantages which can be exploited in the field of protein quantification: (a) This method uses relatively cheap and commercially available reagents which are commonly exploited in the solubilization cocktail prior to any electrophoretic step, this means that no additional steps are needed in the phase of sample preparation. (b) The use of 2-D gel in the present method requires more time compared to LC/MS of an entire digest; on the other hand the use of high resolution separation protocols offers two distinct advantages: it gives a more comprehensive picture of the protein content of the sample, simpler MS data to interpret and the possibility to analyse intact proteins, based on which a number of protein modifications can be spotted at an early stage of analysis. The ability to elute and analyse intact proteins from a 2-D map is, in fact, a unique advantage for those scientists seeking to attribute a post-synthetic modification to a specific protein zone. Such possibility is in reality lost (or greatly hampered) in fractionation strategies based on massive protein digestion prior to the separation process, such as those adopted in the ICAT and other protocols.

It is possible to anticipate that this method [22] will find wide application in the field of quantitative proteomics. For example Turko and Murad [21] used our method for quantifying protein expression in heart mitochondria from diabetic rats treated with streptozotocin.

2.5 BIBLIOGRAPHY

- [1] Sechi, S., *Rapid Commun. Mass Spectrom.* 2002, **16**: 1416-1424.
- [2] Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., Aebersold, R., *Nat. Biotechnol.* 1999, **17**: 994.
- [3] Griffin, T.J., Gygi, S.P., Rist, B., Aebersold, R., Loboda, A., Jilkine, A., Ens, W., Standing, K.G., *Anal. Chem.* 2001, **73**: 978.
- [4] Zhou, H., Ranish, J.A., Watts, J.D., Aebersold, R., *Nat. Biotechnol.*, 2002, **20**: 512.
- [5] Oda, Y., Huang, K., Cross, F.R., Cowburn, D., Chait, B.T., *Proc. Natl. Acad. Sci. USA* 1999, **96**: 6591.
- [6] Conrads, T.P., Alving, K., Veenstra, T.D., Belov, M.E., Anderson, G.A., Anderson, D.J., Lipton, M.S., Pasa-tolic, L., Udseth, H.R., Chrisler, W.B., Thrall, B.D., Smith, R.D., *Anal. Chem.* 2001, **73**: 2123.
- [7] Yao, X., Freas, A., Ramirez, J., Demirev, P.A., Fenselau, C., *Anal. Chem.* 2001, **73**: 2836.
- [8] Mirgorodskaya, O.A., Kozmin, Y.P., Titov, M.I., Korner, R., Sonksen, C.P., Roepstorff, P., *Rapid Commun. Mass Spectrom.* 2000, **14**: 1226.
- [9] Lueking, A., Horn, M., Eickhoff, H., Bussow, K., Lehrach, H., Walter, G., *Anal. Biochem.* 1999, **270**: 103.
- [10] Sechi, S., Chait, B.T., *Anal. Chem.* 1998, **70**: 5150.
- [11] Thannhauser, T.W., Sherwood, R.W., Scheraga, H.A., *J. Protein Chem.* 1998, **17**: 37.
- [12] Friedman, M., *J. Agric. Food Chem.* 1994, **42**: 3.
- [13] Came, A.F., Sally, U., *Methods Mol. Biol.* 1997, **64**: 271.
- [14] Bures, E.J., Hui, J.O., Young, Y., Chow, D.T., Katta, V., Rohde, M.F., Zeni, L., Rosenfeld, R.D., Stark, K.L., Haniu, M., *Biochemistry* 1998, **37**: 12172.
- [15] Galvani, M., Rovatti, L., Hamdan, M., Herbert, B., Righetti, P.G., *Electrophoresis* 2002, **22**: 2066.
- [16] Hamdan, M., Bordini, E., Galvani, M., Righetti, P.G., *Electrophoresis* 2001, **22**: 1633.
- [17] Hamdan, M., Bordini, E., Galvani, M., Righetti, P.G., *Electrophoresis* 2001, **22**: 1633.
- [18] Hamdan, M., Galvani, M., Righetti, P.G., *Mass Spectrom. Reviews* 2001, **20**: 121.
- [19] Goodlett, D.R., Keller, A., Watts, J.D., Newitt, R., Yi, E.C., Purvine, S., Eng, J.K., Von Haller, P., Aebersold, R., Kolker, E., *Rapid Commun. Mass Spectrom.* 2001, **15**: 1214.

[20] Cahill, M.A., Wozny, W., Schwall, G., Schroer, K., Hoelzer, K., Poznanovic, S., Hunzinger, C., Vogt, J.A., Stegmann, W., Matthies, H., Schrattenholz, A., *Rapid Commun. Mass Spectrom.* 2003, **17**: 1283-1290.

[21] Turko, I.V. and Murad, F., *J. Biol. Chem.*, 2003, **278**: 35844-35849.

[22] Gehanne, S., Cecconi, D., Carboni, L., Righetti, P.G., Domenici, E., Hamdan, M., *Rapid Commun. Mass Spectrom.* 2002, **16**: 1692-1698.

CHAPTER 3

NEW STATISTICAL TOOLS FOR COMPARATIVE PROTEOMICS (I): SPOT FUZZYFICATION AND MULTIDIMENSIONAL SCALING

3.1 INTRODUCTION

The comparison of 2-D maps from control and pathological (or treated) samples in order to identify proteins connected with the disease is an intricate process [1-7]. This complexity arises from:

- the complexity of the maps, which may contain thousands of spots;
- the low reproducibility of the 2D-maps, caused by the effect of several experimental factors on the final result (polymerisation conditions, sample pre-treatment, two-dimensional run conditions) [8];
- the small differences which often characterise samples from healthy individuals and diseased patients. Differences which often might be due not simply to the absence or presence of some proteins, but even to subtle changes in their relative amounts.

Today, the comparison of the 2-D maps is performed by different methods [9-11], all based on similar approaches. The first step of the analysis consists in the alignment of the maps, following a warping and matching process by shrinking or widening the spots with respect to the two dimensions. This step is based on the choice of some landmark spots defined by the user. Then each map is “cleaned”, by rejecting the spots which do not show a typical optical density pattern. This step allows the elimination of several spots which do not contain biochemical information, whose presence is caused by the effect of external experimental factors (e.g., staining artefacts). Once the alignment is obtained and the maps have been cleaned, they are matched to one another and the common information is identified.

In spite of all the problems connected with their comparison, 2-D maps are still very powerful for proteome analysis. It is evident that it is necessary to work with replicates of the same experiment, in order to take into account the experimental variability.

In the approach proposed here the comparison of different 2-D maps was approached with a different method:

- the comparison is performed on each replicate, not on a synthetic map;
- the spots in the maps are allowed to become fuzzy, with a broadening of the spots themselves, for simulating the experimental uncertainty in their position, shape and size;
- similarity indexes between the fuzzy maps are calculated;
- the matrix of similarities thus obtained is used as input in a Multidimensional Scaling (MDS) calculation;
- the co-ordinates of each map obtained from the multidimensional scaling are employed in a pattern recognition analysis which enables the clustering of different classes of samples, as belonging to control and diseased, or treated, tissues.

The proposed procedure is first analysed by using simulated maps in order to show its potential applications, then it is utilized for the analysis of real 2D maps.

3.2 THEORY

The method proposed for analysing 2D maps is based on two steps:

- a) the calibration of the method;
- b) its use for the classification of unknown samples.

The calibration produces a model that must be evaluated to check its ability to distinguish between the different classes of subjects. If the model performs satisfactorily the classification of the known samples (supervised analysis), it can be used for the assignment of new samples to the existing classes (usually diseased or treated and control patients).

The calibration phase consists of five steps:

1. *digitalisation* of each map with the creation of a rectangular grid of cells. The grid is matched on the 2D map and a signal varying from 0 (empty cell) to 1 (occupied cell) is assigned to each cell, depending on the degree of occupancy of the corresponding area of the 2D map;
2. *spot fuzzyfication*, *i.e.* a step where the spots are allowed to become more fuzzy so that the cells occupied by the spots can influence the signal of the near cells [12];
3. *calculation of a similarity matrix* (where each similarity is obtained from the matching of the corresponding maps) for the comparison of all the pairs of transformed 2D maps;
4. *multidimensional scaling analysis* of the similarity matrix;

5. *statistical analysis* of the co-ordinates of the 2D maps in the virtual space obtained from multidimensional scaling by cluster analysis and/or classification methods.

3.2.1 DIGITALISATION

A 2D map can be treated as a two-dimensional surface corresponding to a grid of a given step. The digitalisation of the map is obtained by a grid with a pass of 0.001×0.001 m, that produces a map of 200×200 cells.

Each 2D map is thus transformed into a matrix, corresponding to the grid. A 0 value is assigned to the cell corresponding to empty zones of the 2D-maps, a unit value to the area occupied by the spots. In the first application the percentage of black area of the cells (i.e. the degree of occupancy) is not taken into account, only 0 or 1 is used to distinguish between empty and occupied zones. An example of digitalisation is presented in figure 1.

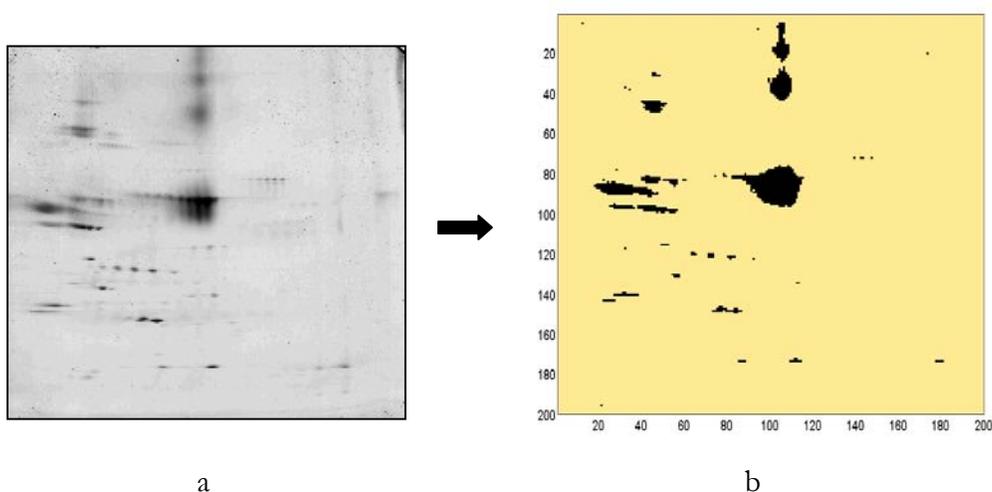


Figure 1. 2D map of sample HEA5 (a) and the correspondent digitalised image (b).

3.2.2 FUZZYFICATION

The spot fuzzyfication is very important since the position, size and shape of the spots on the maps may change in different experimental runs performed on the same biological sample. Moreover, different factors (such as temperature, contact time with development solutions, etc.) can contribute to the appearance of spurious spots not directly attributable to the protein content of the original

sample. This second step is very important since it allows to take into account the low reproducibility of the 2D-maps. In order to avoid this stiffness, we allowed each spot to become a fuzzy entity, thus having an influence on its neighbourhood. Each spot was turned into a two-dimensional probability density distribution, centered on the spot itself. The cells of the grid contain, after this treatment, a numerical value related to the degree of probability of finding a spot in the correspondent cell. The two-dimensional function that was chosen for this purpose was a two-dimensional gaussian function (figure 2).

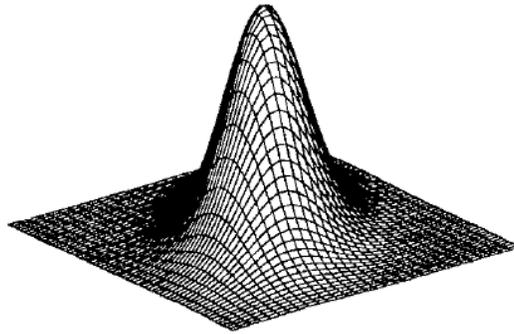


Figure 2. Bi-dimensional Normal distribution.

The influence of the presence of a spot in cell x_p, y_j on the neighbour cell x_k, y_l is calculated by the following two-dimensional gaussian function:

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

where

σ_x, σ_y = standard deviation of the gaussian function along each of the two dimensions;

ρ = correlation between the two dimensions x and y ; fixed at 0 (expected complete independence of the two electrophoretic runs).

In our approach the two parameters σ_x and σ_y are maintained identical, so that the gaussian function presents the same uncertainty with respect to both dimensions. This is based on the assumption of a similar variability of the two electrophoretic runs. Following this statement, the only parameter which shall be analysed for its effect on the final result is $\sigma = \sigma_x = \sigma_y$.

A change of the value of the parameter σ corresponds to a symmetrical change of the effect of each cell on its neighbouring cells (figure 3) on both electrophoretic axes: obviously, a large value for σ shall extend the effect of each spot at a larger distance, making the whole virtual map more fuzzy.

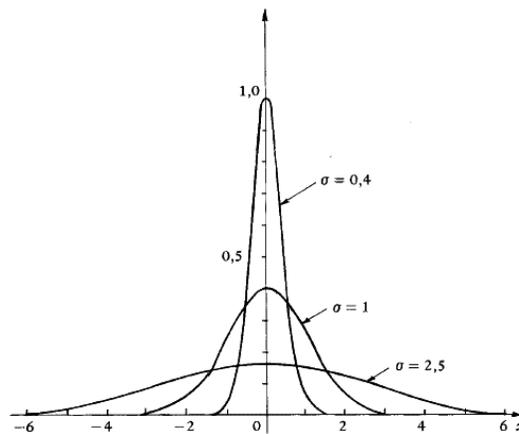


Figure 3. Normal probability density distribution as a function of the σ value.

The value of the signal S_k in each cell x_i, y_j of the virtual grid is given by the sum of the effects of all neighbour cells containing spots:

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

The summation runs on all cells of the grid, but in dependence on the value of the parameter σ , only the neighbour cells produce a significant effect. Each sample is then transformed in a virtual map containing in each cell the sum of the influence of all the spots of the original map. These virtual maps represent a sort of probability of the presence of a spot in each given cell of the grid, and are used for calculating the similarity matrix between every pair of samples.

In principle this method eliminates operator intervention to match spots or clean the 2D maps.

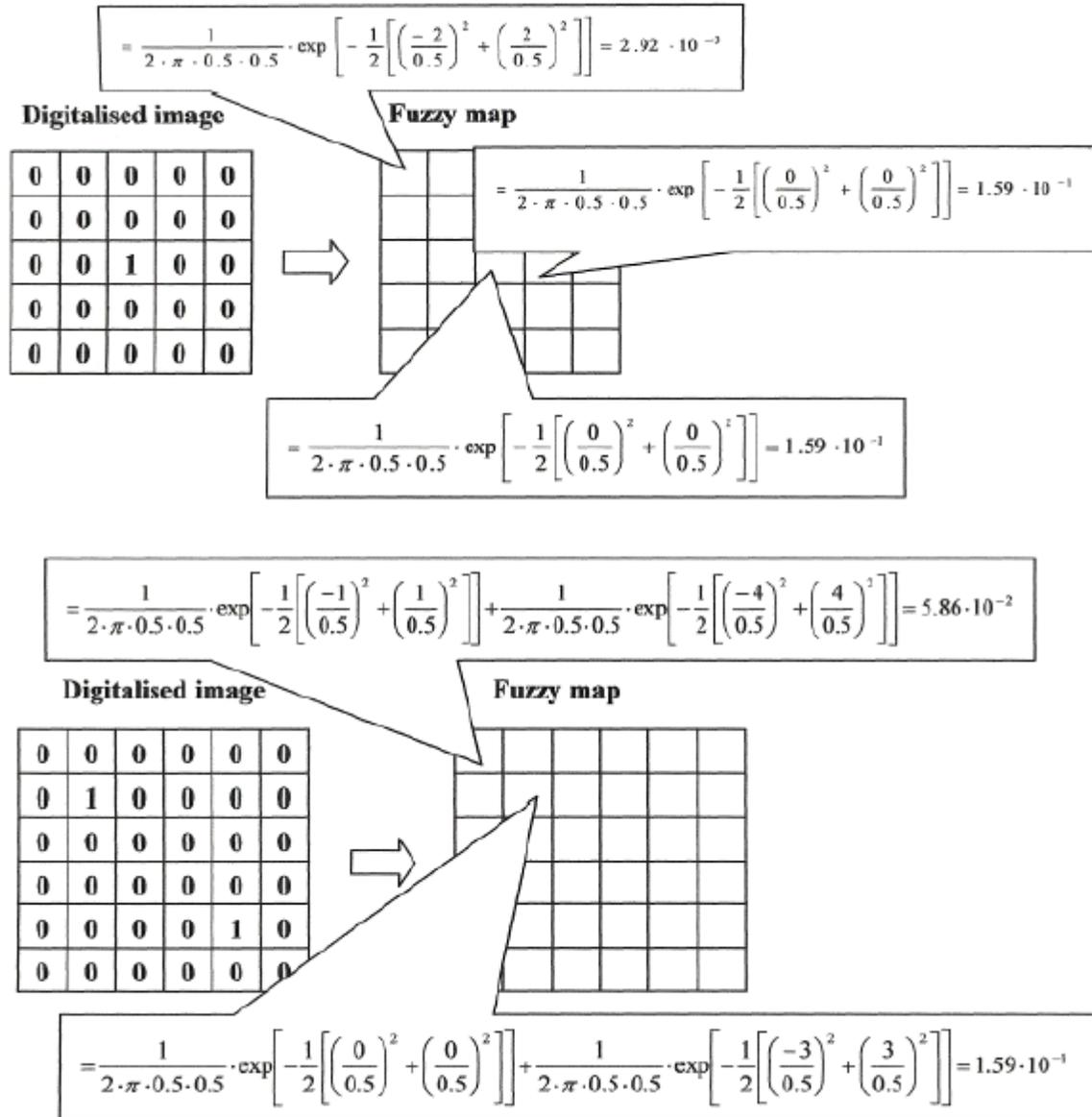


Figure 4. Examples of calculation of the fuzzy matrices with one and two spots in the 2D-map ($\sigma=1.0$).

3.2.3 SIMILARITY INDEX

The similarity between each pair of samples was calculated after performing a match of the potential matrices. The match of the two virtual grids k and l allows the computation of the common signal (SC_{kl}), namely the sum of all the signals present in both maps, and the total signal (ST_{kl}):

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

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

The similarity index can then be computed as the ratio between SC_{kl} and ST_{kl} :

$$S_{kl} = \frac{SC_{kl}}{ST_{kl}}$$

A similarity equal to 1 indicates that the two virtual maps are identical, since the common and total signals are identical. The lower the similarity index, the greater the difference between common and total signals. A value of 0 indicates that there is no match between the signals of the two maps, which corresponds to the minimum possible similarity. The similarity matrices were calculated by an algorithm programmed in Visual Basic 6.0 (Microsoft) and Matlab 6.1 (The Mathworks Inc.).

3.2.4 MULTIDIMENSIONAL SCALING

Multidimensional scaling (MDS) models are based on the idea that a set of ordinal data can be converted into a smaller amount of cardinal information.

Multidimensional scaling is a branch of multivariate data analysis geared towards dimensional reduction and graphical representation of data. Given a set of n objects and a measure of their similarity S_{ij} , Multidimensional Scaling (MDS) consists in the search for a low dimensional space in which the objects are represented by points in the space and such that the distances between the points match as much as possible with the original similarities of the objects [13]. The space is usually Euclidean, but this is not a constraint. The calculation of MDS from the 2D-map similarities was performed by using the Kruskal iterative method [14, 15]. The configuration of points obtained

at convergence of the iterative process, *i.e.* when changes of the points co-ordinates do not introduce any further decrements of the overall matching between real similarities and calculated distances, is considered as the final solution. The search for the co-ordinates is based on the steepest descent minimisation algorithm, where the target function is the so called Stress (S), which is proportional to the sum of squares of differences between calculated and real distances, *i.e.* a measure of the ability of the configuration of points to simulate the distance matrix. MDS calculations were performed by the STATISTICA software (Statsoft inc., Ver. 5.1).

3.2.5 STATISTICAL ANALYSIS

Once the co-ordinates have been obtained, they can be used for a multivariate statistical analysis. The results are often self-evident, so that a visual inspection is sufficient for obtaining the final results.

3.3 EXPERIMENTAL PROCEDURES

The Spot fuzzification and the Multidimensional scaling, of simulated and real samples, were performed by the Department of Environmental and Life Sciences Technologies staff, University of Eastern Piedmont.

The method proposed has been applied on simulated and real 2D maps, the effect of parameter σ was investigated at 5 levels: 0.50, 0.75, 1.00, 1.25, 1.50.

3.3.1 SIMULATED SAMPLES

The simulated 2D maps were used for evaluating the effect on the similarity indexes of the following parameters:

- changes in size and shape of a single spot (10 samples, figure 5a); DIM = dimension;
- changes in position of a single spot (10 samples, figure 5b); POS = position;
- the effect of two spots, with the largest that changes its size and shape (12 samples, figure 5c);
- the effect of two spots, with the smallest that changes its position (12 samples, figure 5d).

This set of 44 images was used to obtain the corresponding fuzzy maps for (40 x 40) grids. In this case we used (40 x 40) grids instead of the 200 x 200 grids adopted for the real samples, in order to decrease the computational effort.

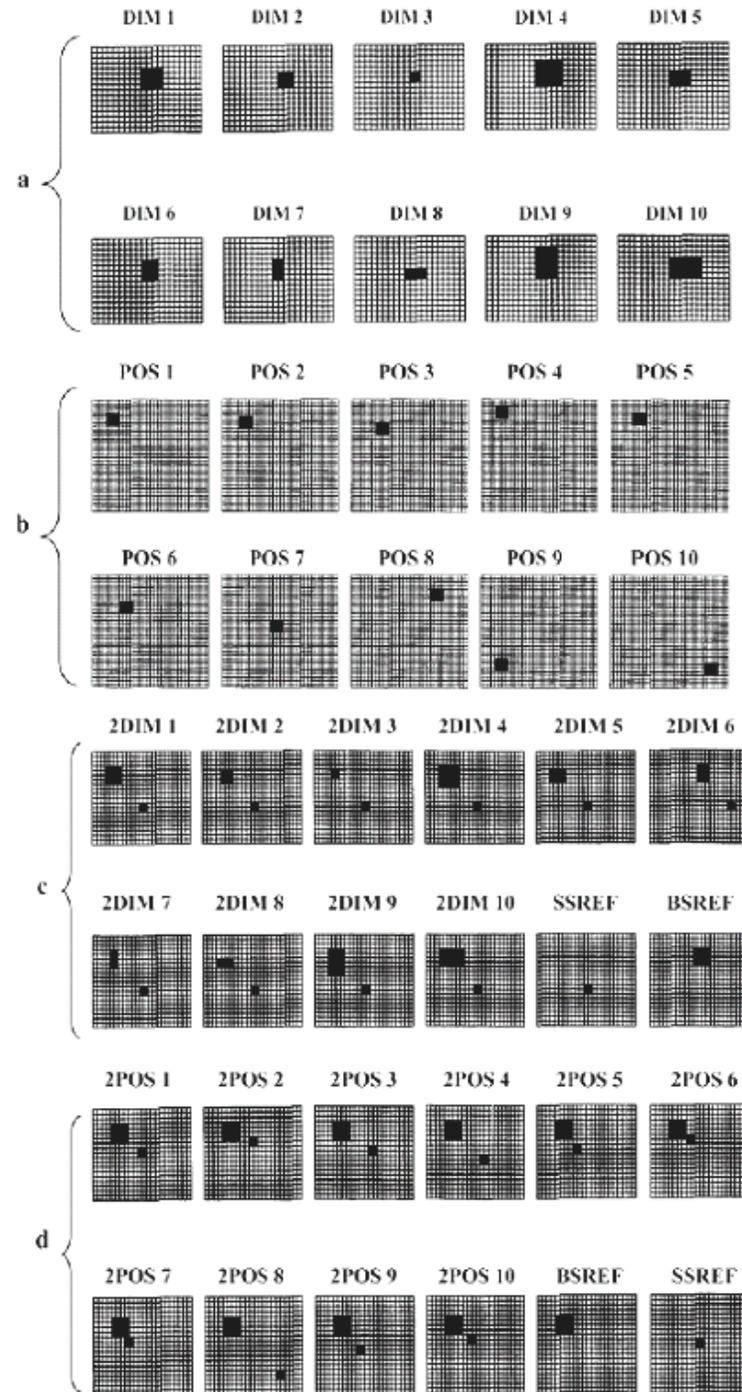


Figure 5. Simulated samples with changes in shape and size of a spot (a); changes in position of a spot (b); changes in shape and size of the largest of two spots (c); changes in position of the smallest of two spots (d).

3.3.2 REAL SAMPLES

The 2D maps used to test the computational method were divided into two groups:

- 1-5 HEA: 2D maps of five control rat serum samples;
- 1-5 ILL: 2D maps of five rat serum samples from nicotine-treated rats.

Figure 6 shows the ten 2-D maps.

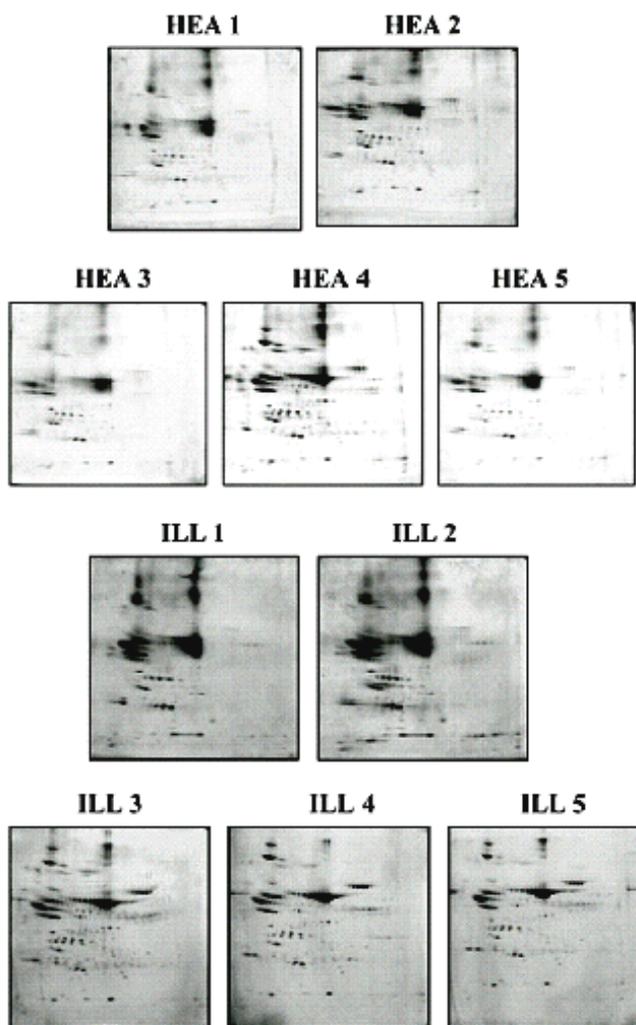


Figure 6. 2D-maps of control (HEA 1 - HEA 5) and nicotine-treated (ILL 1 - ILL 5) rat sera.

As it can be noticed, it is very difficult to distinguish the control 2D maps profile from the ones of serum from nicotine-treated rats by a visual inspection of the image.

In order to obtain the 2D maps presented in figure 6 five Wistar rats were treated for 14 days with a saline solution (control samples) and the other five were treated for the same 14 days with nicotine. The nicotine was administered subcutaneously by injecting 1ml/Kg of a 0.4 mg/ml nicotine solution.

Blood samples were collected on the 14th day (when it is known that nicotine administration begins to induce dependence on treated rats) on rats which were fasted for 12 hours prior to collection in order to avoid interferences due to high concentrations of lipids in the blood. All samples were centrifuged at 4°C to separate from each clot the serum samples (about 200 µL for each blood sample) and they were preserved at -20°C until the analysis was performed. One hundred µL of serum were added with 0.4 mL of a denaturing solution containing 7 M urea, 2 M thiourea, 5 mM TBP (tributylphosphine) and 40 mM Tris. 20 mM IAA (iodoacetamide) was then added and alkylation was continued for one hour. The samples were then submitted to dialysis in order to eliminate the salts present in sera and then the reagents eliminated by the dialysis process were added (7 M urea, 2 M thiourea and 20 mM Tris); 2% CHAPS (3-[(cholamidopropyl)dimethylammonium]-1-propane-sulfonate) was added as a surfactant.

Eighteen-cm long, pH 3-10 non-linear immobilised pH gradient strips (Amersham Pharmacia Biotech) were rehydrated for 8 h with 450 µL of the sample solution (final total protein concentration of 6 mg/mL) containing traces of bromophenol blue to monitor the electrophoretic run. Isoelectric focusing was conducted at 20°C for 60 000 Vh using a Protean IEF Cell (Bio-Rad Laboratories, Hercules, CA, USA) with a low initial voltage and then by applying a voltage gradient up to 10000 V with a limiting current of 50 µA/strip.

For the second dimension, the immobilised pH gradient 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), 4 mM TBP, 4% w/v acrylamide. The immobilised pH gradient strips were then laid on a 7-20% T gradient SDS-PAGE with 0.5% agarose in the cathode buffer [192 mM glycine (Sigma), 15 mM Tris, 0.1% SDS, pH 8.3]. The anode buffer consisted of 375 mM Tris-HCl pH 8.8. Gels were run at 10 °C and 2 mA/gel for 2 hour, 5 mA/gel for 1 hour and 10 mA/gel overnight. Gels were stained for 20 hours with colloidal Coomassie blue G-250 [0.1% Blue G-250 (BDH Laboratory Supplies, Poole, England), 34% v/v methanol, 3% v/v o-phosphoric acid, 17% w/v ammonium sulphate (Sigma)] and destained in 5% acetic acid (Sigma). Images were captured with a GS-710 imaging densitometer (Bio-Rad).

3.4 RESULTS AND DISCUSSION

The study of the effect of changes of the fuzzyfication parameters σ_x , σ_y was performed both on simulated and real samples. The larger the standard deviation along one of the electrophoretic directions, the larger the uncertainty of the spot position, shape and size along the corresponding direction. Figure 7 shows one of the real samples (HEA5) for all the levels chosen for σ_x and σ_y .

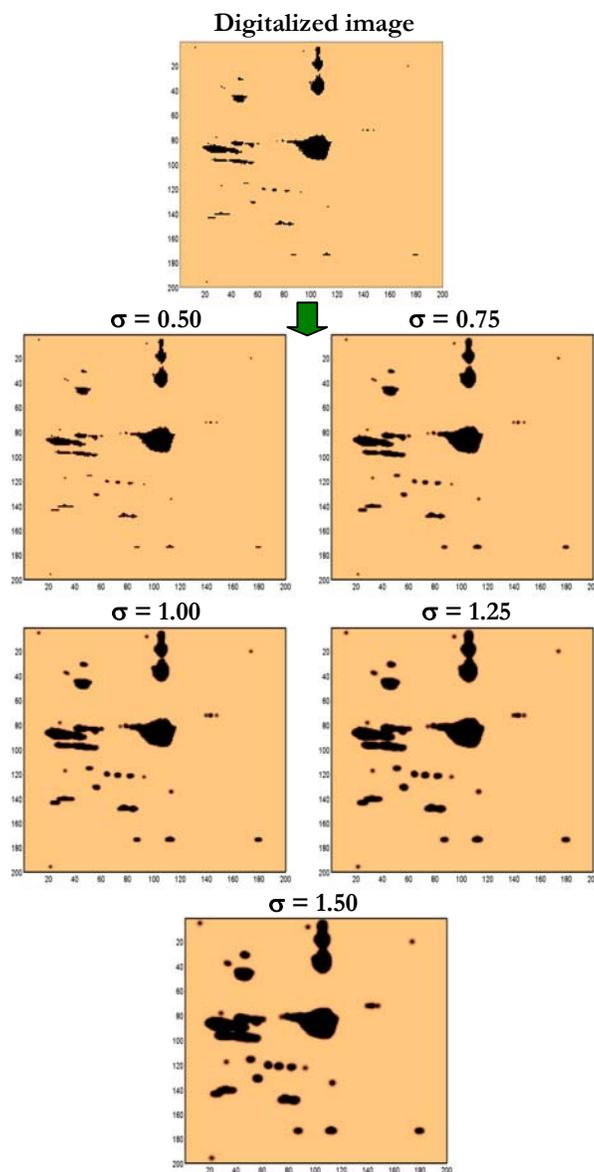


Figure 7. Fuzzy matrices of sample HEA5 at the five levels chosen for σ_x and σ_y .

It is evident that the spots increase their area of influence as the standard deviation increases. In this way it is possible to simulate the low reproducibility of the experimental results, taking into account the different reproducibility for each elution direction. In the real samples series, it is evident how the increase of the two σ values tends to mask the differences between different samples, with the nearest spots merging one into the other. The matching of the fuzzy maps was performed for each group of simulated and real samples and the similarity indexes obtained were analysed separately by multidimensional scaling. The final configuration provided by MDS for the first group of simulated samples, where a spot changes its size and shape, is reported in figure 8a. The result depends heavily on the size of the spot for all σ values. In this case the final coordinates are not very different in dependence on the different σ values, apart from a swap of DIM2 and DIM7 with σ greater than 0.75. The reference spot is the DIM1, which is in a central position in all the maps. All the other spots have been obtained by modifying the spot DIM1. The final configuration respects the observable differences of size and shape between the samples. So DIM5 and DIM6 are the nearest to DIM1, they have both been obtained from DIM1 by eliminating an entire row or column. In all instances, the farthest is DIM3, which represents the smallest spot. In the opposite side of the graph are located the largest spots, obtained from DIM1 by adding some rows and/or columns. The analysis of single spots seems to provide useful information, with a hierarchy of similarity based on the shape and size of the spots, which is reflected on their position in the scatter plots. The result of MDS for the second group, where a square spot changes its position, is reported in figure 8b. The lowest value of σ shows that 3 types of maps do exist: the POS9 and POS10 and a group which includes all other maps. The former two samples are completely different from the reference (POS1). In this case what we would expect is to obtain a representation of the spot centroids, since they do not change in size and shape, but only in their position. Therefore the MDS results should represent their shift along the space of the 2D-map. For low values of σ the method does not recognize that also sample POS8 is an outlier. This probably depends on the presence of some more diverse samples which obscure POS8 features. These highly diverse samples also affect the position of the other ones. Better results are obtained for larger values of σ , by which the 3 extreme samples appear well separated one from another. Moreover sample POS7 assumes a central position, which is in accordance with its equidistance from all other samples in the original map. By enlarging the zone where all other samples are concentrated, POS4 becomes correctly placed on the opposite side of the reference sample with respect to POS3 and POS6, while POS2 and POS5 remain very close

to the reference sample, in agreement with their relative position in the original map. Again the method has been able to provide the correct result. The lowest value of σ does not provide optimal results, which are instead obtained for the other values tested.

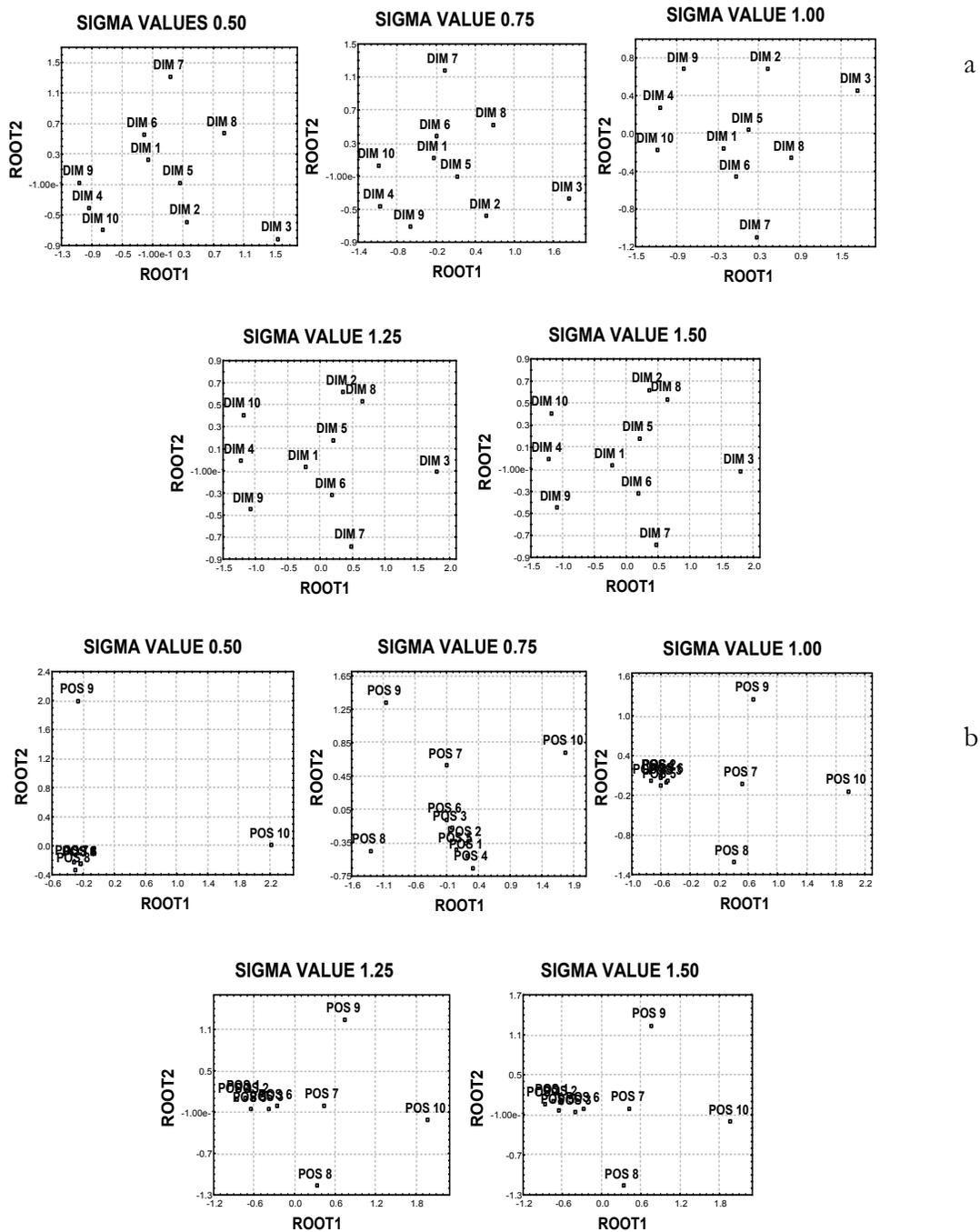


Figure 8. Final MDS configuration for the first (a) and the second (b) group of simulated samples for all σ values.

The third group of simulated samples contains two spots of different size where the larger spot changes size and shape. There is a slight change in the configuration of the points obtained from the MDS treatment (figure 9a). The group of samples showing higher similarity becomes more compact in the plot as the σ value increases. All the scatter plots show the presence of two outliers, represented by the two reference samples which contain only one of the two spots. The major distance is correctly shown by the sample where the larger spot is missing. The other samples stay apart and their position can be easily analysed zooming in the area. This allows the identification of the rules which determine the samples position:

- sample 2DIM3, where the largest spot is reduced in size is the most similar to the sample containing only the small spot;
- 2DIM10 sample, containing the largest large spot is the most similar to the sample containing only the large spot;
- the central position is occupied by sample 2DIM1, which is central since all other samples have been derived from it by changing the size and shape of the largest spot;
- the larger the distance from this sample, the larger the change in size of the spot.

Also in this case it is possible to sort the samples with respect to their similarity, by simply analysing their coordinates in the scatter plot. The low value of σ provides a less defined result. Probably the other configurations ($\sigma = 0.75, 1.0, 1.25, 1.5$) represent better solutions since the difference between the two “single spot” references and the others are better described.

In the case of the fourth group, where the position of the small spot is changed along the samples, there is not a large difference by varying the σ values (figure 9b). The sample containing only the small spot stays aside and represents an outlier. The other reference, where there is only the large spot, is contained among the others, probably because of the higher “weight” of the large spot in the calculation. In all the scatter plots, the x-axis (root 1) takes into account the change in position of the small spot. The sample 2POS1 is in a central position and represents the starting sample for all position changes. The most distant samples are 2POS4 and 2POS3 on one side; these are indeed the most different samples with respect to 2POS1. On the other side, the most distant samples are 2POS5 and 2POS7 where the small spot has moved in the opposite direction with respect to 2POS4 and 2POS3.

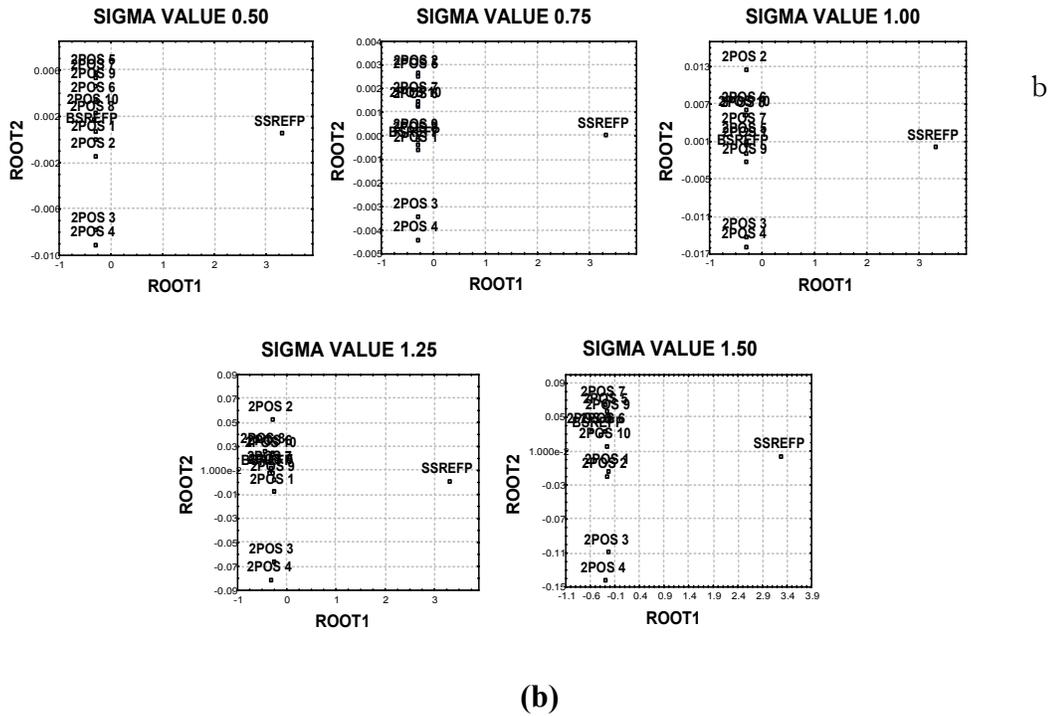
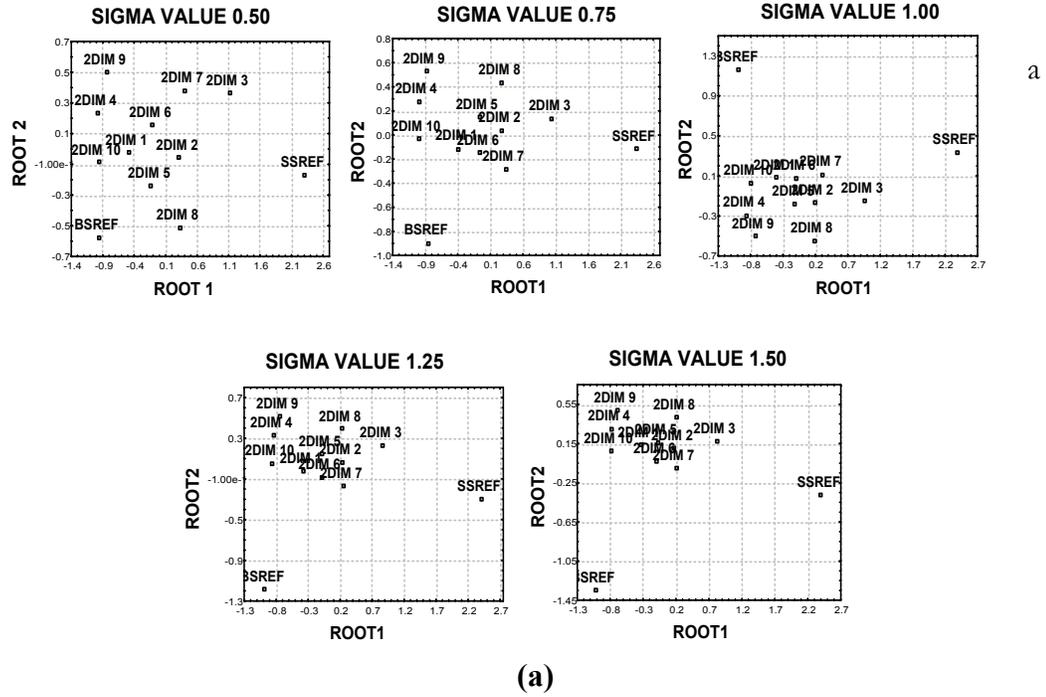


Figure 9. Final MDS configuration for the third (a) and the fourth (b) group of simulated samples for all σ values.

The last calculation was performed on the real samples (figure 10).

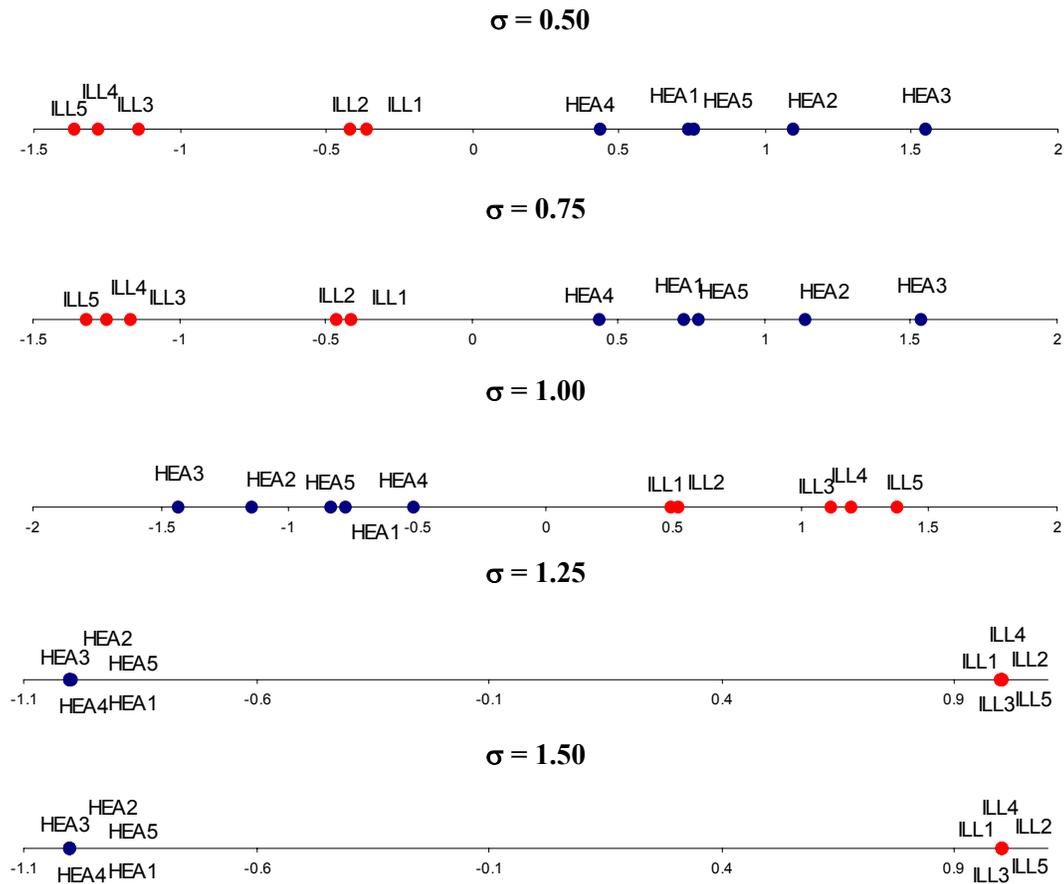


Figure 10. Final MDS configuration for the real samples for all σ values.

From figure 6 it is evident that the 2D-maps are very different one from each other and it would be very difficult, from a simple visual inspection, to state their similarity pattern. MDS provided a clear scheme which changes slightly passing from $\sigma = 0.5$ to the other values. In this case the separation of the two classes of rats, is obtained with all the σ values. In all the cases, it can be observed that it is possible to separate the two classes by mean of only one dimension. Even if the separation is successful for all the σ values considered, it is more effective for the large ones. The best value of the fuzzyfication parameter seems to be $\sigma = 1.25$, which allows an optimal separation of the two classes of samples.

3.5 CONCLUDING REMARKS

A new method for the statistical analysis of sets of 2-D maps, in proteome research, has been developed [16]. The method involves several steps, map digitalisation, fuzzyfication, calculation of a similarity matrix and a multidimensional scaling analysis, and has been applied satisfactorily to the analysis of both simulated and real samples. The analysis of the simulated maps allows an exploration of the effect of the σ parameter, *i.e.* the effect of changing the uncertainty along each electrophoretic direction. The application of the method to a complex dataset constituted by several 2D maps of sera from rats treated with nicotine and controls has shown that this method allows the discrimination between the two classes.

The final aim of the method based on MDS is the diagnosis of the patient condition from a global analysis of a 2D map, so no information is being acquired on the spots responsible for its classification.

3.6 BIBLIOGRAPHY

- [1] Frey, S.J., Larsen, P.M., *Curr. Opin. Biotechnol.* 2001, **5**: 26-33.
- [2] Patton, W.F., *J. Chromatogr. A* 1995, **698**: 55-87.
- [3] Yamaoka, T., *Anal. Chim. Acta* 1998, **372**: 91-98.
- [4] Celis, J.E., *et al.*, *FEBS Lett.* 1996, **398**: 129-34.
- [5] Wilkins, M., *TIBTECH* 2000, **18**: 91-92.
- [6] Naaby-Hansen, S., Waterfield, M.D., Cramer, R., *Trends Pharmacol. Sci.* 2001, **22**: 376-84.
- [7] Patterson, S. D., *Curr. Opin. Biotechnol.* 2000, **11**: 413-18.
- [8] Gianazza, E., Astrua-Testori, S., Caccia, P., Giaccon, P., Quaglia, L., Righetti, P.G., *Electrophoresis* 1986, **7**: 76-83.
- [9] Hoffmann, F., Kriegel, K., Wenk, C., *Discrete Appl. Math.* 1999, **93**: 75-88.
- [10] Lee, K. H., *Trends Biotechnol.* 2001, **19**: 217-22.
- [11] Vihinen, M., *Biomol. Eng.* 2001, **18**: 241-48.
- [12] Marengo, E., Robotti, E., Gianotti, V., Righetti, P.G., *Ann. Chim.* 2003, **93**:105-106.
- [13] Tulp, A., Verwoerd, D., Neefjes, J., *J. Chromatogr. B* 1999, **722**: 141-51.
- [14] Shepard, R.N., *Psychometrika* 1962, **27**: 219-46.
- [15] Kruskal, J.B., *Psychometrika* 1964, **29**: 1-27.
- [16] Marengo, E., Robotti, E., Gianotti, V., Righetti, P.G., Cecconi, D., Domenici, E., *Electrophoresis* 2003, **24**: 225-236.

CHAPTER 4:

NEW STATISTICAL TOOLS FOR COMPARATIVE PROTEOMICS (II): THREE WAY PRINCIPAL COMPONENT ANALYSIS

4.1 INTRODUCTION

In order to develop a new statistical tool for comparative proteomics, in this section three-way Principal Component Analysis has been applied to 2D maps to permit the identification of the different classes of samples present in a dataset. The developed method has been applied to 2D maps of rat sera, constituted by 5 samples of healthy Wistar rat sera and 5 samples of nicotine-treated Wistar rat sera. The method proved to be successful in the identification of the classes of samples present in both the groups of 2D-PAGE images and it allowed to identify the regions of the two-dimensional maps responsible for the differences occurring between the two classes of rat sera samples. Some studies already present in the literature concern the development of methods for the classification of images based on pixel-data [1]; Principal Component Analysis (PCA) [2-5] has been applied to the study of DNA and RNA fragments of several biological systems [6-9] and to the characterisation of proteomic patterns of different classes of tissues [10-15]. These multivariate methods require the previous analysis of the proteomic pattern images by standard softwares (i.e. Melanie III or PD-Quest) to identify the spots, so they present the disadvantage of being submitted to human choice for maps alignment. To avoid this disadvantage, in the approach presented here, all the maps belonging to the same type of sample are maintained during the analysis and the comparison is performed on each sample. In this way, the information about the variability and the reproducibility of the 2D-PAGE maps is maintained and the comparison is performed on all the real samples together and not on “synthetic” maps.

The 2D-PAGE images are first digitalised, transforming each image in a grid containing in each cell the value of the average optic density revealed in the correspondent area of the map. In this way, a three-mode dataset is obtained, with the three modes being the isoelectric point, the molecular mass

and the samples respectively. The digitalised images thus obtained are then analysed by three-way PCA.

4.2 THEORY

The method applied for the comparison of 2D-PAGE maps belonging to different classes, consists of three steps:

1. *digitalisation of the image*. Each image is digitalised producing a grid of (50x50) cells: each cell of the grid contains a value ranging from 0 to 1, according to the intensity of the signal in the correspondent position.
2. *data transformation*. Data transformations have to be applied to the dataset in order to scale all the samples and make them comparable; maximum scaling was chosen as the most suitable for the dataset under investigation.
3. *three-way PCA*. This multivariate technique is applied to the digitalised and scaled images in order to identify the classes of samples present in the dataset and to identify the zones of the maps responsible of the differences occurring between the classes. The maps rebuilt by using the relevant factors are compared in order to identify the differences occurring between the centroids of the two classes of healthy and pathological protein patterns.

4.2.1 DIGITALISATION

A 2D-PAGE slab appears as a transparent polymeric matter with the separated proteins spread all over it as coloured spots. The revelation of the spots is performed, in general, with a solution of an organic colouring agent (Coomassie Blue) or by deposition of silver on the protein surface: the use of these staining solutions generates a colouring intensity which is proportional (within given limits) to the protein concentration. Each 2D-PAGE map is previously scanned with a GS-710 densitometer (Bio-Rad), which transforms the 2D-PAGE in an image (200 x 200 pixels) in which each pixel corresponds to the value of the average optic density in the correspondent area.

The scanned images are then transformed in a grid of 50 x 50 cells, in which each cell contains a value ranging from 0 to 1. Each numerical value corresponds to the colour intensity of the image calculated by averaging the intensities of the pixels which are contained in the correspondent cell. The values smaller than 0.4 were cut off and substituted with null values in order to eliminate the information about the colour intensity of the background.

The choice of a 50 x 50 grid is not a constraint but it was suggested by computational and memory requirements.

4.2.2 DATA TRANSFORMATION

A normalisation 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 digitalised 2D-PAGE maps are scaled one at a time to the maximum value for each map, according to the following mathematical expression:

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

where $x_k(i, j)$ = value of the cell in (i, j) position in the k -th 2D map;

and $\max(x_k)$ = the maximum value in all the cells of the k -th 2D map.

By applying such a transformation to each two-dimensional map, the maximum signal intensity value of every 2D-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 the variation.

4.2.3 THREE-WAY PRINCIPAL COMPONENT ANALYSIS

Three-way Principal Component Analysis (Three-way PCA), based on the Tucker-3 model [16-19], has been used for the identification of the classes of samples present in the 2D maps dataset of rat sera. The interest of three-way PCA is that it allows to take into account the three-way structure of the data set 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. molecular mass), J is the number of columns of the grids (the y coordinates, i.e. pH), and K is the number of samples. The three-way PCA is based on the fact that the observed modes I , J , 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:

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

where x_{ijk} = denotes the elements of the initial matrix \mathbf{X} ,

a_{ip} , b_{jq} and c_{kr} = denote reduced elements of the component matrices \mathbf{A} , \mathbf{B} and \mathbf{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 \mathbf{G} ,

e_{ijk} = error term for element x_{ijk} and is an element of the $I \times J \times K$ array \mathbf{E} ²⁹⁻³⁵.

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, by looking for the common orientation for which the core array is as much as possible body-diagonal. If this condition is sufficiently achieved, i.e., if the elements g_{111} , g_{222} ... are the only elements of the core matrix being significantly different from 0, then the rotated sets of loadings can also be interpreted jointly by overlapping them.

4.3 EXPERIMENTAL PROCEDURES

4.3.1 SAMPLE PREPARATION

The investigated dataset consists of 10 samples belonging to two different groups:

5 samples of Wistar rat serum pertaining to healthy individuals;

5 samples of Wistar rat serum belonging to nicotine treated individuals.

The 10 2D-PAGE maps obtained are represented in figure 1. Looking at the 2D-PAGE images, it is not very easy to distinguish the healthy individuals from the nicotine treated ones by a visual inspection of the image. Moreover, the 2D-maps belonging to the same group show a large variability of the spots number, position, shape and size.

In order to obtain the samples for the two dimensional electrophoresis analysis five Wistar rats were treated for 14 days with a saline solution (control samples) and the other five were treated for the same 14 days with nicotine. The nicotine was administered subcutaneously by injecting 1ml/Kg of a

0.4 mg/ml nicotine solution. Blood samples were collected on the 14th day (when it is known that nicotine administration begins to induce dependence on treated rats) on rats which were fasted for 12 hours prior to collection in order to avoid interferences due to high concentrations of lipids in the blood. All samples were centrifuged at 4°C to separate from each clot the serum samples (about 200 µL for each blood sample) and they were preserved at -20°C until the analysis was performed. One hundred µL of serum were added with 0.4 mL of a denaturing solution containing 7 M urea, 2 M thiourea, 5 mM TBP (tributylphosphine) and 40 mM Tris. 20 mM IAA (iodoacetamide) was then added and alkylation was continued for an hour. The samples were then submitted to dialysis in order to eliminate the salts present in sera and then the reagents eliminated by the dialysis process were restored (7 M urea, 2 M thiourea and 20 mM Tris); 2% CHAPS (3-[[cholamidopropyl]dimethylammonium]-1-propane-sulfonate) was added as a surfactant.

Eighteen-cm long, pH 3-10 non-linear immobilised pH gradient strips (Amersham Pharmacia Biotech) were rehydrated for 8 h with 450 µL of the sample solution (final total protein concentration of 6 mg/mL) containing traces of bromophenol blue to monitor the electrophoretic run. Isoelectric focusing was conducted at 20°C for 60 000 Vh using a Protean IEF Cell (Bio-Rad Laboratories, Hercules, CA, USA) with a low initial voltage and then by applying a voltage gradient up to 10000 V with a limiting current of 50 µA/strip.

For the second dimension, the immobilised pH gradient 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), 4 mM TBP, 4% w/v acrylamide. The immobilised pH gradient strips were then laid on a 7-20% T gradient SDS-PAGE with 0.5% agarose in the cathode buffer [192 mM glycine (Sigma), 15 mM Tris, 0.1% SDS, pH 8.3]. The anode buffer consisted of 375 mM Tris-HCl pH 8.8. Gels were run at 10 °C and 2 mA/gel for 2 hour, 5 mA/gel for 1 hour and 10 mA/gel overnight. Gels were stained 20 hours with colloidal Coomassie blue G-250 [0.1% Blue G-250 (Sigma), 34% v/v methanol, 3% v/v o-phosphoric acid 17% w/v ammonium sulphate (Sigma)] and destained in 5% acetic acid (Sigma). Images were captured with a GS-710 imaging densitometer (Bio-Rad).

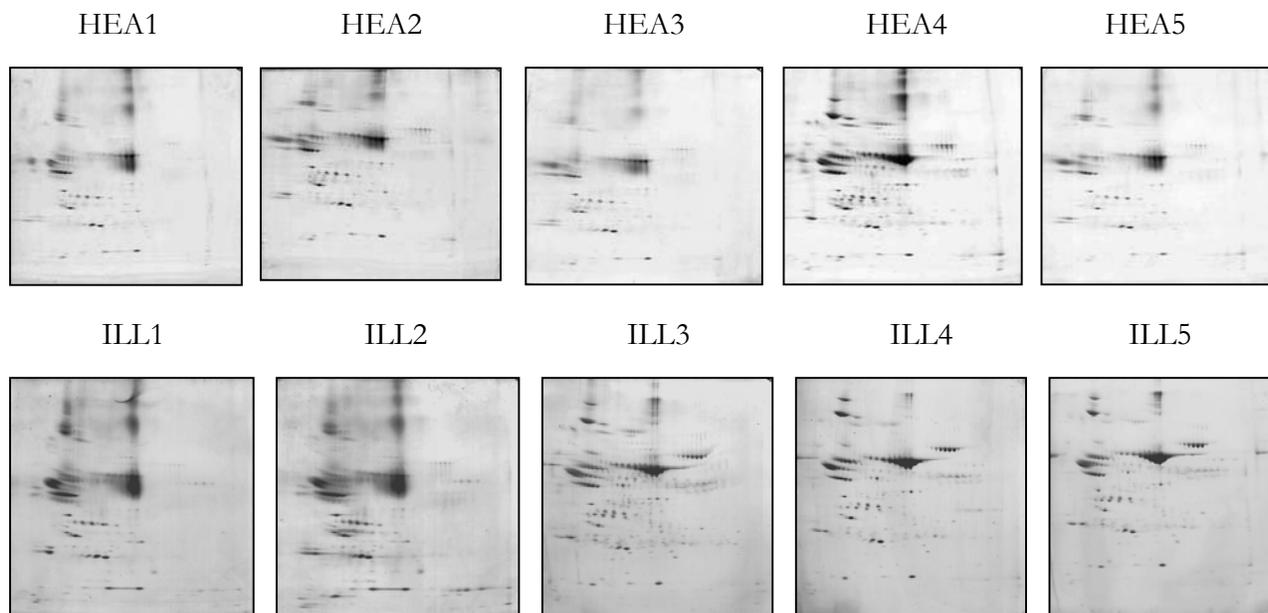


Figure 1. 2D-PAGE maps of the 10 rat serum samples.

4.4 RESULTS AND DISCUSSION

Data transformation and Three-way PCA were performed by the Department of Environmental and Life Sciences Technologies staff, University of Eastern Piedmont.

4.4.1 DATA TRANSFORMATION

When looking at the original 2D-PAGE images in Figure 1, sample HEA2 appears shifted with respect to the other samples of the same class. This peculiarity of sample HEA2 would lead to an incorrect analysis with three-way PCA because the major amount of explained variance would be probably led by this characteristic. Sample HEA2 was then shifted and matched to the other samples belonging to the same class.

Before performing 3-way PCA, a maximum scaling procedure was applied to the dataset; this scaling technique produces some changes in the dataset: in the maximum-scaled patterns, some minor spots can be detected with respect to the original ones; in addition, maximum scaling is fundamental for obtaining a complete independence of the analysis from the staining step, often scarcely reproducible.

4.4.2 THREE-WAY PCA ON THE NORMALISED DATASET

Three-way PCA performed on the normalized dataset gave the results represented in Figures 2 and 3 (two factors were retained for each mode); samples from 1 to 5 correspond to the control rat serum samples, while samples from 6 to 10 correspond to the nicotine-treated rat sera. The first two factors explain 59.8% of the total variance.

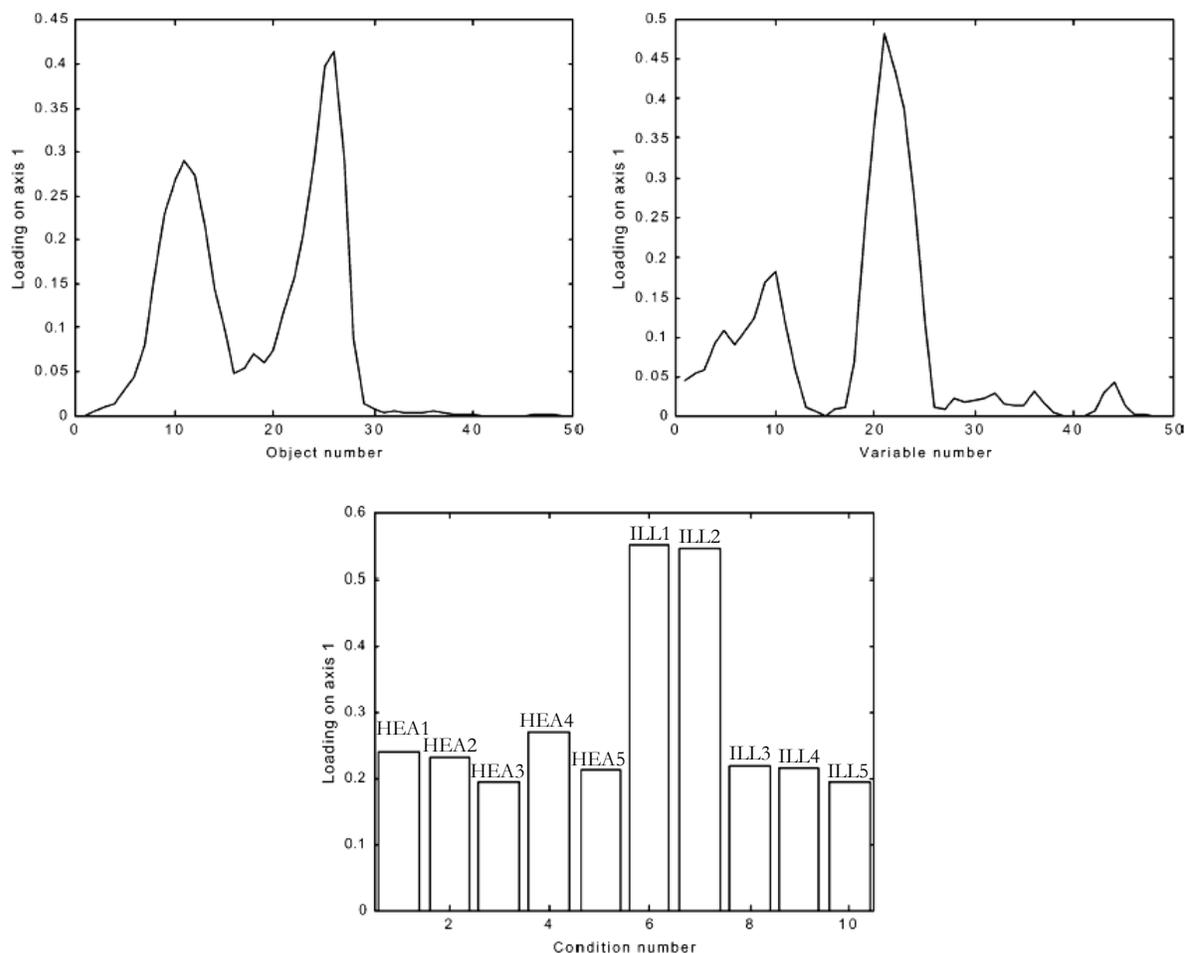


Figure 2. Plots of the three modes for the first entry of the core matrix g_{111} .

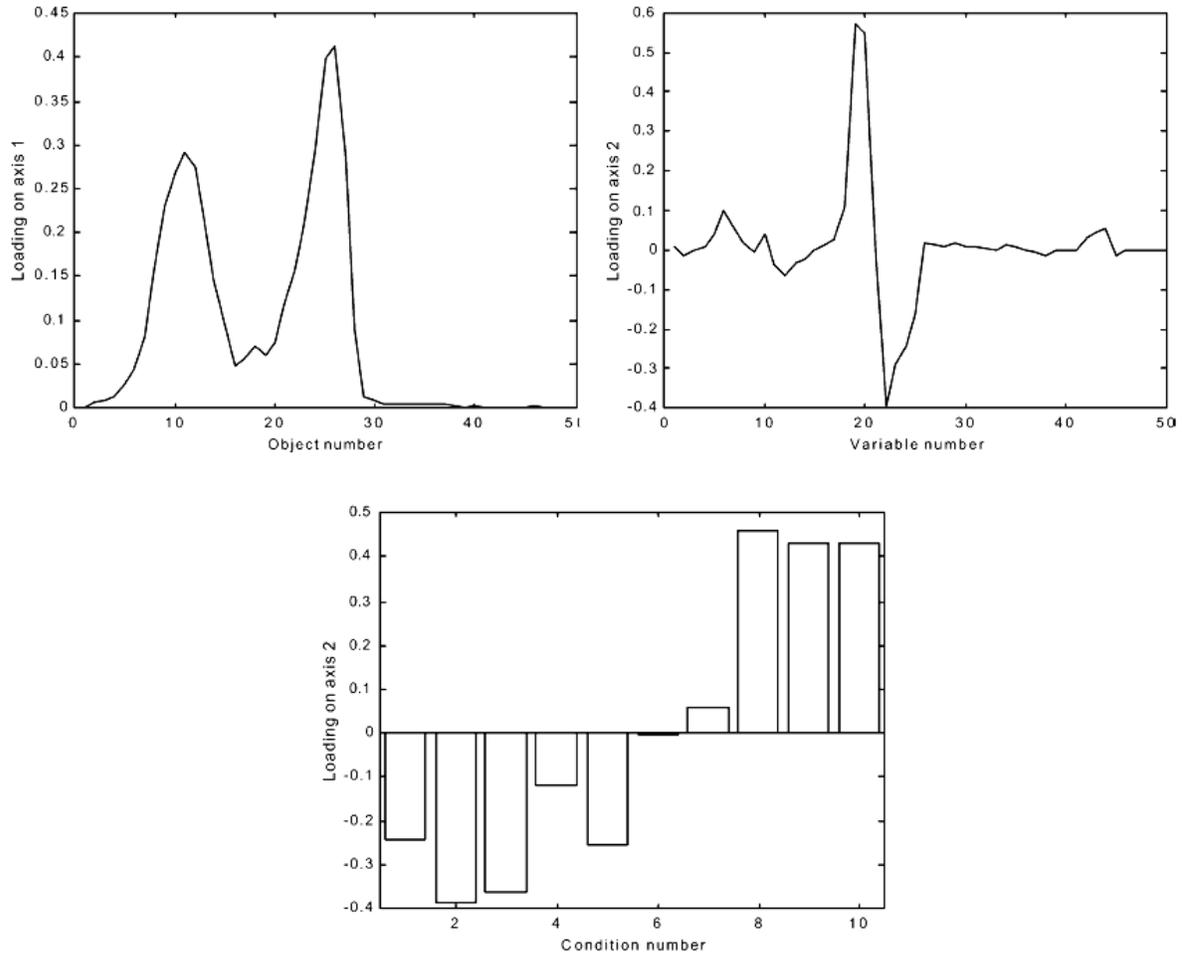


Figure 3. Plots of the three modes for the second entry of the core matrix g_{222} .

After body diagonalization [19] the following core matrix is obtained; the cubic core matrix is reported according to the following unfolding

$$\begin{bmatrix} g_{111} & g_{121} & g_{112} & g_{122} \\ g_{211} & g_{221} & g_{212} & g_{222} \end{bmatrix}$$

$$\begin{bmatrix} 12.76 & -0.22 & 0.13 & 5.10 \\ 0.61 & 0.99 & -2.44 & 1.51 \end{bmatrix}$$

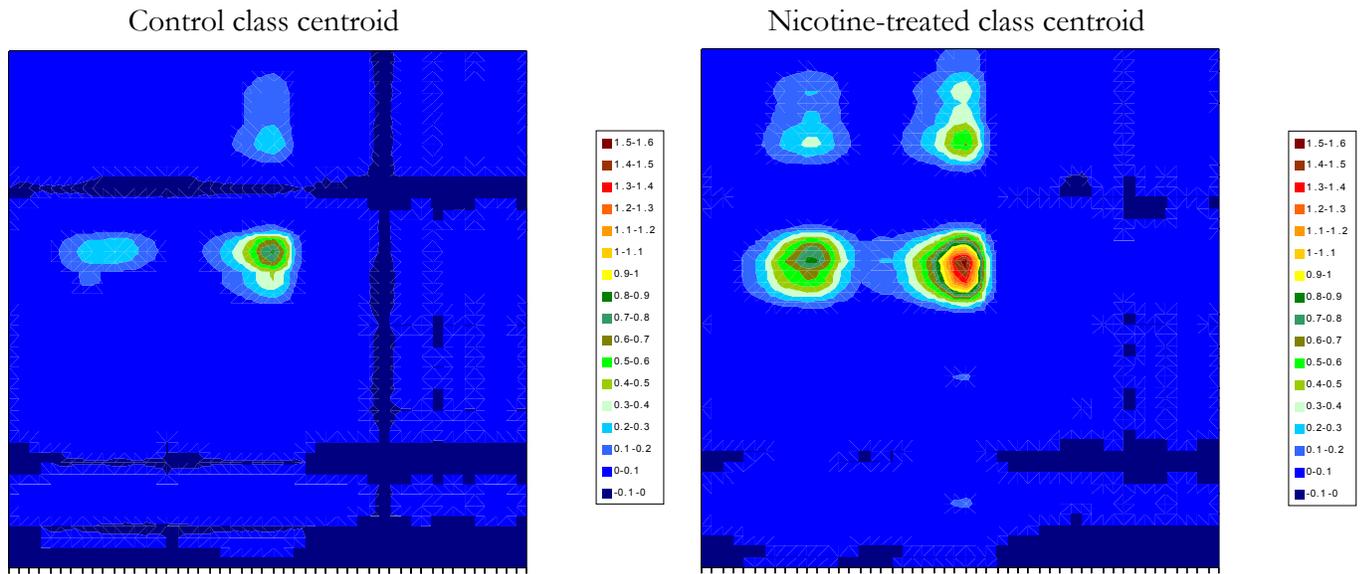
Since it is not super-diagonal, the plots of objects, variables and conditions can be interpreted jointly only for the first factor (the term g_{111} being by far the largest one).

The analysis is performed on one series of combinations at a time: this corresponds to examining one entry of the core matrix at a time. The first one, g_{111} , explains 51.6% of the total variance. The three plots, one for each mode, are represented in Figure 2. The first axis of the mode of conditions clearly discriminates samples 6 and 7 (ILL1 and ILL2) from the others; looking at the plots of the other two modes, it is possible to state that these two samples are characterized by larger values of the objects around 11 and 26 and of the variables around 21 and, to a lesser extent, around 10. By looking at the maps, it is easy to see that these samples are characterized by very intense spots in the rows around 11 and 26 (corresponding to pH 4.5-5 and 6-7) and in the columns around 21 (molecular mass 100-50 KDa).

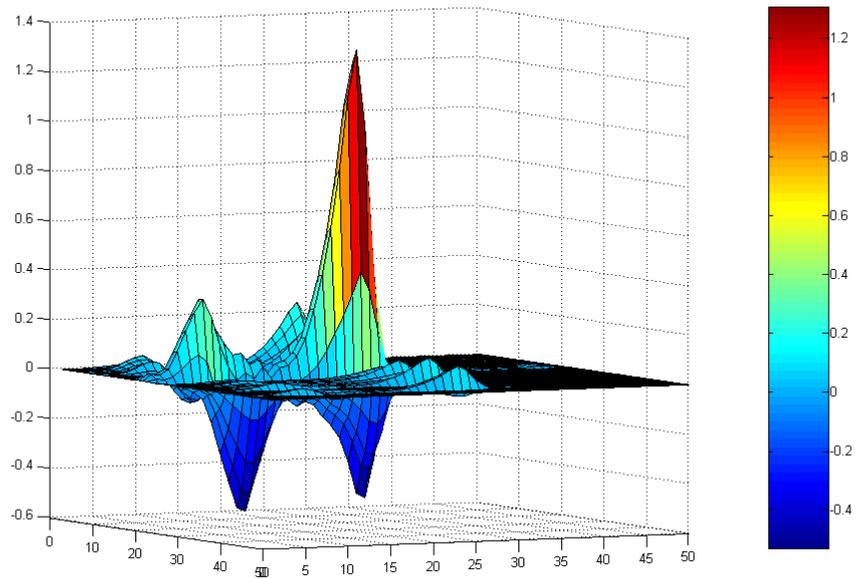
4.4.3 DIFFERENCE ANALYSIS

The fundamental aim of the present study is to identify the regions of the maps responsible for the discrimination of the two classes of samples: the control and the nicotine-treated ones. The identification of the differences occurring between the two classes of samples is performed by using the centroids of each class, in the space defined by variables and objects obtained from three-way principal component analysis. These centroids represent the average information concerning control and treated individuals contained in the first two factors which permit the class discrimination. The centroids can be re-projected in the original space, thus obtaining the corresponding 2D-map images containing only the information accounted for by the first two 3-way factors. The images rebuilt in this way can be compared for identifying the discriminant regions of the 2D-maps. This procedure allows a sort of filtering of the useful (discriminant) information contained in the 2D-maps.

The centroid of the control class is obtained on the basis of all the 5 original maps belonging to this class, while the centroid of the nicotine-treated class is calculated on the basis of only the three samples (ILL3, ILL4, ILL5) which appear the most different from the control class, as pointed out by three-way PCA. The two re-projected maps and their difference are represented in figure 4(a). In figure 4(b), the positive values (towards red) refer to regions which characterise the control sample, while the negative ones (towards blue) refer to regions which characterise the treated specimens. The control samples appear thus richer in spots than the nicotine-treated ones and the two classes show differences due both to the presence/absence of spots and to different relative intensities of the spots.



(a)



(b)

Figure 4. Contour plots of the centroids for the control and nicotine-treated classes (a) and the correspondent map of the differences (b).

4.5 CONCLUDING REMARKS

Three-way PCA was performed on 10 samples of rat sera. Maximum scaling was applied to each dataset before performing three-way PCA in order to eliminate the influence of the staining procedure on the statistical analysis.

The applied method turned out to be a successful tool for the discrimination of the classes of samples present and for the identification of the zones responsible for the differences occurring between the samples belonging to the different classes. This last goal was realised by “differences analysis”, which allowed the identification of the regions that characterise each class of samples for both the considered dataset. It is necessary to stress that this method [20] represents a preliminary approach to the problem of comparing 2D-PAGE maps belonging to different classes: further studies are necessary before being able to use this method for diagnostic/prognostic purposes.

4.6 BIBLIOGRAPHY

- [1] Swets, D.L., Weng, J.J., *IEEE T. Pattern Anal.* 1996, **18**: 831-836.
- [2] Davis, J.C., *Statistics and Data Analysis in Geology*; John Wiley & Sons: New York, 1986.
- [3] Brereton, R.G., *Chemometrics: Application of Mathematics and Statistics to Laboratory Systems*; Ellis Norwood: New York, 1990.
- [4] Box, G.E.P., Hunter, W.G., Hunter, J.S., *Statistics for Experimenters*; Wiley: New York, 1978.
- [5] Massart, D.L., Vandeginste, B.G.M., Deming, S.N., Michotte, Y., Kaufman, L., *Chemometrics: a Textbook*; Elsevier: Amsterdam, 1988.
- [6] Leonard, R.B., Mayer, J., Sasser, M., Woods, M.L., Mooney, B.R., Brinton, B.G., Newcombgrayman, P.L., Carroll, K.C., *J. Clin. Microbiol.* 1995, **33**: 2723-2727.
- [7] Johansson, M.L., Quednau, M., Ahrne, S., Molin, G., *Int. J. Syst. Bacteriol.* 1995, **45**: 670-675.
- [8] 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.
- [9] Boon, N., De Windt, W., Verstraete, W., Top, E.M., *FEMS Microbiol. Ecol.* 2002, **39**: 101-112.
- [10] Aliko, J.E., AkenOva, M.E., Fatokun, C.A., *Genet. Res. Crop Evol.* 1995, **42**: 393-399.
- [11] Anderson, N.L., EsquerBlasco, R., Richardson, F., Foxworthy, P., Eacho, P., *Toxicol. Appl. Pharm.* 1996, **137**: 75-89.
- [12] Vohradsky, J., *Electrophoresis* 1997, **18**: 2749-2754.
- [13] Kovarova, H., Radzioch, D., Hajduch, M., Sirova, M., Blaha, V., Macela, A., Stulik, J., Hernychova, L., *Electrophoresis* 1998, **19**: 1325-1331.
- [14] 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.
- [15] Dewettinck, K., Dierckx, S., Eichwalder, P., Huyghebaert, A., *Lait* 1997, **77**: 77-89.
- [16] Tucker, L.R., *Psychometrika* 1966, **31**: 279-311.
- [17] Kroonenberg, P.M., *Three-Mode Principal Component Analysis*; DSWO Press: Leiden, 1983.
- [18] Geladi, P., *Chemometr. Intell. Lab.* 1989, **7**: 11-30.
- [19] Kroonenberg, P.M., *Chemometr. Intell. Lab.* 1989, **15**: 143-157.
- [20] Marengo, E., Leardi, R., Robotti, E., Righetti, P.G., Antonucci, F., Cecconi, D., *J. Proteome Res.* 2003, **2**: 351-360.

CHAPTER 5

PANCREATIC ADENOCARCINOMA: PROTEOMICS ANALYSIS OF EPIGENETIC EVENTS

5.1 PANCREATIC ADENOCARCINOMA

Pancreatic adenocarcinoma is the fourth leading cause of cancer death in the United States [1]. With a 5-year survival rate of only 3% and a median survival of less than 6 months, a diagnosis of pancreatic adenocarcinoma carries one of the most dismal prognoses in all of medicine [2]. Due to a lack of specific symptoms and limitations in diagnostic methods, the disease often eludes detection during its formative stages. For the 15-20% of patients who undergo potentially curative resection, the 5-year survival is only 20% [3]. Some improvements in surgical outcome occur in patients who also receive chemotherapy and/or radiotherapy, although the impact on long-term survival has been minimal owing to the intense resistance of pancreatic adenocarcinoma to all extant treatments. It is thus evident the need to conduct a more penetrating analysis of pancreatic cancer biology. The identification of signature gene mutations in pancreatic adenocarcinoma was recognized as a valuable starting point, providing a conceptual framework to guide the future analysis of complex aspects of this disease. How these genetic changes translate into the classical biological features of pancreatic cancer cells stands as a key area for increased active investigation. Pancreatic adenocarcinoma is generally thought to arise from pancreatic ductal cells; however, this remains an area of ongoing study [4]. The etiology of pancreatic adenocarcinoma remains poorly defined, although important clues of disease pathogenesis have emerged from epidemiological and genetic studies. Pancreatic adenocarcinoma is a disease that is associated with advancing age [5], rare before the age of 40, it culminates in a 40-fold increased risk by the age of 80. Environmental factors might modulate pancreatic adenocarcinoma risk [5]. On the genetic level, numerous studies have documented an increased risk in relatives of pancreatic adenocarcinoma patients (approximately three fold), and it is estimated that 10% of pancreatic cancers are due to an inherited predisposition [6]. As with most cancer types, important insights have emerged from the study of rare kindreds that show an increased incidence of pancreatic adenocarcinoma. However, unlike familial cancer

syndromes for breast, colon and melanoma, pancreatic adenocarcinoma that is linked to a familial setting has a lower penetrance (<10%) and maintains a comparable age of onset to sporadic cases in the general population.

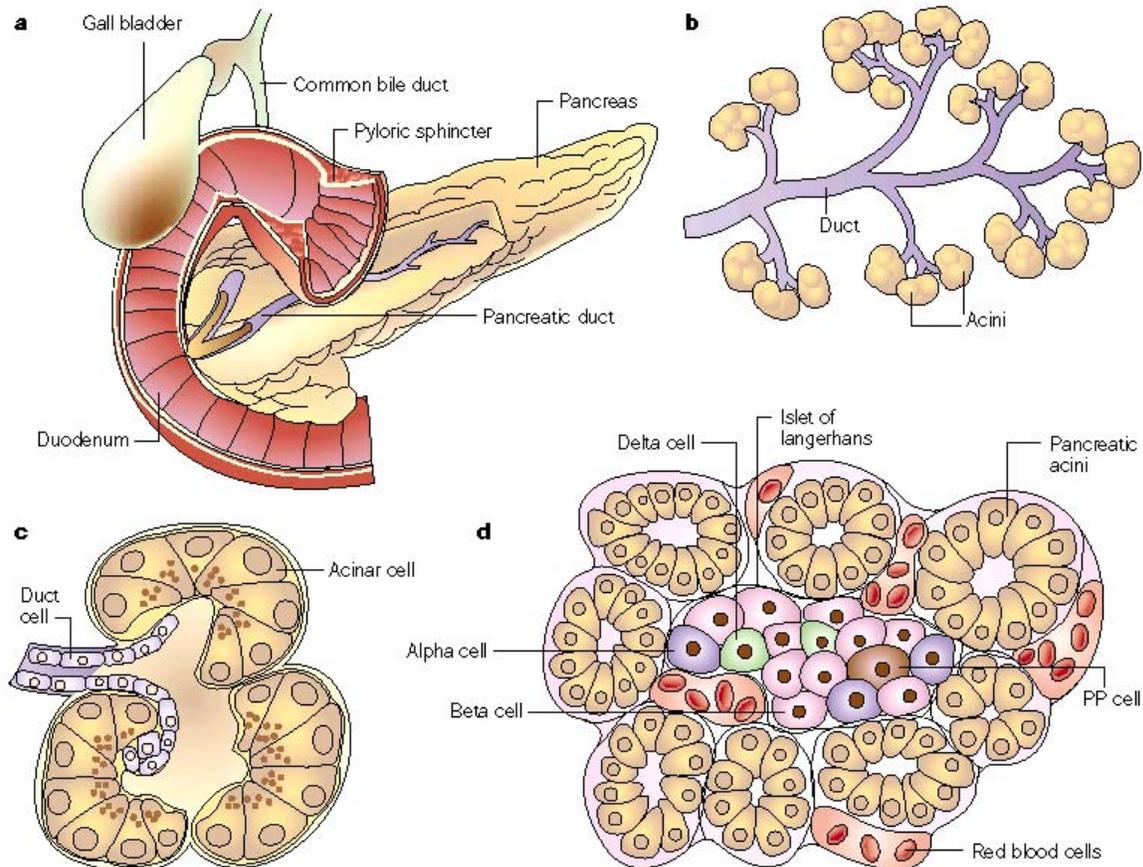


Figure 1. Anatomy of the pancreas. The pancreas is comprised of separate functional units that regulate two major physiological processes: digestion and glucose metabolism. The exocrine pancreas consists of acinar and duct cells. The acinar cells produce digestive enzymes and constitute the bulk of the pancreatic tissue. They are organized into grape-like clusters that are at the smallest termini of the branching duct system. The ducts, which add mucous and bicarbonate to the enzyme mixture, form a network of increasing size, culminating in main and accessory pancreatic ducts that empty into the duodenum. The endocrine pancreas, consisting of four specialized cell types that are organized into compact islets embedded within acinar tissue, secretes hormones into the bloodstream. The α - and β -cells regulate the usage of glucose through the production of glucagon and insulin, respectively. Pancreatic polypeptides and somatostatin that are produced in the PP and δ -cells modulate the secretory properties of the other pancreatic cell types. **a)** Gross anatomy of the pancreas. **b)** The exocrine pancreas. **c)** A single acinus. **d)** A pancreatic islet embedded in exocrine tissue.

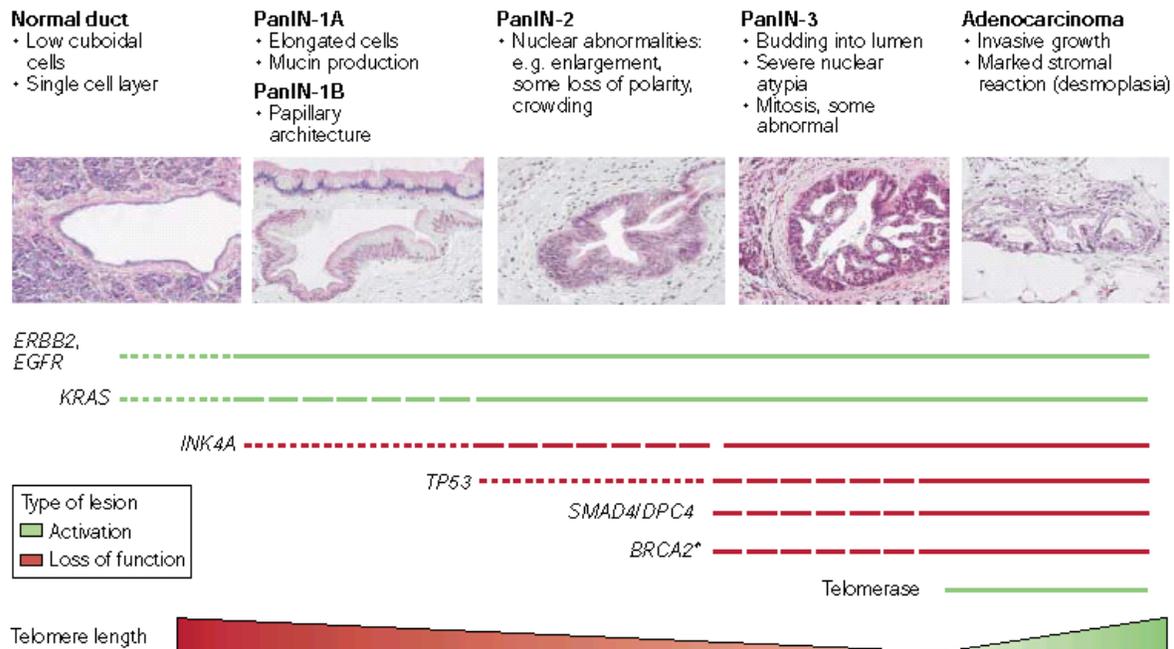


Figure 2. Genetic progression model of pancreatic adenocarcinoma. Pancreatic intraepithelial neoplasias (PanINs) seem to represent progressive stages of neoplastic growth that are precursors to pancreatic adenocarcinomas. The genetic alterations documented in adenocarcinomas also occur in PanIN in what seems to be a temporal sequence, although these alterations have not been correlated with the acquisition of specific histopathological features. The stage of onset of these lesions is depicted. The thickness of the line corresponds to the frequency of a lesion.

A careful molecular and pathological analysis of evolving pancreatic adenocarcinoma has revealed a characteristic pattern of genetic lesions. The pancreatic-duct cell is generally believed to be the progenitor of pancreatic adenocarcinoma. The increased incidence of abnormal ductal structures (now designated pancreatic intraepithelial neoplasia, PanIN) [5, 7] in patients with pancreatic adenocarcinoma, and the similar spatial distribution of such lesions to malignant tumours, are consistent with the hypothesis that such lesions might represent incipient pancreatic adenocarcinoma. Histologically, PanINs show a spectrum of divergent morphological alterations relative to normal ducts that seem to represent graded stages of increasingly dysplastic growth [4] (figure 2). Cell proliferation rates increase with advancing PanIN stages, which is consistent with the idea that these are progressive lesions [7]. A growing number of studies have identified common mutational profiles in simultaneous lesions, providing supportive evidence of the relationship

between PanINs and the pathogenesis of pancreatic adenocarcinoma. Specifically, common mutation patterns in PanIN and associated adenocarcinomas have been reported for *KRAS* and for *CDKN2A* [8]. In addition, similar patterns of loss of heterozygosity (LOH) at chromosomes 9q,17p and 18q (harbouring *CDKN2A*, *TP53* and *SMAD4*, respectively) have been detected in coincident lesions, and studies have consistently shown an increasing number of gene alterations in higher-grade PanINs [9-12].

5.2 EPIGENETIC REGULATION OF GENES

Knowledge of the molecular events that occur during the early stages of cancer has advanced rapidly. The initiation and development of cancer involves several molecular changes, which include epigenetic alterations. Epigenetics is the study of modifications in gene expression that do not involve changes in DNA nucleotide sequences. Modifications in gene expression through methylation of DNA and remodelling of chromatin via histone proteins are believed to be the most important of the epigenetic changes. The study of epigenetics offers great potential for the identification of biomarkers that can be used to detect and diagnose cancer in its earliest stages and to accurately assess individual risk.

Epigenetics has an important role in biological research and affects many different areas of study including cancer biology [13, 14], viral latency [15-18], activity of mobile elements [19], somatic gene therapy [20-25], cloning and transgenic technologies, genomic imprinting [26, 27], and developmental abnormalities [26, 27]. The definition of epigenetics varies among investigators. One definition describes epigenetics as the study of mitotically heritable changes in gene expression that are not caused by alteration of the DNA sequence [28]. Another view is that epigenetics concerns the inheritance of information on the basis of differential gene expression, whereas genetics focuses on the information inherited through gene sequence [29]. The mechanism of inheritance for epigenetic modification has yet to be identified. However, regardless of the definition used, the most important difference between an epigenetic mechanism and a genetic mechanism is that epigenetic changes can be reversed by chemical agents (therapeutic interventions).

Various factors can modify mammalian cells resulting in an epigenetically transformed phenotype, without changing the DNA sequence information of the cell, including radiation, tobacco smoke,

stress, hormones (such as oestradiol), base analogues, cadmium, arsenic, nickel, reactive oxygen species, and various other chemicals.

Two major steps in the epigenetic regulation of gene expression are the deacetylation of histones, causing a change in the structure of chromatin, and the methylation of the promoter region of the gene.

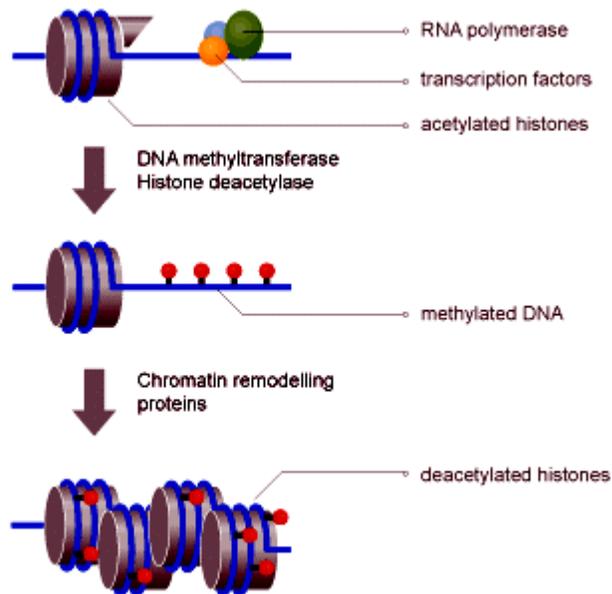


Figure 3. Epigenetic modifications that abolish gene expression. Top panel – open chromatin is characterized by non-methylated DNA and histones with acetylated tails. This allows the assembly of transcription factors and transcription by RNA polymerase. Middle panel – DNA methyltransferase activity results in the methylation of DNA. This may directly block binding by transcription factors and prevent transcription. It may also recruit methyl-binding domain proteins that have associated histone deacetylases. Bottom panel – DNA methylation and histone deacetylation result in the condensation of chromatin into a compact state that is inaccessible by transcription factors.

5.2.1 HISTONE DEACETYLATION

The DNA of all eukaryotes is packaged into chromatin, which is made up of histone proteins around which the DNA is coiled. Histones have amino-terminal extensions called ‘tails’, which undergo many covalent modifications that are important in both the organisation of chromosomes and the regulation of specific genes. Histone methyltransferases direct site-specific methylation of aminoacid residues such as Lys4 and Lys9 in the tail of the histone protein H3. Methylation of Lys4

is important for the maintenance of the structure of euchromatic domains, which are diffuse areas of chromatin where genes are freely accessible and generally active. By contrast, methylation of H3 Lys9 is associated with the initiation and propagation of heterochromatic domains where chromatin is densely packed and the genes are generally inactive [30]. The dense, heterochromatic domains are flanked by inverted repeats that act as boundary elements which prevent the tightly packed structure spreading to neighbouring euchromatic regions.

There is evidence that histone proteins and their associated covalent modifications contribute to a mechanism that can alter the structure of chromatin. Such changes could lead to inherited differences in transcriptional activation (or inactivation) states or to the stable propagation of chromosomes by a specialised centromere higher-order structure. Histone acetylation is a dynamic process that is regulated by two classes of enzymes: the histone acetyltransferases and histone deacetylases. Although many recent studies have focused on promoter-specific acetylation and deacetylation, these mechanisms are part of a broader, more dynamic acetylation mechanism that profoundly affects many nuclear processes. During mitosis, histone acetyltransferases and deacetylases are unable to acetylate or deacetylate chromatin *in situ* despite remaining catalytically active when isolated from mitotic cells and assayed *in vitro*. Thus, these enzymes do not stably bind to the genome to function as an epigenetic mechanism of selective postmitotic gene activation. However, evidence does support a role for spatial organisation of these enzymes within the nucleus. Furthermore, their relation to euchromatin and heterochromatin postmitotically in the reactivation of the genome also is important for the active organised structure.

5.2.2 DNA METHYLATION

In the normal mammalian genome, methylation occurs only at cytosines 5' to guanosines at CpG dinucleotides. Many CpG dinucleotides have been depleted from the eukaryotic genome throughout evolution by spontaneous deamination. The remaining CpGs have a very high frequency of methylation, which facilitates changes in chromatin structure that block the transcription of particular genes, by rendering them inaccessible to cellular transcription machinery. However, throughout the genome, short stretches of CpG-rich DNA exist. These regions, known as CpG islands, are not highly methylated despite being rich in cytosine and guanine, and are generally located in the promoter regions of “housekeeping” genes which are essential for cell function. The lack of methylation in promoter regions may be a prerequisite for active transcription of the genes under their control.

Many tumour suppressor genes are inactivated, in particular tumour types have highly methylated promoters; under normal conditions, the promoter regions are unmethylated and the gene is transcribed [31-37]. Whether methylation is the initiating or secondary event in gene silencing has not been established. However, irrespective of its role in the initiation of cancer development, methylation is an important marker for epigenetically mediated loss-of-gene function. Furthermore, these events are of comparable importance to gene mutations for the initiation and propagation of carcinogenesis. Although the functional importance of hyper-methylation is apparent, the molecular mechanisms involved are still unclear. The integration of DNA methylation with chromatin organisation and the regulation of histone acetylation and deacetylation may be important parts of the overall effect.

Furthermore several studies have shown low DNA methylation of proto-oncogenes in cancer cells [38-40]. For example, low DNA methylation of *Raf*, *c-Myc*, *c-Fos*, *c-H-Ras*, and *c-K-Ras* associated with neoplasia have been reported in rodent liver [40, 41]. Although some studies have identified hypomethylation of *RAS* in human cancers [42-44], other studies do not support these findings and some researchers have suggested that DNA methylation is irrelevant to *RAS* expression [38, 45]. A significant inverse correlation was found between methylation and the degree of expression of the *BCL-2* gene in human B-cell chronic lymphocytic leukaemia [39]. Hypomethylation of the third exon of the *c-MYC* gene has been reported in a various human cancers [46].

Another emerging concept is that deacetylation of histone proteins is the first step in the recruitment of methyltransferase to the CpG islands, resulting in hypermethylation of the promoter.

For cancer prevention strategies to be developed, factors that regulate deacetylation have to be identified. Extensive research is needed in this area, especially because targets for chemoprevention could emerge from these studies. Furthermore, epigenetic regulations occur early in cancer progression, thus providing an opportunity for the development of interventions to prevent further progression.

5.3 HISTONE DEACETYLATION (I): PROTEOMIC PROFILING OF PANCREATIC DUCTAL CARCINOMA CELL LINES TREATED WITH TRICHOSTATIN-A

5.3.1 INTRODUCTION

The epigenetic events may be modulated by pharmacological intervention and several drugs have entered into clinical trials during the last decade [47, 48]. Those compounds that interfere with the acetylation/deacetylation pattern of histones seem to show particular promise.

Chromatin modifying factors may be involved in cell proliferation and cancer [49, 50]. Generally, transcriptional efficiency is correlated with the relative activities of histone acetyltransferases (HAT) and histone deacetylases (HDAC). Acetylation is believed to separate the basic N-termini of histones from DNA that then becomes more accessible to transcription factors [51, 52]. Thus, histone acetylation leads to gene activation, while histone deacetylation leads to a tighter histone-DNA interaction and, accordingly, to gene repression. The role of histone acetylation and deacetylation in the genesis of cancer has been shown in recent studies [53]. During the last decade, a number of HDAC inhibitors have been identified that induce cultured tumor cells to undergo growth arrest, differentiation and/or apoptotic cell death [54]. These agents also inhibit the growth of cancer cells in animal models and several agents have been shown to slow tumor growth in animals at nontoxic doses. It has also been suggested that the action of HDAC inhibitors on gene expression is selective and restricted to a relatively small number of genes based on the results using differential display analysis of tumor cells cultured with trichostatin-A (TSA). In that study, only the expression of a small number of genes (8 of 340) was altered compared with untreated cells. Among these, two key genes involved in cell proliferation, namely *c-myc* and *p21*, were shown to have decreased or increased expression, respectively.

The present data demonstrate that the growth of the human pancreatic adenocarcinoma cell line Paca44 is strongly inhibited by TSA at submicromolar concentrations, and that the cellular mechanisms underlying this effect consists in cell cycle arrest at the G2 phase and apoptotic cell death. In order to understand these effects, the differential protein expression profiles of the pancreatic adenocarcinoma cell line Paca44, following treatment with TSA (a potent inhibitor of histone deacetylase), was investigated by using proteomics tools.

5.3.2 EXPERIMENTAL PROCEDURES

Cell culture and cell proliferation assay were performed by Department of Neurological and Visual Sciences staff, Section of Biochemistry, University of Verona.

5.3.2.1 Cells and growth conditions

Paca44 human adenocarcinoma cells were grown in RPMI 1640 supplemented with 20 mM glutamine and 10% FBS (BioWhittaker, Italy) and were incubated at 37°C with 5% CO₂. Cells were seeded at a density of 25x10³/cm².

5.3.2.2 Cell proliferation assay

Cells were plated in 96-well cell culture plates at subconfluency (4x10³ cells/well) and after 24 h treated with 0.2 μM TSA for 40 h. Cells were stained with Crystal Violet (Sigma), solubilized in PBS with 1% SDS, and measured photometrically (A_{595nm}) to determine cell viability. Three independent experiments were performed.

5.3.2.3 Cell cycle analysis

Cell cycle distribution was analyzed by using propidium iodide (PI)-stained cells. Briefly, 10⁶ cells were treated with 0.2 μM TSA for 48 h, washed with PBS, incubated with 0.1% sodium citrate dihydrate, 0.1% Triton X-100, 200 μg/mL RNase A, 50 μg/mL propidium iodide (Roche Molecular Biochemicals) and analyzed on a FACScalibur flow cytometer (Becton Dickinson). The percentage of cells in the various stages of the cell cycle was determined using the ModFitLT software program.

5.3.2.4 Apoptosis detection

Apoptosis was assessed by detection of histone-associated DNA fragments of cells treated for 40 h with 0.2 μM TSA. Cells were analyzed using the Cell-Death Detection ELISA^{Plus} assay (Roche Molecular Biochemicals). The assay is based on a quantitative sandwich-enzyme-immunoassay principle using monoclonal antibodies anti-histone-biotin and anti-DNA-peroxidase (POD). POD was determined photometrically (A_{405nm} - A_{492nm}) with ABTS as substrate. This allows the apoptotic determination of the enrichment of nucleosomes in the cytoplasmic fraction of cell lysates.

5.3.2.5 Two-dimensional gel electrophoresis

Protein extraction from cells untreated and treated with 0.2 μ M TSA for 18 h was performed with lysis buffer (40 mM Tris, 1% NP40, 1 mM Na_3VO_4 , 1 mM NaF, 1 mM PMSF, protease inhibitor cocktail). Cells were left in lysis buffer for 30 min in ice. After centrifugation at 14,000 x g at 4°C (for removal of particulate material) the protein solution was collected and stored at -80°C until used. Seventeen cm long, pH 3-10 immobilized pH gradient strips (IPG; Bio-Rad Labs., Hercules, CA, USA) were rehydrated for 8 h with 450 μ L of 2-D solubilizing solution (7 M urea, 2 M thiourea, 5 mM tributylphosphine, 40 mM Tris and 20 mM iodoacetamide) containing 2 mg/mL of total protein from Paca 44 cells. Isoelectric focusing (IEF) was carried out with a Protean IEF Cell (Biorad, Hercules, CA, USA), with a low initial voltage and then by applying a voltage gradient up to 10000 V with a limiting current of 50 mA/strip. The total product time x voltage applied was 70000 Vh for each strip and the temperature was set at 20°C. For the second dimension, the IPGs strips were equilibrated for 26 min by rocking in a solution of 6 M urea, 2% SDS, 20% glycerol, 375 mM Tris-HCl, pH 8.8. The IPG strips were then laid on an 8-18%T gradient SDS-PAGE with 0.5% agarose in the cathode buffer (192 mM glycine, 0.1% SDS and Tris to pH 8.3). The anodic buffer was a solution of 375 mM Tris HCl, pH 8.8. The electrophoretic run was performed by setting a current of 2 mA for each gel for 2 hours, then 5 mA/gel for 1 h, 10 mA/gel for 20 h and 20 mA/gel until the end of the run. During the whole run the temperature was set at 11°C. Gels were stained overnight with Colloidal Coomassie blue (0.1% Coomassie Brilliant Blue G, 34% v/v methanol, 3% v/v phosphoric acid and 17% w/v ammonium sulphate), while destaining was performed with a solution of 5% acetic acid until a clear background was achieved.

5.3.2.6 Protein pattern analysis

The 2-DE gels were scanned with a GS-710 densitometer (Bio-Rad), and analyzed with the software PDQuest Version 6.2 (Bio-Rad, Laboratories, Hercules, CA, USA). A match set was created from the protein patterns of the two independent cellular extracts (control cell line, TSA-treated cell line). A standard gel was generated out of the image with the highest spot number. Spot quantities of all gels were normalized to remove non expression-related variations in spot intensity, so 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 allowed the study of proteins that were significantly increased or

decreased in TSA-treated cell line. 51 spots were found to be differently expressed in TSA-treated Paca 44, of which 29 spots were up-regulated (with a significance level α of 0.05) and 22 spots down-regulated (with $\alpha = 0.05$).

5.3.2.7 Protein identification by mass spectrometry

The spots of interest were carefully excised from the gel with a razor blade, placed in Eppendorf tubes, and destained by washing three times for 20 min in 50% v/v acetonitrile, 2.5 mM Tris, pH 8.5. The gel pieces were dehydrated at room temperature and covered with 10 μ L of trypsin (0.04 mg/mL) in Tris buffer (2.5 mM, pH 8.5) and left at 37°C overnight. The spots were crushed and peptides were extracted in 15 μ L of 50% acetonitrile, 1% v/v formic acid. The extraction was conducted in an ultrasonic bath for 15 min. The sample was centrifuged at 8000 x g for 2 min, and the supernatant was collected.

The extracted peptides were loaded onto the 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 reflectron mode, with an accelerating voltage of 20 kV. Peptide masses were searched against SWISS-PROT, TrEMBL and NCBI nr databases by utilizing the ProteiLynx program (Micromass) (<http://www.expasy.ch/tools/peptide.html>) or NCBI nr database by using the ProFound (<http://129.85.19.192>) program (<http://prospector.ucsf.edu/ucsfhtml3.4/msfit.htm>).

5.3.3 RESULTS

Fig. 4 shows that TSA is able to inhibit proliferation of Paca44 cells (Fig. 4a) via mechanisms involving apoptotic cell death (Fig. 4b) and cell cycle arrest at the G2 phase (Fig. 4c).

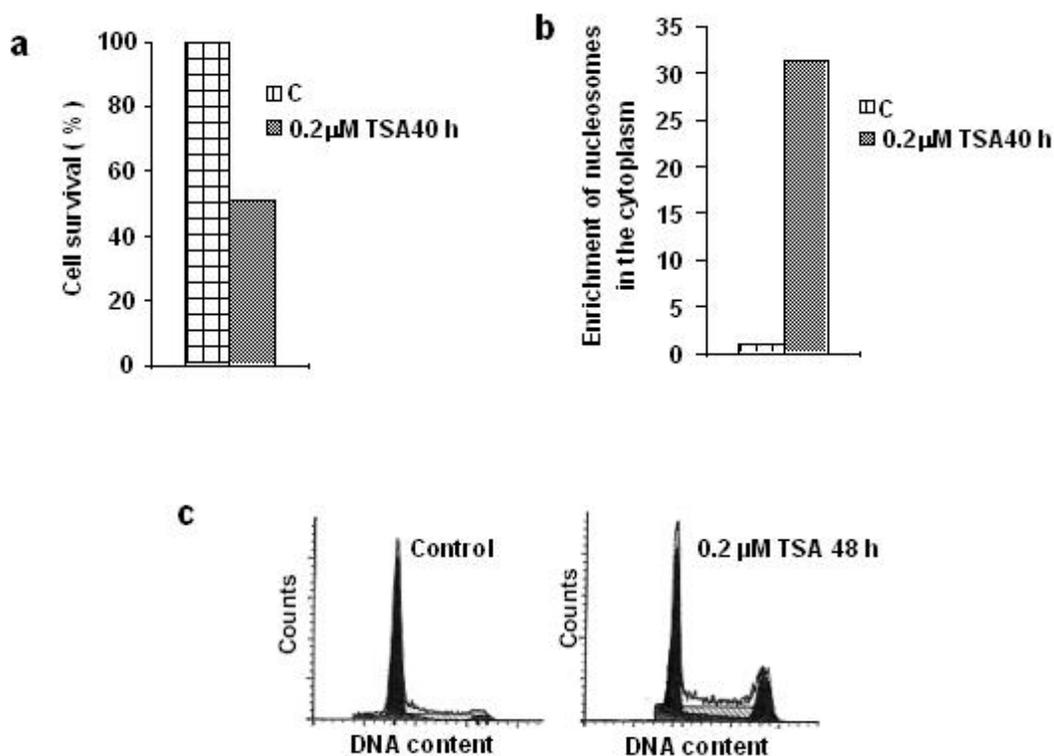


Figure 4. Effect of 0.2 μM TSA on growth of Paca 44 adenocarcinoma cell line. The cells were seeded in 96-well plates (a and b) or in 60 mm diameter plates (c) and incubated overnight. Cells were further incubated in the absence or presence of TSA for the time indicated in each panel. 1a: proliferation. Cell proliferation was determined using the Crystal Violet colorimetric assay as described in materials and methods. Values are the means of triplicate wells. 1b: apoptosis. The cells were analyzed by detection of histone-associated DNA fragments in the cytoplasmic fraction of cell lysates as described in materials and methods. The values are the means of triplicate wells. 1c: cell cycle. The cell cycle distribution was analyzed by a flow cytometer after DNA staining with propidium iodide. Similar results were obtained from three independent experiments.

Fig. 5 shows a standard 2-D map of the Paca44 cell line, stained with colloidal Coomassie Blue. About 700 polypeptide spots could be revealed in the pH 3-10 interval with this medium-sensitivity stain.

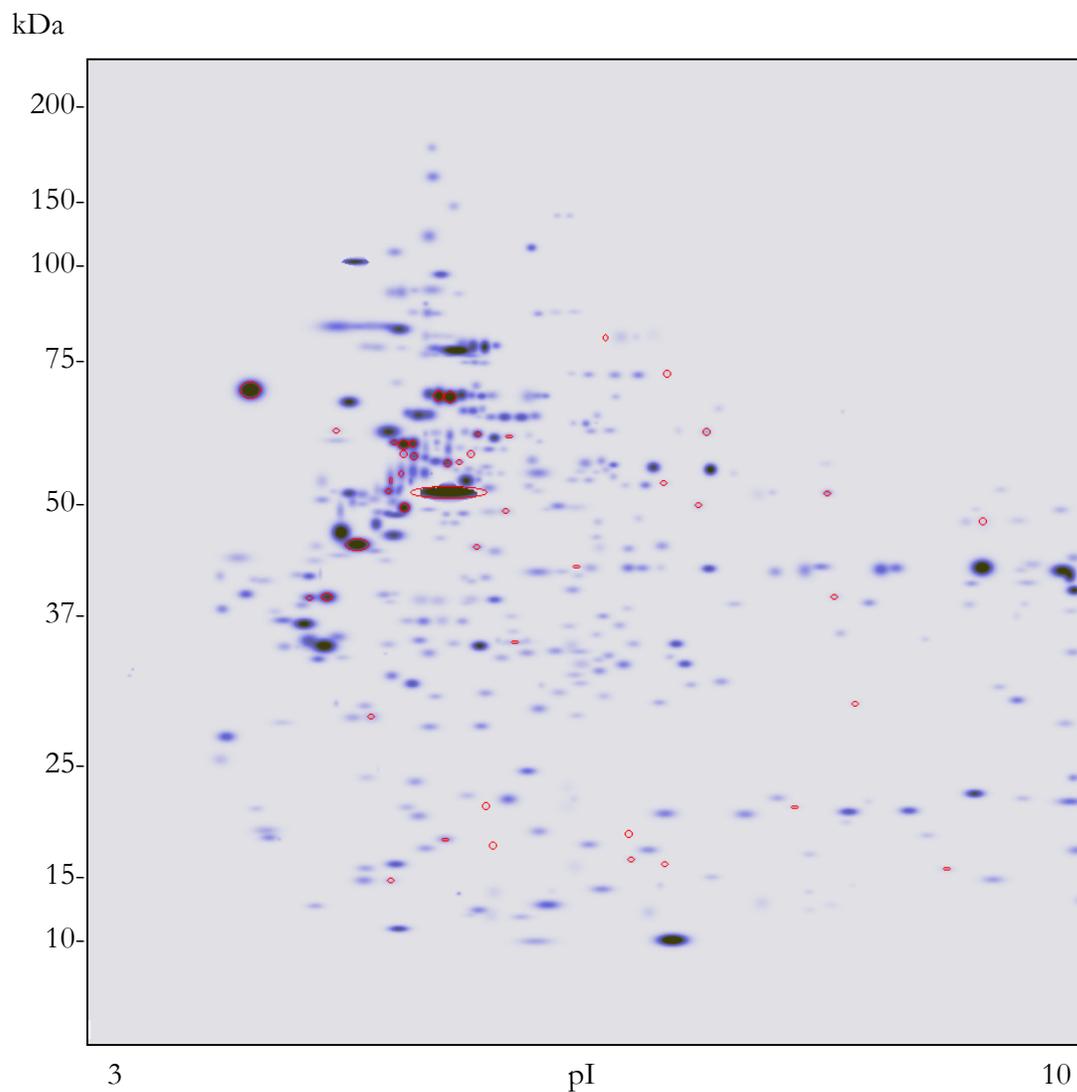


Figure 5. Master map of the pancreatic adenocarcinoma cell line Paca44. The 51 spots differently expressed after trichostatin A treatment are marked in red.

After matching the master map of the control Paca44 cells with the master map of the TSA-treated cells, 51 polypeptide chains, highlighted in red, were found to be differentially expressed after treatment with the anti-tumoral drug. All these spots were eluted and treated as described in section 2.6. In Fig. 6, we have marked the 22 spots which could be identified by MS analysis, in red those down-regulated, in blue those up-regulated.

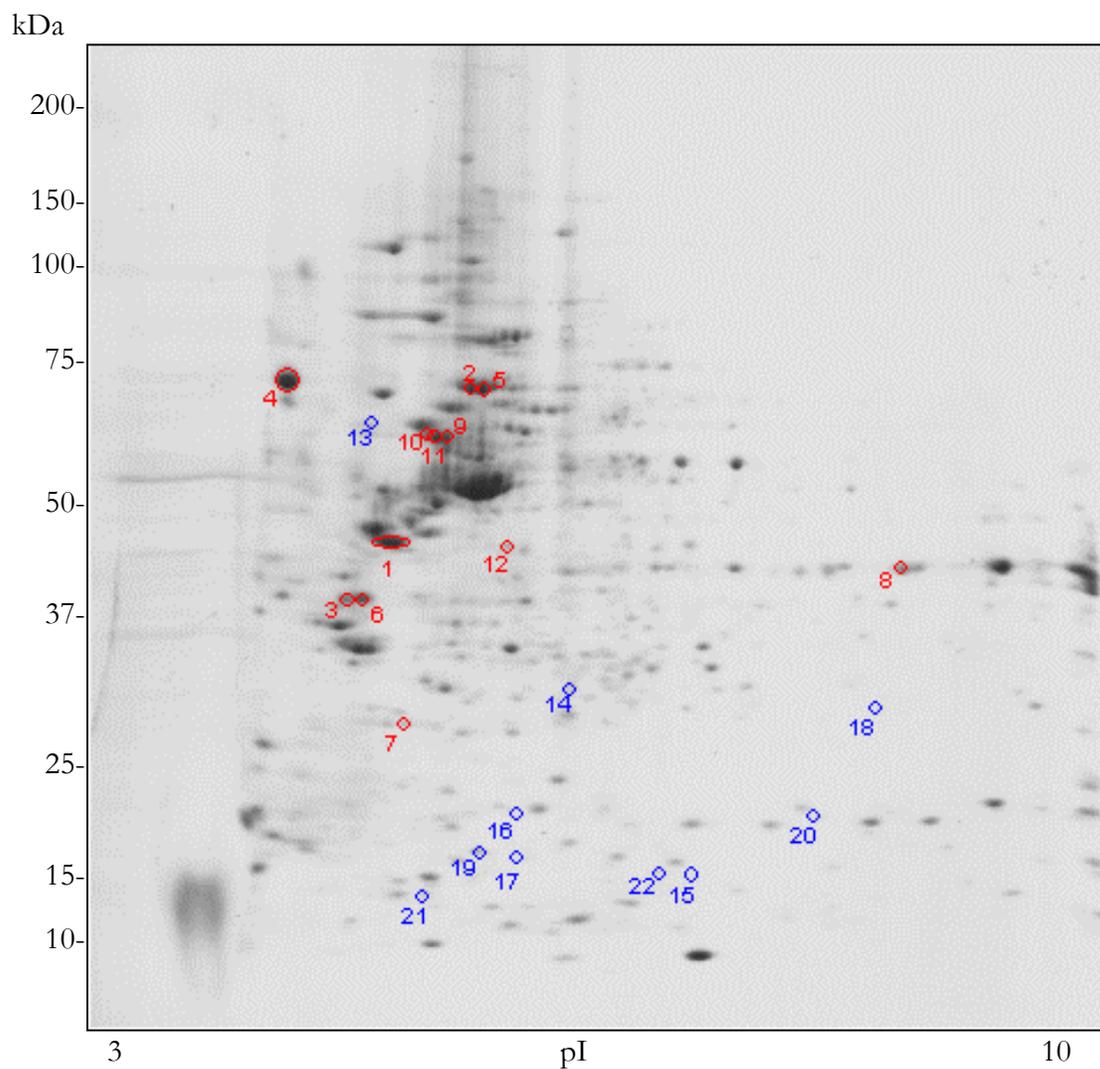


Figure 6. Two-dimensional master map of Paca44 cells in the pH 3-10 IPG interval. The identified spots are numbered, in red the proteins down-regulated, in blue the proteins up-regulated. The numbered spots have been identified by peptide fingerprinting and MS analysis (see Table 1 for protein classification).

Their identification, together with experimental and theoretically predicted pI and Mr values, are given in Table 1.

Spot number	2D Gel		Databank								
	Exp. Mr (Da)	Exp. pI	Theor. Mr.	Theor. pI	Z-Score	MOWSE-Score	protein name	Accession Number	%coverage	No. of peptides	VARIATION
1	~ 40000	~ 5,2	32554	4,7	2,03	4.79 E12	Nucleophosmin (NPM) (Nucleolar phosphoprotein B23) (Numatrin)	P06748	55,1	18	decreased 3
2	~ 72000	~ 5,7	57926	5,3	2,4	6.77 E31	60 kDa heat shock protein, mitochondrial precursor (Hsp60) (60 kDa chaperonin)	P10809	69,4	42	decreased 2
3	~ 35000	~ 4,5	24504	4,7	2,37	8.28 E12	Tropomyosin alpha 4 chain (Tropomyosin 4)	P07226	73,4	25	decreased 2
4	~ 75000	~ 4,2	46466	4,3	2,38	1.81 E19	Calreticulin precursor (CRP55) (Calregulin)	P27797	62	25	decreased 2
5	~ 75000	~ 6	57926	5,3	2,38	1.46 E27	60 kDa heat shock protein, mitochondrial precursor (Hsp60) (60 kDa chaperonin)	P10809	60	36	decreased 2
6	~ 35000	~ 4,9	32798	4,7	2,36	4.83 E14	Tropomyosin alpha 3 chain (Tropomyosin 3)	P12324	65	31	decreased 3
7	~ 26000	~ 5,4	19582	4,9	2,41	1.09 E9	Translationally controlled tumor protein (TCTP)	P13693	47	14	decreased 3
8	~ 37000	~ 9	37406 & 35899	9,0 & 8,6	2,4	5.69 E10 & 1.17 E10	Heterogeneous nuclear ribonucleoproteins A2/B1 & Glyceraldehyde 3-phosphate dehydrogenase, liver (EC 1.2.1.12).	P22626 & P04406	54 & 48	16 & 16	decreased 2
9	~ 55000	~ 5,6	51736 & 46142	5,0 & 5,0	2,35	3.38 E12 & 4.19 E9	ATP synthase beta chain, mitochondrial precursor (EC 3.6.3.14) & Protein disulfide isomerase A6 precursor (EC 5.3.4.1)	P06576 & Q15084	41 & 41	16 & 13	decreased 3
10	~ 55000	~ 5,4	51736 & 49671	5,0 & 4,8	2,32	4.24 E11 & 1.20 E10	ATP synthase beta chain, mitochondrial precursor (EC 3.6.3.14) & Tubulin beta-1 chain	P06576 & P07437	38 & 42	15 & 14	decreased 2
11	~ 55000	~ 5,5	51736	5	2,41	1.63 E15	ATP synthase beta chain, mitochondrial precursor (EC 3.6.3.14)	P06576	53	24	decreased 2
12	~ 37000	~ 5,5	34274	5,9	2,23	2.57 E9	60S acidic Ribosomal Protein PO	P05388	47	16	decreased 2
13	~ 60000	~ 4,6	47626	4,7	2,35	1.24 E5	Chromatin assembly factor 1 subunit C	Q09028	20	9	increased 2
14	~ 30000	~ 6,5	25190	6,2	2,43	6.67 E7	Growth factor receptor-bound protein 2	P29354	43,8	16	increased 4
15	~ 18000	~ 7	10606	6,3	1,72	1.26 E5	Cytochrome c oxidase polypeptide Vb	P10606	40	9	increased 5
16	~ 20000	~ 6,2	16310 & 17160	5,5 & 5,7	1,91 & 2,01	2.22 E9 & 5.89 E7	ARP2/3 complex 16 kDa subunit (P16-ARC) & Stathmin (Phosphoprotein p19) (pp19) (Oncoprotein 18)	O15511 & P16949	86 & 77	14 & 24	increased 8
17	~ 18000	~ 6	14276	5,7	1,77	1.54 E5	Programmed cell death protein 5 (TFAR19 protein)	O14737	60,8	8	increased 4
18	~ 26000	~ 8,5	22096	8,2	2,38	6.57 E8	Peroxiredoxin 1 (EC 1.11.1.-)	Q06830	68	15	increased 2
19	~ 18000	~ 5,8	16040	5,5	2,34		Deduced protein product shows significant homology to coactosin	NCBIInr: AAA88022.1	57	17	increased 2
20	~ 20000	~ 8	16363	8,5	1,98	1,92 E9	UEV protein (ubiquitin-conjugating E2 enzyme variant)	NCBIInr: AAH28673	75	13	increased 2
21	~ 13000	~ 5	11606	4,8	2,35	2.33 E7	Thioredoxin	P10599	69	10	increased 2
22	~ 13500	~ 6,5	13654	6,4	2,38	1,02 E6	Hint Protein	P49773	59	6	increased 3

Table 1. Identified proteins from the Paca44 cell line 2-D gel.

Fig. 7 gives an example of a gel area (central, low-Mr area of Fig. 6) in which some up-regulated proteins could be found. The spot marked with a question mark in panel A represents a polypeptide which was 10-fold up-regulated, which gave good MS values of the tryptic digest, but which could not be identified in any of the data bases presently available.

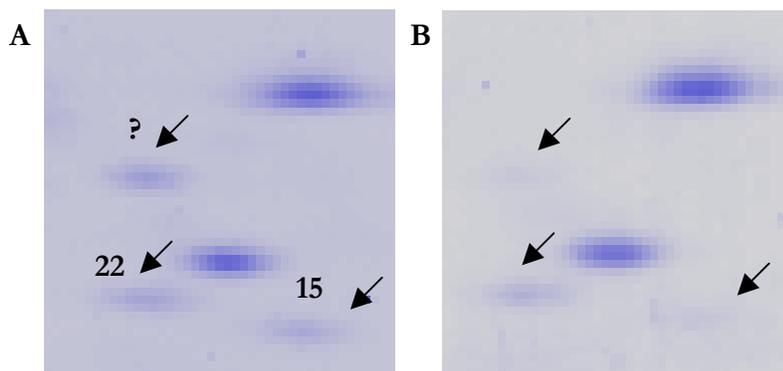


Figure 7. Comparison of 2-D gel patterns of some proteins in Paca44 cells treated with TSA (A) and in Paca44 control (B). The corresponding spot numbers are shown in Table 1; the spot with a question mark corresponds to a 10 fold up-regulated protein which could not be identified in any of the databases available, although it gave good MS spectra.

5.3.4 DISCUSSION

Differential analysis of protein expression is becoming, in the modern proteome world, a rapidly growing field for profiling a number of pathological states, such as cancer growth. The most widely accepted techniques, by and large, are still two-dimensional maps, with charge and mass coordinates [55], followed by spot excision and mass spectrometry identification of tryptic digests [56]. Although differential profiling can be today performed by a vast number of techniques (for a review, see [57]), in conventional 2-D maps the preferred method is still statistical analysis performed on master maps created from two independent cellular extracts (control *vs.* pathological or treated cells). The available software packages, although labour-intensive, produce reliable results, widely accepted by the scientific community.

To date only two studies have used a proteomic approach in the field of pancreatic diseases [92, 93]. The first one [92] addressed a technical issue and was devoted to preparation of pancreatic juice for 2-D map analysis, thus is not related to the present work. The second one [93] reported the effect of the cytotoxic agent daunorubicin on protein expression of pancreatic cancer cells. In the present

study we show that TSA is able to inhibit cell proliferation of the pancreatic adenocarcinoma cell line Paca 44 by cell growth arrest and apoptosis. We attempted to address the molecular basis of this effect by the study of protein expression profiles of cells before and after TSA treatment. This is the first paper addressing the issue of gene expression variation after HDAC inhibition, which is associated with cell cycle arrest and apoptotic cell death.

Among the 22 proteins which were identified by MS analysis, of particular interest are the two down-regulated proteins nucleophosmin and TCTP, as well as the up-regulated proteins PDCD5 and stathmin. Their role will be briefly discussed below.

Nucleophosmin (NPM), a 3-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 [58-60]. NPM appears to be present in most human tissues, with especially robust expression in pancreas and testis and lowest expression in lung [61]. Interestingly, a fusion protein, containing the amino-terminal 117 amino acid portion of NPM, joined to the entire cytoplasmic portion of the receptor tyrosine kinase ALK (anaplastic lymphoma kinase) has been found to be involved in oncogenesis in the case of non-Hodgkin's lymphoma [62].

Translationally-controlled tumour protein (TCTP), a 3-fold down-regulated polypeptide, seems to be involved in tumour reversion, *i.e.* in the process by which some cancer cells lose their malignant phenotype. In a recent study, Tuynder *et al.* [63], have shown that TCTP is strongly down-regulated in the reversion process of human leukemia and breast cancer cell lines. They thus hypothesized that tumour reversion is a biological process in its own right, involving a cellular reprogramming mechanism, overriding genetic changes in cancer, which triggers an alternative pathway leading to suppression of tumorigenicity. TCTP was also recently found, for the first time, in rat and human testes by Guillame *et al.* [64]. Interestingly, the mRNA of TCTP was recently reported to be over-expressed in human colon cancer by Chung *et al.* [65].

The programmed cell death protein 5 (PDCD5, also designated as TFAR19), here found to be up-regulated by a factor of four, is a recently discovered protein involved in the regulation of cell apoptosis [66-68]. The level of this protein in cells undergoing apoptosis, is significantly increased

compared with normal cells. Thus, its up-regulation in TSA-treated cell lines, as here reported, is consistent with our finding of apoptotic cell death of Paca 44 by TSA treatment.

Stathmin (oncoprotein 18, OC18) was up-regulated by the TSA treatment by a factor of eight. Stathmin is a p53-regulated member of a novel class of microtubule-destabilizing proteins known to promote microtubule depolymerization during interphase and late mitosis [69]. Thus, high levels of stathmin could induce growth arrest between the end of G2 phase and the beginning of mitosis boundary [70, 71]. This again is highly consistent with the observation that Paca44 showed a cell cycle arrest at the G2 phase. It is of interest to note that over-expression of stathmin, via its effect of inhibiting polymerization of microtubules, permits increased binding to these structures of vinblastin, a well known chemotherapeutic agent, during treatment of human breast cancer [69]. Due to its effect of inhibiting cell proliferation via a mitotic block, the up-regulation of stathmin here reported appears to be consistent with the anti-tumoural activity of TSA [72].

5.4 HISTONE DEACETYLATION (II): MULTIVARIATE STATISTICAL ANALYSIS OF PROTEOMICS PROFILING OF PANCREATIC DUCTAL CELL LINES TREATED WITH TRICHOSTATIN-A

5.4.1 INTRODUCTION

For what concerns the comparison of 2D-PAGE maps on the basis of the spot volume, Principal Component Analysis (PCA) has been applied since the middle eighties by Anderson *et al.* [73] in USA and by Tarroux *et al.* [74] in France. Recently, it has been applied to the study of DNA and RNA fragments of several biological systems [75-78] and to the characterisation of proteomic patterns of different classes of tissues [79-84]. Another recent application of PCA is for the characterisation of the anticancer activity of bohemine, a new omoleucine-derived synthetic cycline-dependent kinase inhibitor, by Kovarova *et al.* [85].

In the present study, PCA is applied on a dataset constituted by 18 samples, belonging to two different pancreatic ductal cell lines (Paca44 and T3M4) before and after the treatment with trichostatin-A.

This approach is focused on the identification of the differences that depend on the treatment and on the different cell lines (comparison between control/treated samples and Paca44/T3M4 cell lines) and on the validation of the results obtained by proteomics analysis of Paca44 cell lines treated with trichostatin-A (see chapter 5.3).

5.4.2 THEORY

5.4.2.1 Principal Component Analysis

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 residual variance contained in the original dataset: the first PC explains the maximum amount of variance contained in the overall dataset, while the second one explains the maximum residual variance. The PCs are then calculated hierarchically, so that experimental noise and random variations are contained in the last PCs. The PCs, which are expressed as linear combinations of the original variables, are orthogonal one 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*. The graphical representation of scores by means of PCs allows the identification of groups of samples showing a similar behaviour (samples close one to the other in the graph) or different characteristics (samples far from each other). By looking at the corresponding loading plot, it is possible to identify the variables which are responsible for the analogies or the differences detected for the samples in the score plot. From this point of view, PCA is a very powerful visualisation tool, which allows the representation of multivariate datasets by means of only few PCs, identified as the most relevant.

5.4.2.2 Cluster Analysis

Cluster analysis techniques allow to investigate the relationships between the objects or the variables of a dataset, in order to recognise the existence of groups. The most used methods belong to the class of the agglomerative hierarchical methods [86], 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 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 the linking method.

5.4.3 EXPERIMENTAL PROCEDURES

The dataset was constituted by 18 2D-maps, divided into four classes:

4 replicate 2D-maps of a Paca44 cell line pool;

5 replicate 2D-maps of a T3M4 cell line pool;

4 replicate 2D-maps of a Paca 44 cell pool treated for 48 hours with trichostatin-A;

5 replicate 2D-maps of a T3M4 cell pool treated for 48 hours with trichostatin-A.

Because were used pools of cell lines (grown under the same conditions) 4-5 replicates of 2D-maps for each sample were deemed amply sufficient for reducing variability due to experimental errors. This strategy is common practice in today's proteome analysis. Fig. 8 represents an example, for each class, of the experimental 2D-maps obtained.

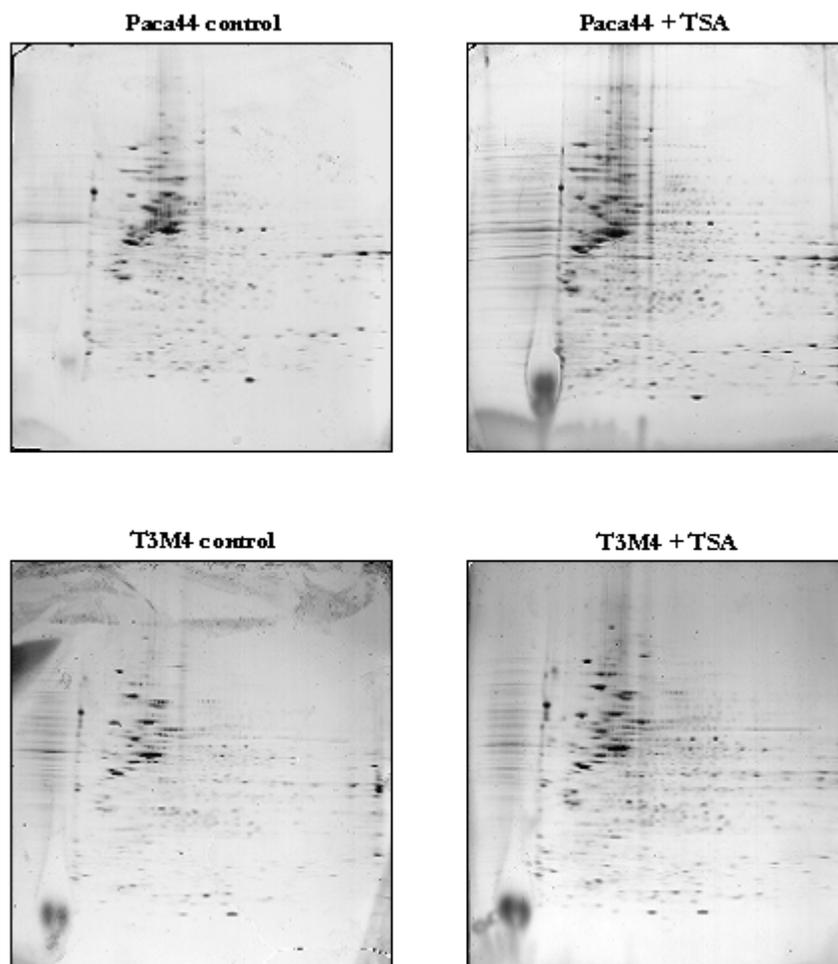


Figure 8. 2D-PAGE maps of the real samples of pancreatic human cancer: examples of control Paca44 cells, treated Paca44 cells, control T3M4 cells and treated T3M4 cells.

5.4.3.1 Softwares

PCA was performed with UNSCRAMBLER (Camo Inc., ver. 7.6, Norway). Cluster Analysis was performed with STATISTICA (Statsoft Inc., ver. 5.1, USA). Graphical representations were performed with both UNSCRAMBLER and STATISTICA. The 2D-PAGE maps were scanned with a GS-710 densitometer (Bio-Rad Laboratories, Hercules, CA, USA), and analyzed with the software PDQuest Version 6.2 (Bio-Rad Laboratories, Hercules, CA, USA).

5.4.3.2 Cell treatment with TSA

Paca44 and T3M4 cells were grown in RPMI 1640 supplemented with 20 mM glutamine and 10 % (v/v) FBS (BioWhittaker, Italy) and were incubated at 37 °C with 5 % (v/v) CO₂. Subconfluent cells were treated with 0.2 μM TSA for 48 h.

5.4.3.3 Cell lysis

Protein extraction from cells was performed with lysis buffer (40 mM Tris, 1 % v/v NP40, 1 mM Na₃VO₄, 1 mM NaF, 1 mM PMSF, protease inhibitor cocktail). Cells were left in lysis buffer for 30 min in ice. After centrifugation at 14.000 x g at 4 °C for removal of particulate material, the protein solution was collected and stored at -80 °C until used.

5.4.3.4 Two-dimensional gel electrophoresis

Two-dimensional gel electrophoresis was performed as described in Section 5.3.2.5.

5.4.3.5 Protein identification by mass spectrometry

Protein identification by MS was performed as described in Section 5.3.2.7.

5.4.4 RESULTS AND DISCUSSION

5.4.4.1 Protein pattern analysis with the PDQuest software

The 2-D gels of all the samples (Paca44 control and TSA-treated, T3M4 control and TSA-treated) were scanned with a GS-710 densitometer (Bio-Rad), and analyzed with the software PDQuest. A match-set was created from the protein patterns of the 18 replicate 2D-maps. A standard gel was generated out of the image with the highest spot number. Spot quantities of all gels were normalized to remove non expression-related variations in spot intensity, so 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 Optical Density (OD). The analysis with the PDQuest allowed two types of comparisons: between the two different cell lines (Paca44 vs. T3M4), and between the control and TSA-treated cell lines (control *vs.* TSA-treated), in order to detect protein variations that were at least two-fold. The Student's T-test analysis allowed the identification of 60 spots up-regulated (with a significance level α of 0.05) and 45 spots down-regulated (with α =

0.05) in the T3M4 cell line with respect to PaCa44 cell line; and 11 spots up-regulated (with a significance level α of 0.05) and 2 spots down-regulated in TSA-treated cell lines in respect to the control samples.

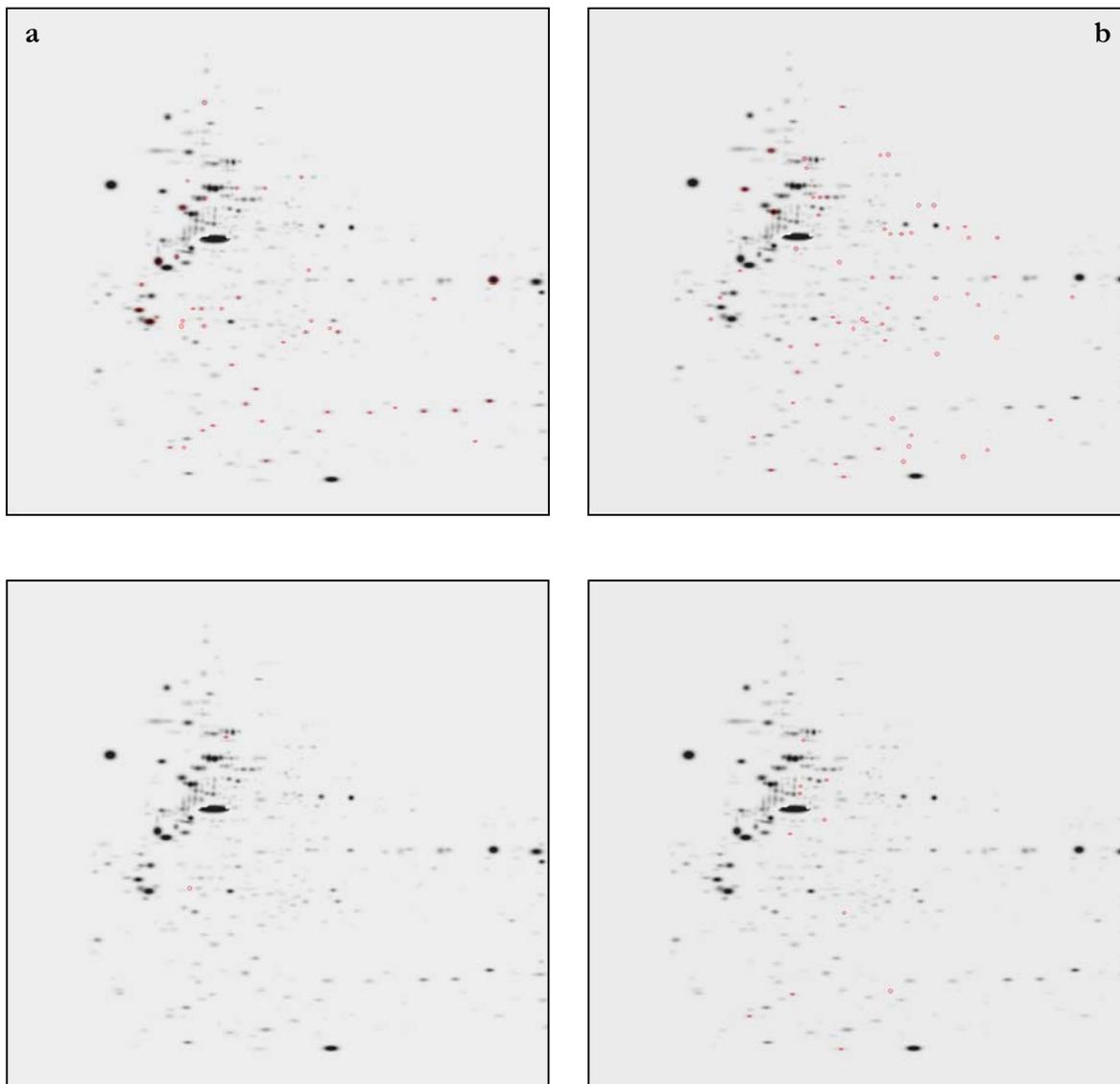


Figure 9. Results of a comparison between the two different cell lines. (a) the 45 spots with lower optical density in T3M4 (thus more intense in Paca44 cell line) are marked in red; (b) the 60 spots with higher optical intensity in T3M4 are marked in red; (c) the two spots down-regulated in TSA-treated cell lines are marked in red; (d) the 11 spots up-regulated in TSA-treated cell lines are marked in red.

Figs. 9-a and 9-b show the results of comparison between the two different cell lines (Paca44 and T3M4). In Fig. 9-a the 45 spots with a higher optical density in Paca44 (which were thus less intense in the T3M4 cell line) are marked in red, whereas in Fig. 9-b are marked in red the 60 spots with a higher optical density in T3M4.

Figs. 9-c and 9-d show the results of comparison between the control and TSA-treated cell lines. In Fig. 9-c the 2 spots down-regulated in TSA-treated cell lines (which were thus more intense in the control) are marked in red, while in Fig. 9-d are marked in red the 11 spots up-regulated in TSA-treated cell lines.

5.4.4.2 Principal Component Analysis

The differential analysis performed by PDQuest on the 18 2D-maps allowed the identification of 435 spots. The matching procedure produced a dataset constituted by 435 variables (the optical density of each matched spot) and 18 objects (the 18 samples) thus giving a data matrix of dimensions 18 x 435. All variables were autoscaled before performing PCA. The autoscaling procedure transforms the variables so that they all have a null average value and a unit variance: this last feature is fundamental since it allows all the variables to bring the same amount of information to the overall dataset. In the present case, the autoscaling procedure is particularly suitable: it gives the small spots and the large ones the same relevance, thus enhancing the detection of the differences between the four classes of 2D-maps, which are often to be searched for among the smallest spots (less abundant proteins) rather than among the largest ones (more abundant proteins). Since PCA was performed on a singular matrix, having more variables than samples, a NIPALS algorithm was used for PCs calculation. The results of PCA are given in table 2.

	Explained variance (%)	Cumulative explained variance (%)
PC1	37.59	37.59
PC2	12.90	50.49
PC3	8.27	58.75
PC4	5.42	64.17
PC5	4.70	68.87
PC6	4.33	73.20
PC7	3.47	76.67

Table 2. Results of PCA performed on the overall dataset: percentage of explained variance and percentage of cumulative explained variance.

The first three PCs explain more than 58 % of the total variance contained in the original dataset and were considered for the successive analysis.

Fig. 10 represents the score plots of the first three principal components.

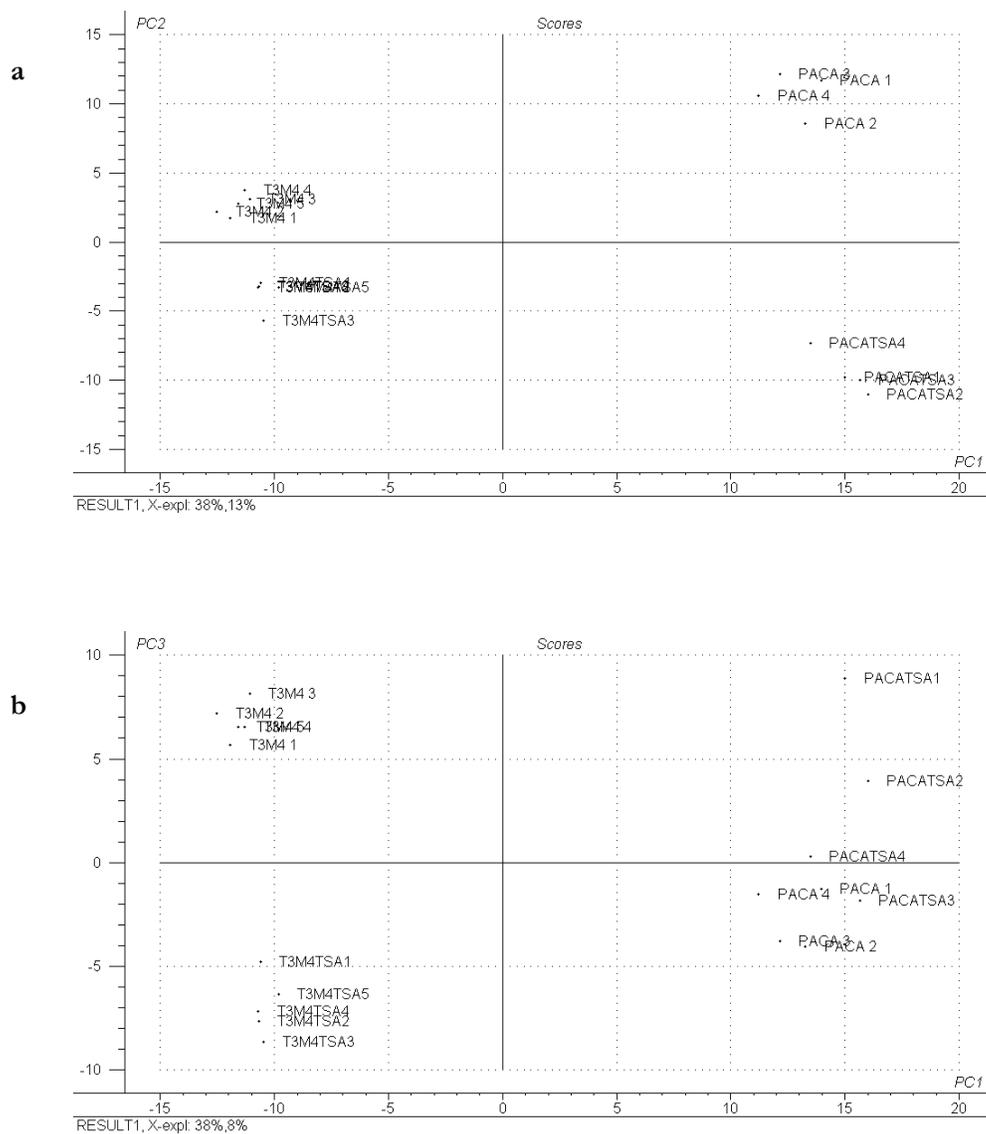


Figure 10. Score plots of the first three PCs. (a) PC2 versus PC1, and (b) PC3 versus PC1.

The first score plot (Fig. 10a) shows the sample co-ordinates on PC_2 and PC_1 ; the four classes of samples appear completely separated along the two PCs. In fact, at large positive values along PC_1 , there are the samples belonging to the cell line Paca44, while the samples belonging to the second

cell line, T3M4, are grouped at large negative scores on the same PC. The information about the cell type differences is then explained by the first PC.

The information about the differences occurring between control and TSA-treated samples is instead explained by the second PC: the TSA-treated samples of both cell lines appear in fact at large negative scores on PC₂ and the control samples at large positive scores on the same PC. However, PC₂ is dominated by the relative down-regulation of the spots in the Paca44 cell line, since the samples belonging to this cell line show the largest variations of the scores on the second PC. So PC₂ can be considered as describing the TSA general effect, with a larger contribution due to the Paca44 cell line. The first two PCs are thus able to account for the differences occurring between the four classes investigated; however also PC₃ explains a significant amount of variance, worth of being interpreted. The score plot of PC₃ versus PC₁ (Fig. 10-b) shows the control samples belonging to T3M4 cell line at large positive scores on PC₃ together with the TSA-treated ones of the Paca44 cell line; the other two groups of samples (TSA-treated samples of the T3M4 cells, and control samples of Paca44 cells) being grouped at large negative scores on the same PC. However, as for PC₂, the third PC is dominated by the relative down-regulation of the spots in the T3M4 cell line, since the samples belonging to this cell line show the largest change of the scores on PC₃. The third PC mainly describes therefore the information about the effect of TSA on T3M4 cell line, *i.e.* a complementary information with respect to that accounted for by PC₂.

Fig. 11 reports the cell survival of the two cell lines which present a similar sensitivity to a 48 h treatment with 0.2 μM Trichostatin-A. With respect to this further information, the third PC mainly accounts for the sensitivity of T3M4 cell line to the treatment with TSA.

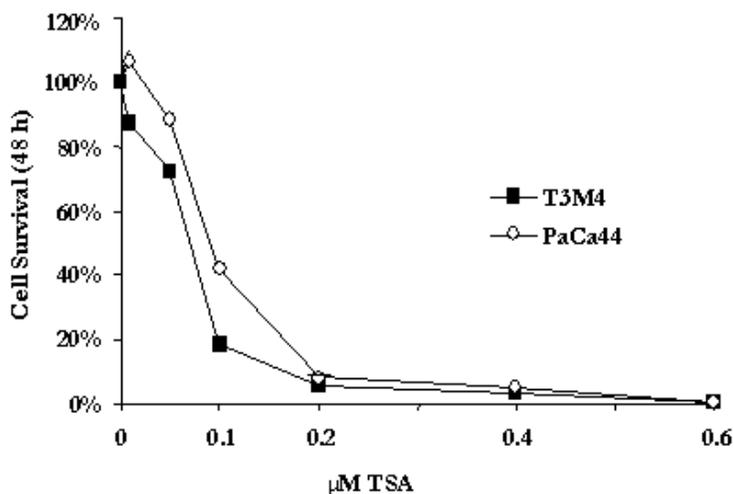


Figure 11. Cell survival after a 48-h treatment with 0.2 μM Trichostatin A for Paca44 and T3M4 cell lines.

From the previous considerations, it is possible to state that the first three PCs allow a synthetic and exhaustive representation of the investigated dataset, in facts:

- PC_1 explains the information related to the two cell lines;
- PC_2 carries the information about the TSA effect, (mainly for the Paca44 cell line);
- PC_3 carries the information about the sensitivity to TSA, (mainly for the T3M4 cell line).

The loadings of the three significant PCs provide information on the spots responsible for the regulatory effect, *i.e.* they can allow the identification of the differences occurring between the 2D-PAGE maps of the four groups of samples.

For this purpose, the loading plots of these three components are displayed in figures 12, 13 and 14, where the two central maps report the spots as circles centred in the x-y positions revealed by PDQuest analysis. The red-coloured spots have a large positive loading on the correspondent PC, while the blue ones identify the spots with large negative loadings. So the spots are represented in a colour scale where the increasing red or blue tone is proportional to its loading. The colour of the spots changes from those which show a small influence (small positive or negative loading, light red or light blue coloured) towards those which show a large influence (large positive or negative loadings, dark blue or red); the spots marked as a black circle do not have a relevant loading on the first three PCs.

Fig. 12 shows the loadings of the spots on PC_1 : the red coloured circles identify those spots showing a larger optical density in the Paca 44 cells or spots which are identified in this cell line but not in the other one; the blue coloured circles identify the spots more intense in the T3M4 cell line or those which are present in this line but missing in Paca 44 cells. The two examples of real samples on top of Fig. 12 are characterised by large optical densities of the red coloured spots and small values of the blue coloured ones (Paca 44). On the bottom of Fig. 12 there are two examples of real samples of T3M4 cell line, characterised by large optical densities of the blue coloured spots and small values of the red coloured ones.

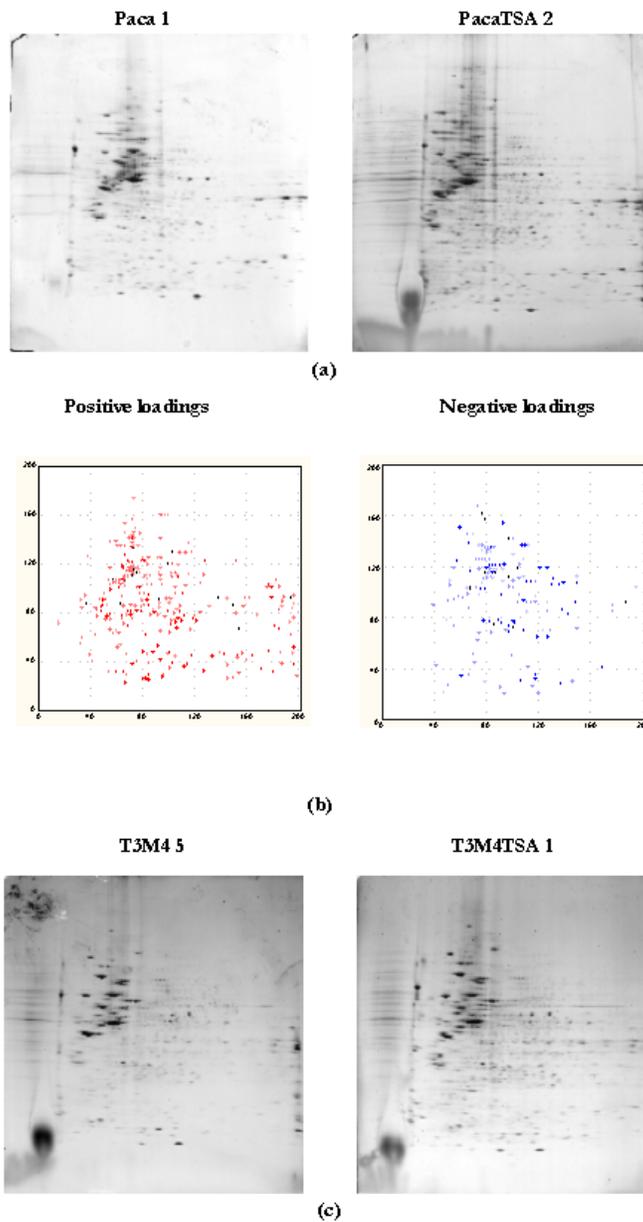


Figure 12. Loading plots of PC1 (coloured maps) with two examples of 2D-PAGE maps of real samples (Paca44 control and TSA-treated) characterised by large values of the red-coloured spots (top) and two examples of 2D-PAGE maps of real samples (T3M4 control and TSA-treated) characterised by large values of the blue-coloured spots (bottom).

Fig. 13 represents the loadings plots of the PC2: the red coloured spots identify spots characterised by a larger optical density in the diseased samples or missing in the TSA-treated ones, while the blue coloured spots represent the spots with a larger optical density in the TSA-treated samples or

missing in the control ones (of both cell lines). Figs. 13 (top and bottom maps) represents two examples of control samples of the two cell lines (characterised by large values of the red coloured spots) and two examples of TSA-treated samples of the two cell lines (characterised by large values of the blue coloured spots).

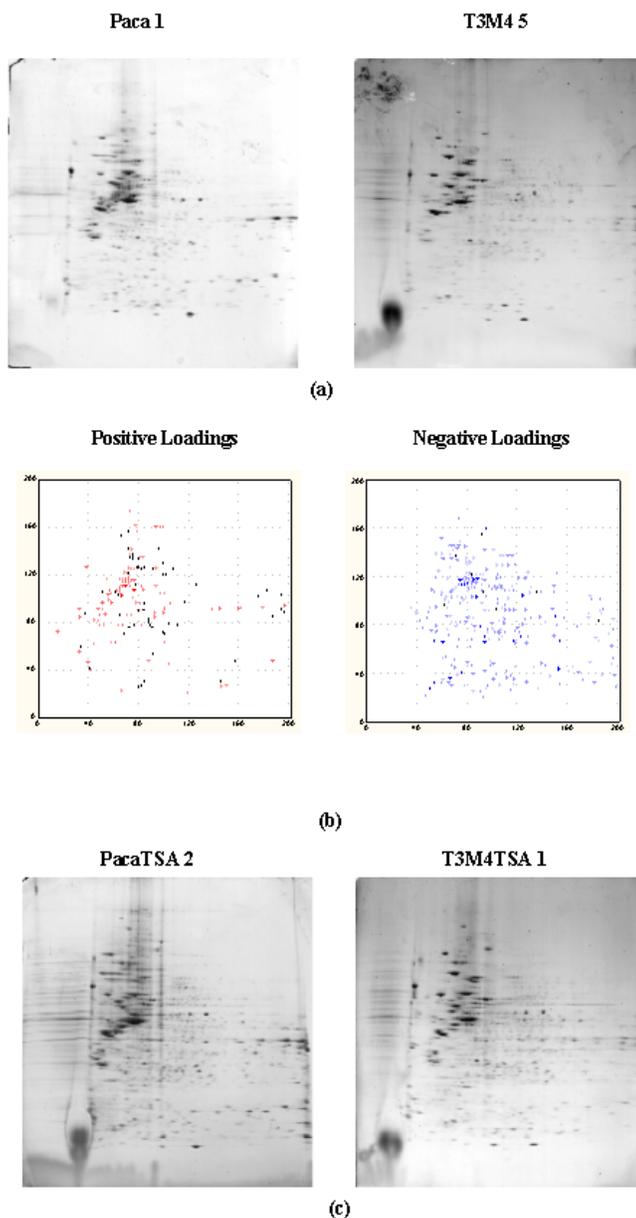


Figure 13. Loading plots of PC2 (coloured maps) with two examples of 2D-PAGE maps of Paca44 and T3M4 control samples, characterised by large values of the red-coloured spots (top) and two examples of 2D-PAGE maps of Paca44 and T3M4 TSA-treated samples, characterised by large values of the blue-coloured spots (bottom).

The loading plots of the PC3 are represented in Fig. 14: the red coloured circles identify those spots showing a larger optical density in the control T3M4 cells and in the TSA-treated Paca44 cells or spots which are absent in the other two classes of samples; the blue coloured circles show instead spots which show a larger optical density in the control Paca44 cells and the TSA-treated T3M4 cells or absent in the other two classes.

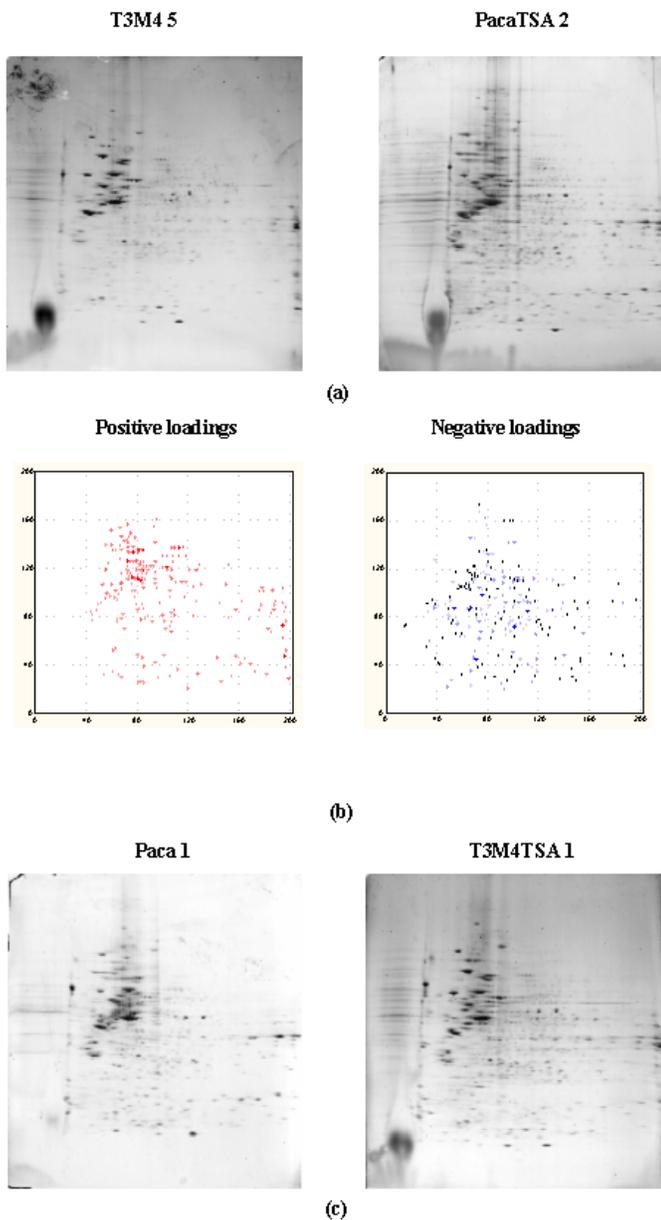


Figure 14. Loading plots of PC3 (coloured maps) with two examples of 2D-PAGE maps of Paca44 treated and T3M4 control samples, characterised by large values of the red-coloured spots (top) and two examples of 2D-PAGE maps of Paca44 control and T3M4 treated samples, characterised by large values of the blue-coloured spots (bottom)

The conclusions driven by means of PCA show a very good agreement with those driven by PDQuest analysis of the 2D-PAGE maps (see chapter 5.3). The spots identified by PDQuest as the most characterising ones were also identified by means of PCA; in this last case, however, a larger number of spots was identified. Analysis of 2D-PAGE maps by dedicated softwares usually allows the identification of only those spots which exhibit at least a the two-fold variations in the protein content. PCA is a robust tool, which allows the detection of variation lower than the classical two-fold, since the changes due to the natural variability of the experimental steps are explained by the last PCs, which are not taken into account. The total information obtained by PCA is then larger than that obtained by dedicated softwares; for example, in the present case, the existence of the three patterns identified by PC_1 - PC_3 (Figs. 12, 13 and 14) could not be achieved by conventional PDQuest analysis.

5.4.4.3 Cluster Analysis

Since the first three principal components are able to separate the four classes of samples present in the dataset and to account for the reasons of the differences occurring between them, they are used for performing a cluster analysis, to verify how the samples are grouped by means of the first three PCs. The cluster analysis was performed by calculating a dendrogram with the Ward method; the distances were computed using the Euclidean distance.

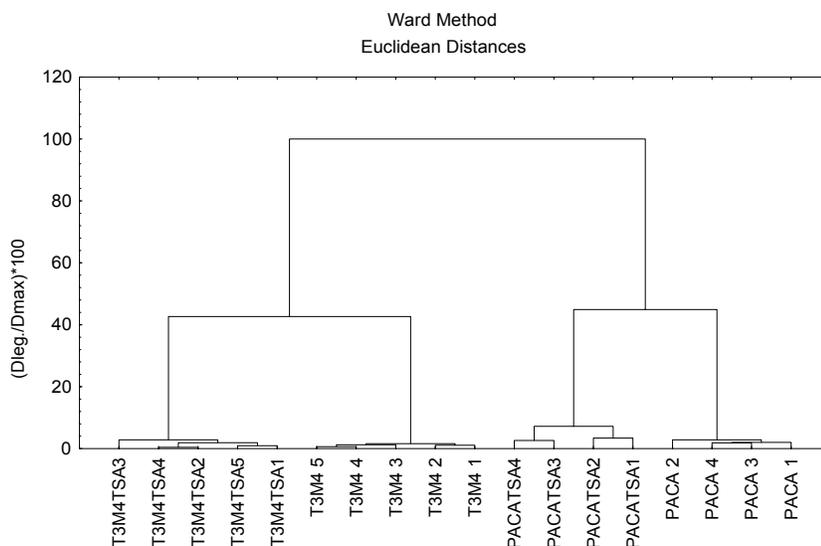


Figure 15. Dendrogram calculated on the basis of the first three PCs (Ward method, Euclidean distances).

Fig. 15 reports the obtained dendrogram; the ordinate label $(D_{leg}/D_{max}) * 100$ is a per cent dissimilarity scale expressing the linking distance (D_{leg}) of the groups of objects as a fraction of the maximum possible distance (D_{max}).

The samples are separated in two main groups: the first constituted by Paca44 cells line samples and the other given by the T3M4 cells line samples. The two groups are then separated in two sub-groups each, at a normalised distance of more than 40%: both the cell lines considered are correctly separated in control and TSA-treated samples. The dendrogram obtained by considering the first three PCs is then able to correctly separate the four classes of samples thus confirming the conclusions just driven by mean of PCA.

5.4.4.4 Mass-spectrometry

Mass-spectrometry analysis was performed only on the Paca44 cell line, due to the small size of the samples belonging to the T3M4 cell line. Some of the differentially expressed spots between control and TSA-treated Paca44 samples were identified by MALDI-TOF, as reported in table 3. The identified spots are represented in Fig. 16 as black squares and the SSP is indicated near each spot.

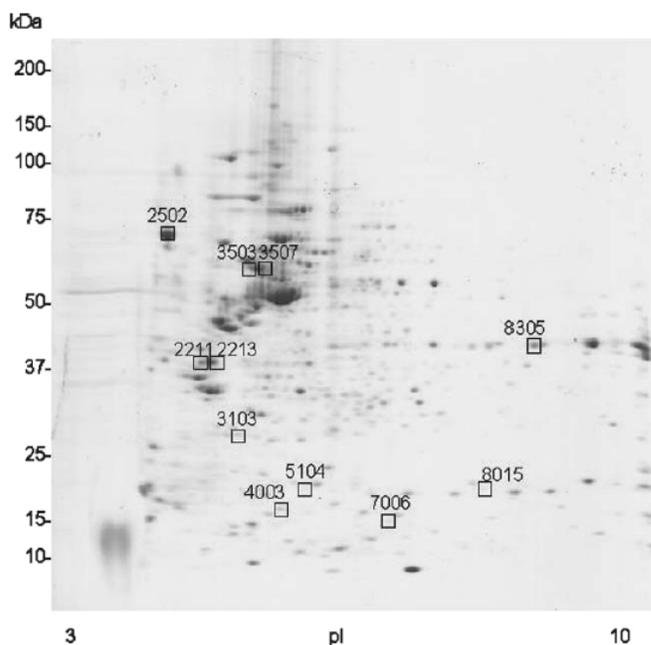


Figure 16. Spots identified by MS analysis: the number near each spot identifies the SSP number

Spot SSP	2D gel		Databank					Protein name	Accession number	Coverage (%)	No. of peptides	Variation
	Exp. Mr (Da)	Mr	Exp. pI	Theor. Mr.	Theor. pI	Z- Score	MOWSE- score					
2211	≈35,000	≈4.5	24,504	4.7	2.37	8.28	E12	Tropomyosin alpha four chain (Tropomyosin 4)	P07226	73.4	25	Decreased 2
2502	≈75,000	≈4.2	46,466	4.3	2.38	1.81	E19	Calreticulin precursor (CRP55) (Calregulin)	P27797	62	25	Decreased 2
2213	≈35,000	≈4.9	32,798	4.7	2.36	4.83	E14	Tropomyosin alpha three chain (Tropomyosin 3)	P12324	65	31	Decreased 3
3103	≈26,000	≈5.4	19,582	4.9	2.41	1.09	E9	Translationally controlled tumor protein (TCTP)	P13693	47	14	Decreased 3
8305	≈37,000	≈9	37,406 and 35,899	9.0 and 8.6	2.4	5.69 and 1.17	E10 E10	Heterogeneous nuclear ribonucleoproteins A2/B1 and Glyceraldehyde 3-phosphate dehydrogenase, liver (EC 1.2.1.12)	P22626 and P04406	54 and 48	16 and 16	Decreased 2
3507	≈55,000	≈5.6	51,736 and 46,142	5.0 and 5.0	2.35	3.38 and 4.19	E12 E9	ATP synthase beta chain, mitochondrial precursor (EC 3.6.3.14) and protein disulfide isomerase A6 precursor (EC 5.3.4.1)	P06576 and Q15084	41 and 41	16 and 13	Decreased 3
3503	≈55,000	≈5.4	51,736 and 49,671	5.0 and 4.8	2.32	4.24 and 1.20	E11 E10	ATP synthase beta chain, mitochondrial precursor (EC 3.6.3.14) and Tubulin beta-1 chain	P06576 and P07437	38 and 42	15 and 14	Decreased 2
5104	≈20,000	≈6.2	16,310 and 17,160	5.5 and 5.7	1.91 and 2.01	2.22 and 5.89	E9 E7	ARP2/3 complex 16 kDa subunit (P16-ARC) and Stathmin (Phosphoprotein p19) (pp19) (Oncoprotein 18)	O15511 and P16949	86 and 77	14 and 24	Increased 8
4003	≈18,000	≈5.8	16,040	5.5	2.34			Deduced protein product shows significant homology to coactosin	NCBIInr: AAA88022.1	57	17	Increased 2
8015	≈20,000	≈8	16,363	8.5	1.98	1.92	E9	UEV protein (ubiquitin-conjugating E2 enzyme variant)	NCBIInr: AAH28673	75	13	Increased 2
7006	≈13,500	≈6.5	13,654	6.4	2.38	1.02	E6	Hint Protein	P49773	59	6	Increased 3

Table 3. Summary of the identified proteins from Paca44 cell line 2D gels. For spot numbers, refer to Fig. 9.

5.4.4.5 Biological significance of some interesting identified proteins

Among the proteins which were identified by MALDI-TOF analysis, of particular interest are the down-regulated translationally-controlled tumour protein (TCTP) as well as the up-regulated protein stathmin (OP18). Their role will be briefly discussed below.

Translationally-controlled tumour protein (TCTP), a 3-fold down-regulated polypeptide, seems to be involved in tumour reversion, *i.e.* in the process by which some cancer cells lose their malignant phenotype. In a recent study, Tuynder *et al.* [87], have shown that TCTP is strongly down-regulated in the reversion processes of human leukemia and breast cancer cell lines.

Stathmin (Oncoprotein 18, OP18) was 8-fold up-regulated by the TSA treatment. Stathmin is a p53-regulated member of a novel class of microtubule-destabilizing proteins known to promote microtubule depolymerization during interphase and late mitosis [88]. Thus, high levels of stathmin could induce growth arrest at the G2 to mitotic boundary [89-90]. This suggest a cell cycle arrest at

the G2 phase of Paca44 cell treated with TSA. Due to its effect of inhibiting cell proliferation via a mitotic block, the up-regulation of stathmin here reported appears to be consistent with the anti-tumoural activity of TSA.

5.4.5 CONCLUDING REMARKS

PCA is applied here to a dataset constituted by 18 samples belonging to control and TSA-treated pancreatic ductal carcinoma cell lines (Paca44 and T3M4). PCA turned out to be a successful tool for the identification of the classes of samples present in the dataset; moreover, the loadings analysis allowed the identification of the regulatory spots, which characterise each group of samples. So, the treatment of both cell lines with Trichostatin-A showed an appreciable effect on the proteomic pattern of the control samples. The separation of the samples in four groups by mean of the first three PCs was also confirmed by Cluster Analysis.

The conclusion driven by PCA [91] resulted in good agreement with those obtained from the application of the differential analysis provided by PDQuest and confirmed the results shown in chapter 5.3.

5.5 DNA METHYLATION: PROTEOMIC PROFILING OF PANCREATIC DUCTAL CARCINOMA CELL LINES TREATED WITH 5-AZA-2'-DEOXYCYTIDINE

5.5.1 INTRODUCTION

Methylation of DNA occurs predominantly at the cytosine of CpG dinucleotides by DNA methyltransferase (DNMT) activities and is generally responsible for transcriptional repression. Recently, a link has been established between DNA methylation and histone deacetylation by studies showing that DNMT1 and MeCPs (methyl-CpG-binding proteins) interact physically with histone deacetylases [94-96] and that repression by methylated CpGs is partially relieved by TSA. Many tumor suppressor genes contain abnormal methylation of CpG islands in their promoters and appear to be silenced at the transcriptional level. To date, most of the studies investigating the induction of gene expression by the DNMT1 inhibitor 5-aza-2'-deoxycytidine (DAC) in pancreatic cancer have focused on the reactivation of a few genes, such as *p16*, *DAP kinase*, *RARBeta* [97]. The tumor suppressor gene *p16* is silenced by promoter methylation in up to 20% of both xenografted [98, 99] and primary pancreatic cancers. However, the identification of other genes selectively hypermethylated in this cancer may be of great importance for diagnostic purpose and for further understanding the biology of this tumor.

In the present study, we have investigated the growth susceptibility of the human pancreatic cancer cell line PaCa44 to DAC and determined the effect of this drug on protein expression, via proteomic profiling, as implemented by 2-D map analysis [55].

5.5.2 EXPERIMENTAL PROCEDURES

Cell culture and cell proliferation assay were performed by Department of Neurological and Visual Sciences staff, Section of Biochemistry, University of Verona.

5.5.2.1 Cells and growth conditions

PaCa44 human pancreatic adenocarcinoma cells were grown in RPMI 1640 supplemented with 20 mM glutamine and 10% fetal bovine serum (FBS; BioWhittaker, Italy) and were incubated at 37°C with 5% CO₂. Cells were seeded at a density of 12 x 10³/cm².

5.5.2.2 Cell proliferation assay

Cells were plated in 96-well cell culture plates at subconfluency (2×10^3 cells/well).

Cells were treated with 2.5 μM DAC for 24 h and cell proliferation was evaluated from 3 to 17 days after the beginning of the treatment. Cells were stained with Crystal Violet (Sigma), solubilized in PBS with 1% SDS, and measured photometrically ($A_{595\text{nm}}$) to determine cell viability. The crystal violet staining procedure is a simple and reproducible assay of cytotoxicity based on the growth rate reduction reflected by the colorimetric determination of the stained cells [125]. Three independent experiments were performed.

5.5.2.3 Treatment of cells for proteomic analysis and cell lysis

Preliminary results using different concentrations of DAC showed that a 2.5 μM DAC treatment for 24 h was able to induce a 50% arrest of cell growth 3 days after beginning of the treatment, and that maximal demethylation of DNA is obtained 6 days after treatment. Thus, we have chosen 2.5 μM concentration and 6 days after beginning of treatment as a good compromise to obtain the best cytotoxic effect and a reasonable amount of material for performing proteomic analysis. Protein extraction from cells untreated and treated with 2.5 μM DAC for 24 h and harvested after 6 days from the beginning of the treatment was performed with lysis buffer (40 mM Tris, 1% NP40, 1 mM Na_3VO_4 , 1 mM NaF, 1 mM PMSF, protease inhibitor cocktail). Cells were left in lysis buffer for 30 min on ice. After centrifugation at $14,000 \times g$ at 4°C , for removal of particulate material, the protein solution was collected and stored at -80°C until use.

5.5.2.4 Two-dimensional gel electrophoresis

Seventeen cm long, pH 3-10 immobilized pH gradient strips (IPG; Bio-Rad Labs., Hercules, CA, USA) were rehydrated for 8 h with 450 μL of 2-D solubilizing solution (7 M urea, 2 M thiourea, 5 mM tributylphosphine, 40 mM Tris and 20 mM iodoacetamide) containing 2 mg/mL of total protein from PaCa 44 cells [100]. Isoelectric focusing (IEF) was carried out with a Protean IEF Cell (Biorad, Hercules, CA, USA), with a low initial voltage and then by applying a voltage gradient up to 10000 V with a limiting current of 50 mA/strip. The total product time \times voltage applied was 70000 Vh for each strip and the temperature was set at 20°C . For the second dimension, the IPGs strips were equilibrated for 26 min by rocking in a solution of 6 M urea, 2% SDS, 20% glycerol, 375 mM Tris-HCl, pH 8.8. The IPG strips were then laid on an 8-18%T gradient SDS-PAGE with 0.5%

agarose in the cathode buffer (192 mM glycine, 0.1% SDS and Tris to pH 8.3). The anodic buffer was a solution of 375 mM Tris HCl, pH 8.8. The electrophoretic run was performed by setting a current of 2 mA for each gel for 2 hours, then 5 mA/gel for 1 h, 10 mA/gel for 20 h and 20 mA/gel until the end of the run. During the whole run the temperature was set at 11°C. Gels were stained overnight with Colloidal Coomassie blue (0.1% Coomassie Brilliant Blue G, 34% v/v methanol, 3% v/v phosphoric acid and 17% w/v ammonium sulphate), whereas destaining was performed with a solution of 5% acetic acid until a clear background was achieved. Five replicas for each condition (control and DAC-treated cells) were made. In addition, the same experiments were repeated twice.

5.5.2.5 Protein pattern analysis

The 2-DE gels were scanned with a GS-710 densitometer (Bio-Rad), and analyzed with the software PDQuest Version 7.1 (Bio-Rad, Laboratories, Hercules, CA, USA). A match set was created from the protein patterns of the two independent cellular extracts (control cell line, DAC-treated cell line). A standard gel was generated out of the image with the highest spot number. Spot quantities of all gels were normalized to remove non expression-related variations in spot intensity, so the raw quantity of each spot in a gel was divided by the total quantity of all the spots, in the same gel, that have been included in the standard. The results were evaluated in terms of spot OD (optical density). Statistical analysis of PDQuest allowed the study of proteins that were significantly increased or decreased in DAC-treated cell line. Forty-eight spots were found to be differently expressed in DAC-treated PaCa44, of which 13 spots were up-regulated (with a significance level α of 0.05) and 35 spots down-regulated (with $\alpha = 0.05$).

5.5.2.6 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 5mM 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. The extract was brought to dryness in Speedvac and then resuspended with 10 μ L of a H_2O 0.1 %TFA solution. The extracted peptides were further purified by Zip-Tip

pipetting tips (Millipore). The obtained solutions 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. 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.

5.5.3 RESULTS

The effect of the DNMT1 inhibitor DAC on the proliferation of the PaCa44 pancreatic adenocarcinoma cell line was examined by measuring cell viability with a colorimetric assay. Treatment with 2.5 μM DAC for 24 h inhibited cell proliferation for at least 10 days (Fig. 17).

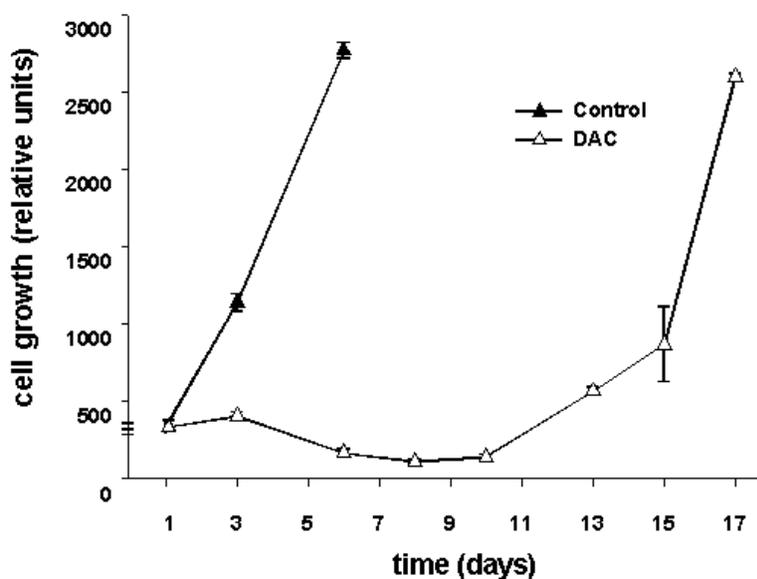


Figure 17. Effect of DAC on growth of the pancreatic cancer cell line PaCa44. Cells were seeded in 96-well plates at a density of 2×10^3 cells/well in the absence (control) or presence of 2.5 μM DAC for 24 h. Cells were then cultured for 17 days and the culture medium was changed three times a week. At various times, cell viability was measured photometrically as described in the materials and methods. Values are means of quadruplicate wells for each time point. The relative units of cell growth, in the ordinate, correspond to the optical density at 595 nm \times 1000.

Microscopic examination of the cells at the sixth day after treatment showed important morphological alterations such as distinctive polynucleated cells, larger dimension and heterogeneous shape which may suggest an extensive change of the gene expression profile (Fig. 18 a, b).

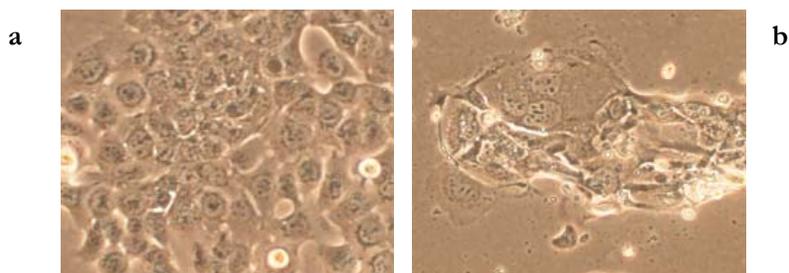


Figure 18. Changes induced in cell morphology by treatment with DAC after 6 days. Photomicrographs of unstained PaCa44 cells were performed with OLYMPUS-IX50, 10X-20X. (a) untreated cells; (b) DAC treated cells.

A standard 2-D map of the total protein extract from PaCa44 cell line, stained with colloidal Coomassie Blue, is shown in Fig. 19. About 700 polypeptide spots were revealed in the pH 3-10 interval with this medium-sensitivity stain.

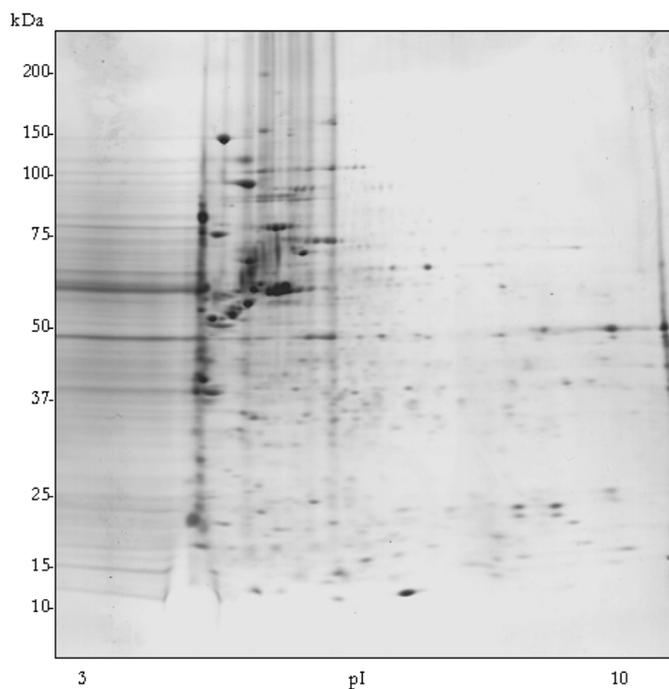


Figure 19. Master map of the pancreatic adenocarcinoma cell line Paca44, developed in the IPG 3-10 interval.

After matching the master map of the control PaCa44 cells with the master map of cells at 6 days after DAC treatment, a total of 45 polypeptide chains were found to be differentially expressed: 32 were down-regulated and 13 up-regulated.

All these spots were eluted and submitted to mass spectrometry analysis, by which thirty-six proteins could be identified (Fig. 20).

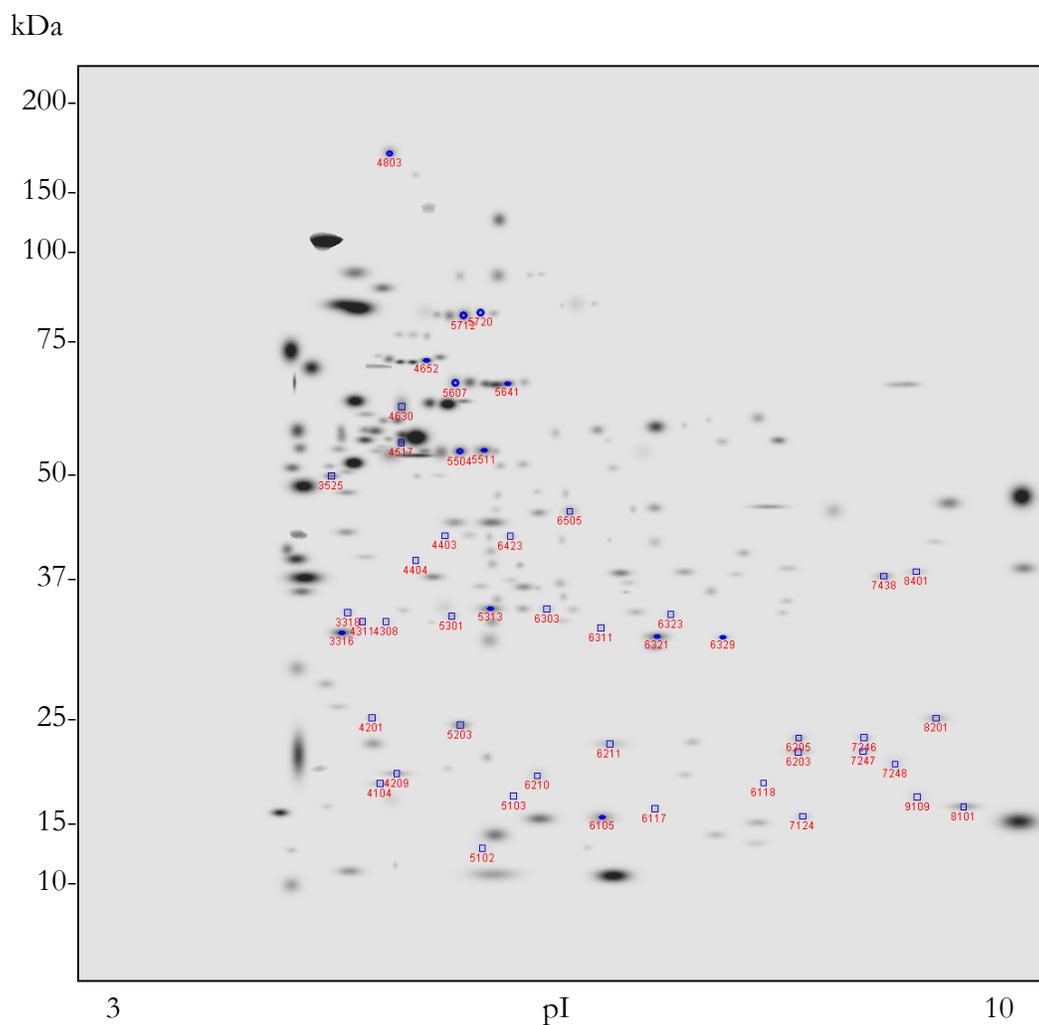


Figure 20. Standard two-dimensional map of Paca44 cells in the pH 3-10 IPG interval. The identified spots are marked across the pattern, with a square the proteins down-regulated, with a circle those up-regulated. Entry numbers refer to proteins identified by peptide fingerprinting and MS analysis (see Table 4 for their classification).

Of these thirty-six proteins, two were silenced, 22 were down regulated, and the remaining 12 were up regulated.

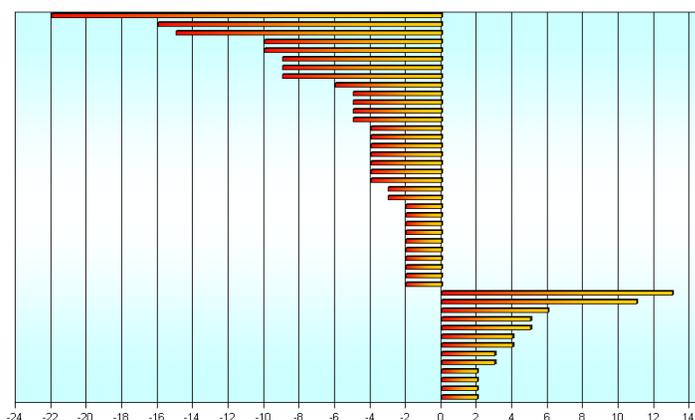
Their identification, experimental and theoretically predicted pI, and Mr values are given in Table 4.

Spot ID	Protein	Gene name	Swiss-Prot accession number	MW (expt/pred)	pI (expt/pred)	No. peaks matched (% coverage)	MASCOT score	Z-Score	Trend in DAC-treated cell	Observations and Reported Functions
8201	Cofilin	CFL1	P23528	18710/25000	8,5/9	10 (49)	89	2,37	OFF	tumor-related protein, whose expression is affected by DNA methylation; actin-modulating protein.
6118	Profilin I	PFN1	P07737	14923/18000	8,7/8	8 (46)	68	2,39	OFF	actin-binding protein involved in maintaining cell integrity, cell mobility, and tumor cell metastasis.
4209	Coactosin-like protein	COTL1	Q14019	15935/20000	5,54/5	7 (45)	102	2,31	decreased 22-fold	human pancreatic cancer antigen; actin-binding protein.
6211	Peptidyl-propyl cis-trans isomerase A	PPIA	P05092	17881/22000	8,2/6,5	16 (50)	94	2,38	decreased 16-fold	novel human hepatocellular carcinoma marker; overexpressed in metastatic tumors.
6117	Cystatin B	CSTB	P04080	11140/15000	7,7/7	7 (61)	110	2,3	decreased 15-fold	inhibitors of lysosomal cysteine proteinases (cathepsins); involved in tumor cell invasion and metastasis.
6210	no data in database								decreased 10-fold	
8101	no data in database								decreased 9-fold	
3525	Rho GDP-dissociation inhibitor 2	ARHGDI2	P52566	22988/50000	5,1/4,5	14 (61)	115	2,17	decreased 9-fold	inhibitor for the Ras-related Rho family GTPases; decreased in drug-induced apoptosis by proteolysis mediated by caspase-3.
7246	Peptidyl-propyl cis-trans isomerase A	PPIA	P05092	17881/22000	8,2/8,6	19 (65)	139	2,38	decreased 9-fold	novel human hepatocellular carcinoma marker; overexpressed in metastatic tumors.
9109	FK506-binding protein	FKBP1A	P20071	11820/15000	8,5/8,9	7 (40)	77	1,7	decreased 6-fold	physiologic regulator of the cell cycle, cell from mice knock-out manifest cell cycle arrest in G1 phase.
6205	Peptidyl-propyl cis-trans isomerase A	PPIA	P05092	17881/22000	8,2/8,4	18 (67)	96	2,32	decreased 5-fold	novel human hepatocellular carcinoma marker; overexpressed in metastatic tumors.
4308	Rho GDP-dissociation inhibitor 2	ARHGDI2	P52566	22988/32000	5,1/4,9	8 (40)	82	1,68	decreased 5-fold	inhibitor for the Ras-related Rho family GTPases; decreased in drug-induced apoptosis by proteolysis mediated by caspase-3.
3318	Rho GDP-dissociation inhibitor 1	ARHGDI1	P52565	23207/32000	5/4,6	10 (32)	82	1,67	decreased 4-fold	inhibitor for the Ras-related Rho family GTPases; decreased in drug-induced apoptosis by proteolysis mediated by caspase-3.
4104	Tubulin-Specific Chaperone A	TBCA	O75347	12716/16000	5,25/5	11 (61)	76	2,42	decreased 4-fold	involved in dimerization of the tubulin subunits, in degradation of defective tubulin subunits and in regulating tubulin polymerization.
4404	Proteasome activator complex subunit 2. Chloride intracellular channel protein	PSME2 CLIC1	Q9UL46 O00299	27362/37000 26923/37000	5,5/5,4 5,1/5,4	18 (59) 9 (43)	181		decreased 4-fold	modulates the proteasome-catalyzed production of antigenic peptides presented to the immune system on MHC class I molecules. chloride channel in nuclear membrane.
5203	Stathmin (phosphoprotein p19) Oncoprotein 18	STMN1	P16949	17171/24000	6/5,5	10 (49)	86	2,06	decreased 4-fold	microtubule-destabilizing proteins upregulated in neoplastic cells; downregulation in malignant cells interferes with their progression through cell cycle and abrogates their transformed phenotype.
7247	Peptidyl-propyl cis-trans isomerase A	PPIA	P05092	17881/22000	8,2/8,6	18 (71)	105	2,32	decreased 4-fold	novel human hepatocellular carcinoma marker; overexpressed in metastatic tumors.
6323	GTP-binding nuclear protein RAN	RAN	P17080	24423/32000	7,6/7,1	10 (48)	75	2,24	decreased 3-fold	involved in Cdc2/cyclin B activation and entry into mitosis when this process is coupled to the progression of S-phase.

6311	Proteasome subunit beta type 2	PSMB2	P49721	22836/30000	7/6,5	10 (54)	63	1,82	decreased 3-fold	subunit of proteasome involved in an ATP/ubiquitin-dependent nonlysosomal proteolytic pathway.
7248	Single-stranded DNA-binding Protein	SSBP1	Q04837	17260/20000	9,8/8,9	11 (70)	162	2,31	decreased 2-fold	participate in DNA replication in eukaryotic cells and possibly are intracellular regulators of proliferation.
4201	60S ribosomal protein L12	MRPL12	P52815	16394/25000	9,3/4,8				decreased 2-fold	component of the ribosomal elongation factor binding domain.
7438	Adenylate Kinase isoenzyme 2	AK2	P54819	26347/36000	8,2/8,6	17 (59)	87	2,28	decreased 2-fold	proapoptotic protein releases from the mitochondrial intermembrane space into the cytoplasm during apoptosis.
6505	Annexin I	ANXA1	P04083	38582/40000	7/6,3	19 (54)	127	2,4	decreased 2-fold	calcium- and phospholipid-binding protein upregulated in pancreatic carcinoma cell lines, involved in cell proliferation, mitogenic signal transduction and metastasis.
5301	Proteasome subunit beta type 4	PSMB4	P28070	29192/32000	5,9/5,5	11 (46)	79	2,33	decreased 2-fold	subunit of proteasome involved in an ATP/ubiquitin-dependent nonlysosomal proteolytic pathway
6203	Peptidyl-propyl cis-trans isomerase A	PPIA	P05092	17881/22000	8,2/8,4	18 (58)	67	2,26	decreased 2-fold	novel human hepatocellular carcinoma marker; overexpressed in metastatic tumors.
4403	Annexin III	ANXA3	P12429	36375/38000	5,8/5,5	17 (52)	132	2,36	decreased 2-fold	acts like inositol 1,2-cyclic phosphate 2-phosphohydrolase.
5720	Stress-70 Protein	HSPA9B	P38646	73681/80000	6,1/5,6	20 (28)	140	2,37	increased 13-fold	chaperone involved in cell proliferation, differentiation and tumorigenesis.
6329	Superoxide dismutase	SOD2	P04179	24722/30000	8,6/7,6	9 (37)	100	2,38	increased 11-fold	tumor suppressor gene in human pancreatic cancer, overexpression may be effective in growth suppression of pancreatic cancer.
6321	Superoxide dismutase	SOD2	P04179	24722/30000	8,6/7	10 (43)	107	2,19	increased 6-fold	tumor suppressor gene in human pancreatic cancer, overexpression may be effective in growth suppression of pancreatic cancer.
5641	Protein disulfide isomerase A3	GRP58	P30101	56782/60000	6,3/5,8	30 (51)	246	2,41	increased 5-fold	chaperone in the endoplasmic reticulum lumen, may regulate signaling by sequestering inactive and activated Stat3.
4652	60 kDa heat shock protein	HSPD1	P10809	61055/65000	5,8/5,4	21 (37)	151	2,38	increased 4-fold	chaperone that accelerates the maturation of pro-caspase-3 by upstream activator proteases during apoptosis.
6105	Beta-2-microglobulin	B2M	P01884	13706/15000	6,1/6,5	10 (60)	76	2,35	increased 4-fold	induces caspase-dependent apoptosis and cell cycle arrest.
5511	Actine	ACTB	P02570	41710/52000	5,3/5,6	14 (41)	105	2,38	increased 3-fold	involved in various types of cell motility.
4803	150kDa oxygen-regulated protein	HYOU1	Q9Y4L1	111336/160000	5,2/5	26 (25)	158	2,28	increased 3-fold	endoplasmic reticulum-associated protein; overexpression in pancreatic beta cells is correlated with insulin secretion.
5712	Stress-70 Protein	HSPA9B	P38646	73681/80000	6,1/5,5	18 (30)	153	2,4	increased 2-fold	chaperone involved in cell proliferation, differentiation and tumorigenesis.
3316	SET Protein	SET	Q01105	32103/31000	4,1/4,6	9 (31)	97	2,37	increased 2-fold	regulate G(2)/M transition by modulating cyclin B-CDK1 activity; interacts in vivo and in vitro with the cell cycle inhibitor p21(Cip1).
5607	Protein disulfide isomerase A3	GRP58	P30101	56782/60000	6,3/5,5	20 (36)	170	2,36	increased 2-fold	chaperone in the endoplasmic reticulum lumen, may regulate signaling by sequestering inactive and activated Stat3.
5504	Actine	ACTB	P02570	41710/52000	5,3/5,5	21 (51)	176	2,31	increased 2-fold	involved in various types of cell motility.

Table 4. Identified proteins from the Paca44 cell line 2-D gel.

The histogram in Fig. 21 displays the total changes detected of both down-regulated and up-regulated polypeptides.



decreased/increased proteins in DAC-treated PACA cell line

Figure 21. Histogram showing changes in concentration of Paca44 cell proteins after DAC treatment. The changes are computed as DAC-treated/control cells. A total of 45 proteins were found to change significantly, of which 32 were down-regulated (negative values) and 13 up-regulated (positive values).

Fig. 22 gives an example of a few spots exhibiting marked changes, such as spot No. 8201 (cofilin) that is silenced by DAC treatment.

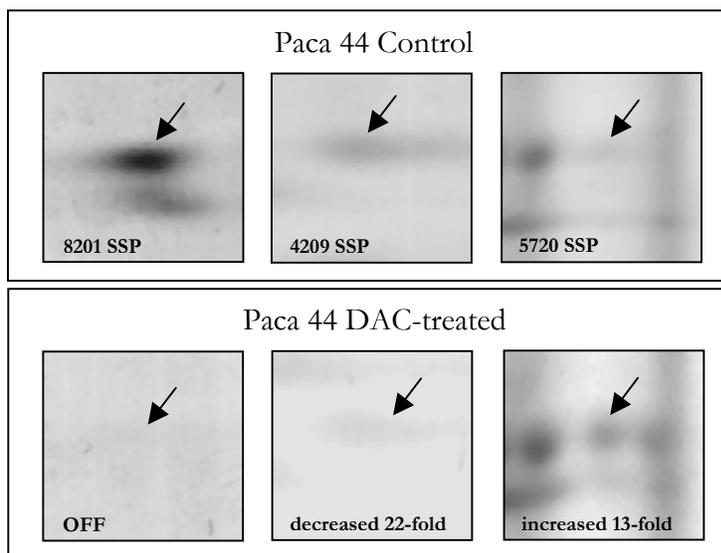


Figure 22. Comparison of 2-D gel patterns of some proteins in Paca44 cells treated with DAC (lower row) and in Paca44 control (upper row). The corresponding spot numbers are reported in Table 4.

DAC treatment caused up- or down-regulation in a number of proteins. The classification of these proteins according to the different biological processes in which they are involved is detailed in Table 5 and summarized in Fig. 23. It is of interest that most of the proteins which appear to be strongly influenced by DAC treatment are mostly involved in metabolism, cell growth and cell communication.

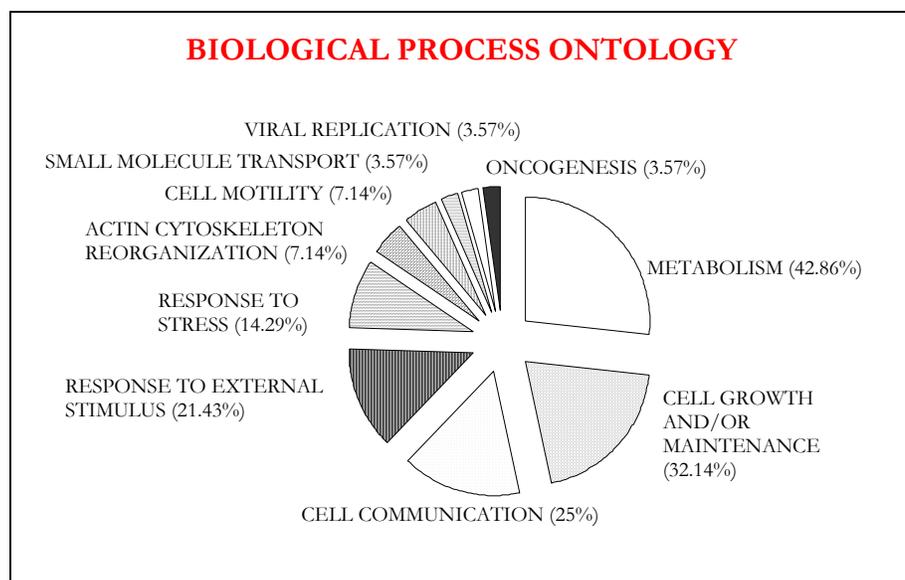


Figure 23. Classification of the proteins modulated by DAC treatment according to the biological function in which they are involved. obtained by the program FatiGO (<http://fatigo.bioinfo.cnio.es>). Note that single proteins may belong to more categories (see also Table 5), which explains the total sum being substantially larger than 100%.

- METABOLISM (42.86%):	Down-regulated: ANXA1, FKBP1A, PPIA, RAN, SSBP1, PSMB2, MRPL12, PSMB4.	Up-regulated: SET, HSPD1, GRP58, SOD2.
- CELL GROWTH AND/OR MAINTENANCE (32.14%):	Down-regulated: PFN1, CLIC1, SSBP1, STMN1, RAN, TBCA.	Up-regulated: SET, HSPD1, GRP58.
- CELL COMMUNICATION (25%):	Down-regulated: ANXA1, CFL1, STMN1, RAN, ARHGDIA, ARHGDIB.	Up-regulated: GRP58.
- RESPONSE TO EXTERNAL STIMULUS (21.43%):	Down-regulated: ANXA1, PPIA, PSME2, ARHGDIB.	Up-regulated: B2M, SOD2.
- RESPONSE TO STRESS (14.29%):	Down-regulated: ANXA1, PPIA.	Up-regulated: HYOU1, SOD2.
- ACTIN CYTOSKELETON REORGANIZATION (7.14%):	Down-regulated: CFL1, ARHGDIB.	Up-regulated: /
- CELL MOTILITY (7.14%):	Down-regulated: ANXA1.	Up-regulated: ACTB.
- SMALL MOLECULE TRANSPORT (3.57%):	Down-regulated: CLIC1.	Up-regulated: /
- VIRAL REPLICATION (3.57%):	Down-regulated: PPIA.	Up-regulated: /
- ONCOGENESIS (3.57%):	Down-regulated: /	Up-regulated: SET.

Table 5. Biological process ontology.

5.5.4 DISCUSSION

Differential analysis of protein expression is increasingly being used for profiling a number of pathological states, including cancer. The most widely accepted technique is still two-dimensional mapping, with charge and mass coordinates [55], followed by spot excision and mass spectrometry identification of tryptic digests [101]. Although differential profiling can be performed by a vast number of techniques (for a review, see [57]), in conventional 2-D maps the preferred method is still statistical analysis performed on master maps created from two independent cellular extracts (control *vs.* pathological or treated cells). The available software packages, although labor-intensive, produce reliable results, widely accepted by the scientific community.

In the present study we showed that DAC inhibits cell proliferation of the pancreatic adenocarcinoma cell line PaCa44, and attempted to address the molecular basis of this effect by the study of protein expression profiles of cells before and after DAC treatment. The significance of some of our findings is discussed below.

5.5.4.1 Silenced proteins

The proteomic analysis showed that 32 proteins were down-regulated at 6 days after DAC treatment. We speculate that at this time indirect effects, such as induction of repressors, may cause reduction in protein expression.

Two proteins were found to be silenced upon DAC treatment: cofilin and profilin 1 (Table 4). Cofilin is a widely distributed actin- modulating protein that has the ability to bind along the side of F (filamentous)-actin and to depolymerize it in a pH-dependent manner. Our data appear to be in agreement with those of Kanai *et al.* [102], who have recently reported strongly reduced mRNA expression for cofilin in DAC-treated human stomach and colorectal cancer cells. Of interest, Sinha *et al.* [103] reported that cofilin was over-expressed in pancreatic adenocarcinoma cell lines, which may suggest that suppression of cofilin might lead to cancer regression. Profilin 1 is a small actin-binding protein that is involved in diverse functions, such as maintaining cell structure integrity, cell motility, growth factor signal transduction and metastasis [104, 105]. The suppression of this protein in DAC-treated PaCa44 cells is consistent with the role of this drug in inhibiting tumor cell growth.

5.5.4.2 Down-regulated proteins

This is the largest set of modulated species, since it accounts for 32 polypeptides out of a total of 45 differentially expressed. Among these, six appear to be of particular relevance: coactosin-like protein (CLP), peptidyl-prolyl cis-trans isomerase A (PPIA), cystatin B, Rho GDP-dissociation inhibitor (Rho GDI-2), stathmin and annexin 1.

Coactosin-like protein (CLP,-22 folds) is a human filamentous actin (F-actin) binding protein [106]. CLP binds to actin filaments with a stoichiometry of 1:2 (CLP:actin subunits). Additionally, it binds to 5-lipoxygenase in a 1:1 molar ratio [107]. A recent report by Nakatsura *et al.* [108] indicated CLP as a tumor-associated antigen, suggesting this protein as a candidate for a vaccine for immunotherapy of cancer patients.

Peptidyl-prolyl cis-trans isomerase A (PPIA) is thought to control mitosis by binding to cell cycle regulatory proteins and altering their activity. It appears that the gene coding for this protein is almost exclusively overexpressed in aggressive metastatic or chemoresistant tumours [109]. In addition, PPIA has been recently proposed as a novel candidate marker for hepatocellular carcinoma [110]. Interestingly, all five isoforms of the PPIA are down-regulated upon DAC treatment (-16 fold for the main isoform, -9, -5, -4 and -2 folds for four additional, minor isoforms, see Table 4).

Cystatin B (also called stefin B) belongs to the cystatin superfamily of cysteine protease inhibitors and target cysteine proteases, such as cathepsin B. Cystatin B has been implicated in malignant progression [111] and alterations in its expression, processing and localization has been observed at various levels in malignant human tumors [112]. Progressively higher levels of cystatin B have been associated, e.g., with short survival in patients with colorectal cancer [113]. In human squamous cell carcinoma of the lung and in a number of other cancer types cystatin B has been found to be significantly increased as compared to normal tissues, and proved to be a prognostic factor [114]. This latter observation together with our findings that the level of cystatin B is markedly decreased (-15 fold) after DAC treatment suggests a role for this protein in maintaining tumor cell growth.

All of the three Rho GDP-dissociation inhibitor (Rho GDI-2) isoforms (-9 fold for the main isoform, -5 and -4 folds for two additional isoforms) were down regulated by DAC. Rho GDI-2 is a cytosolic protein participating in the regulation of both the GDP/GTP cycle and the membrane association/dissociation cycle of Rho/Rac proteins. Rho GDI proteins are in general over-expressed in breast tumors and human bladder cancer and their increased synthesis correlates with malignancy [115, 116]. In agreement with our findings, Kovarova *et al.* [117] have reported, by 2-D map analysis, followed by MALDI-MS, that Rho GDI proteins are significantly down-regulated in CEM T-lymphoblastic leukemia cell lines after treatment with bohemin.

Stathmin (-4 folds) is a member of a novel class of microtubule-destabilizing proteins that regulate the dynamics of microtubule polymerization and depolymerization. Stathmin is expressed at high levels in a wide variety of human cancers. Inhibition of stathmin expression in malignant cells interferes with their orderly progression through the cell cycle and abrogates their transformed phenotype. Thus, stathmin provides an attractive molecular target for disrupting the mitotic apparatus and arresting the growth of malignant cells [118]. The down regulation of stathmin upon DAC treatment gives support to the findings of Mistry and Atweh [118], who suggest stathmin as a therapeutic target in cancer therapy.

Annexin 1 is a member of a family of calcium- and phospholipids binding proteins related by amino acid sequence homology. Annexins 1 and 2 are substrates for protein tyrosine kinases. Both these proteins appear to be involved in mitogenic signal transduction and cell proliferation. All evidence so far gathered suggests that over-expression of annexins occurs in a variety of malignancies and well correlates with their progression [119-122].

5.5.4.3 Up-regulated proteins

The most interesting among the thirteen up regulated proteins is superoxide dismutase (SOD, +11 and +6 folds for a minor isoform), which is a component of the antioxidant system located in the mitochondrion. It has been recently shown that SOD could be a tumor suppressor in human pancreatic cancer, and suggested that delivery of the SOD gene might be used for pancreatic cancer gene therapy [123, 124].

5.5.5 CONCLUDING REMARKS

In conclusion, the identification in a representative pancreatic cancer cell line of proteins, whose expression is altered by DNMT1 inhibition, may serve to understand the biology of the pancreatic adenocarcinoma and the molecular mechanisms involved in the response to DAC treatment [125]. This information may represent a powerful tool for pancreatic cancer diagnosis and therapy with DNA methylation inhibitors.

5.6 BIBLIOGRAPHY

- [1] Niederhuber, J.E., Brennan, M.F., Menck, H.R., *Cancer* 1995, **76**: 1671-1677.
- [2] Warshaw, A.L., Fernandez-del Castillo, C., *N. Engl. J. Med.* 1992, **326**: 455-465.
- [3] Ahrendt, S.A., Pitt, H.A., *Oncology* 2002, **16**: 725-734.
- [4] Hruban, R.H. *et al.*, *Am. J. Surg. Pathol.* 2001, **25**: 579-586.
- [5] Anderson, K.E., Potter, J.D., Mack, T.M. in: Schottenfeld, D. & Fraumeni, J.J. (eds.), *Cancer Epidemiology and Prevention*, Oxford University Press, New York, 1996, 725-771.
- [6] Lynch, H.T. *et al.*, *Semin. Oncol.* 1996, **23**: 251-275.
- [7] Klein, W.M., Hruban, R.H., Klein-Szanto, A.J., Wilentz, R.E., *Mod. Pathol.* 2002, **15**: 441-447.
- [8] Moskaluk, C.A., Hruban, R.H., Kern, S.E., *Cancer Res.* 1997, **57**: 2140-2143.
- [9] Yamano, M. *et al.*, *Am. J. Pathol.* 2000, **156**: 2123-2133.
- [10] Luttges, J. *et al.*, *Am. J. Pathol.* 2001, **158**: 1677-1683.
- [11] Wilentz, R.E. *et al.*, *Cancer Res.* 2000, **60**: 2002-2006.
- [12] Heinmoller, E. *et al.*, *Am. J. Pathol.* 2000, **157**: 83-92.
- [13] Baylin, S.B., *Science* 1997, **277**: 1948-49.
- [14] Jones, P.A., Baylin, S.B., *Nat. Rev. Genet.* 2002, **3**: 415-28.
- [15] Takacs, M., Salamon, D., Myohanen, S., *et al.*, *Biol. Chem.* 2001, **382**: 699-705.
- [16] Tierney, R.J., Kirby, H.E., Nagra, J.K., *et al.*, *J. Virol.* 2000, **74**: 10468-79.
- [17] Robertson, K.D., *Curr. Topics Microbiol. Immunol.* 2000, **249**: 21-34.
- [18] Tao, Q., Swinnen, L.J., Yang, J., *et al.*, *Am. J. Pathol.* 1999, **155**: 619-25.
- [19] Hagan, C.R., Rudin, C.M., *Am. J. Pharmacogenomics* 2002, **2**: 25-35.
- [20] Nelson, W.G., De Marzo, A.M., Deweese, T.L., *et al.*, *Ann. N. Y. Acad. Sci.* 2001, **952**: 135-44.
- [21] Rideout, W.M., Eggan, K., Jaenisch, R., *Science* 2001, **293**: 1093-98.
- [22] Nelson, W.G., De Marzo, A.M., DeWeese, T.L., *Urology* 2001, **57**: 39-45.
- [23] Howell, C.Y., Bestor, T.H., Ding, F., *et al.*, *Cell* 2001, **104**: 829-38.
- [24] El-Osta, A., Wolffe, A.P., *Gene Expr.* 2000, **9**: 63-75.
- [25] Zuccotti, M., Garagna, S., Redi, C.A., *J. Endocrinol. Invest.* 2000, **23**: 623-29.
- [26] Feinberg, A.P., *Proc. Natl. Acad. Sci. USA* 2001, **98**: 392-4.
- [27] Feinberg, A.P., *Curr. Top Microbiol. Immunol.* 2000, **249**: 87-99.
- [28] Lengauer, C., Issa, J.P., *Mol. Med. Today* 1998, **4**: 102-03.
- [29] Jones, P.A., Laird, P.W., *Nat. Genet.* 1999, **21**: 163-67.

- [30] Rice, J.C., Allis, C.D., *Curr. Opin. Cell Biol.* 2001, **13**: 263-73.
- [31] Esteller, M., Catusas, L., Matias-Guiu, X., *et al.*, *Am. J. Pathol.* 1999, **155**: 1767-72.
- [32] Esteller, M., Hamilton, S.R., Burger, P.C., *et al.*, *Cancer Res.* 1999, **59**: 793-97.
- [33] Esteller, M., Sanchez-Cespedes, M., Rosell, R., *et al.*, *Cancer Res.* 1999, **59**: 67-70.
- [34] Tycko, B., *J. Clin. Invest.* 2000, **105**: 401-07.
- [35] Akhtar, M., Cheng, Y., Magno, R.M., *et al.*, *Cancer Res.* 2001, **61**: 2399-403.
- [36] Ahluwalia, A., Yan, P., Hurteau, J.A., *et al.*, *Gynecol. Oncol.* 2001, **82**: 261-68.
- [37] Ahluwalia, A., Hurteau, J.A., Bigsby, R.M., Nephew, K.P., *Gynecol. Oncol.* 2001, **82**: 299-304.
- [38] Barbieri, R., Mischiati, C., Piva, R., *et al.*, *Anticancer Res.* 1989, **9**: 1787-91.
- [39] Hanada, M., Delia, D., Aiello, A., *et al.*, *Blood* 1993, **82**: 1820-28.
- [40] Ray, J.S., Harbison, M.L., McClain, R.M., Goodman, J.I., *Mol. Carcinog.* 1994, **9**: 155-66.
- [41] Rao, P.M., Antony, A., Rajalakshmi, S., Sarma, D.S., *Carcinogenesis* 1989, **10**: 933-37.
- [42] Vachtenheim, J., Horakova, I., Novotna, H., *Cancer Res.* 1994, **54**: 1145-48.
- [43] Feinberg, A.P., Vogelstein, B., *Biochem. Biophys. Res. Commun.* 1983, **111**: 47-54.
- [44] Feinberg, A.P., Vogelstein, B., *Nature* 1983, **301**: 89-92.
- [45] Chandler, L.A., DeClerck, Y.A., Bogenmann, E., Jones, P.A., *Cancer Res.* 1986, **46**: 2944-49.
- [46] Stephenson, J., Akdag, R., Ozbek, N., Mufti, G.J., *Leuk. Res.* 1993, **17**: 291-93.
- [47] Aparicio, A., Weber, J.S., *Curr. Opin. Investig. Drugs* 2002, **3**: 627-633.
- [48] Kelly, W.K., O'Connor, O.A., Marks, P.A., *Expert Opin. Investig. Drugs* 2002, **11**: 1695-1713.
- [49] Kouzarides, T., *Curr. Opin. Genet. Dev.* 1999, **9**: 40-48.
- [50] Archer, S.Y., Hodin, R.A., *Curr. Opin. Genet. Dev.* 1999, **9**: 171-174.
- [51] Tsukiyama, T., Wu, C., *Curr. Opin. Genet. Dev.* 1997, **7**: 182-191.
- [52] Gregory, P.D., Horz, W., *Eur. J. Biochem.* 1998, **251**: 9-18.
- [53] Jacobson, S., Pillus, L., *Curr. Opin. Genet. Dev.* 1999, **9**: 175-184.
- [54] Marks, P.A., Richon, V.M., Rifkind, R.A., *J. Natl. Cancer Inst.* 2000, **92**: 1210-1216.
- [55] Righetti, P.G., Stoyanov, A.V., Zhukov, M.Y., *The Proteome Revisited*, Elsevier, Amsterdam, 2001.
- [56] James, P. (ed.), *Proteome Research: Mass Spectrometry*, Springer, Berlin, 2001.
- [57] Hamdan, M., Righetti, P.G., *Mass Spectrom. Reviews* 2002, **21**: 287-302.
- [58] Colombo, E., Marine, J.C., Danovi, D., Falini, B., Pelicci, P.G., *Nat. Cell Biol.* 2002, **4**: 529-533.
- [59] Okuda, M., *Oncogene* 2002, **21**: 6170-6174.
- [60] Yang, C., Maignel, D.A., Carrier, F., *Nucleic Acid Res.* 2002, **30**: 2251-2260.

- [61] Shackleford, G.M., Ganguly, A., MacArthur, C.A., *BMC Genomics* 2001, **2**: 8-9.
- [62] Bischof, D., Pulford, K., Mason, D.Y., Morris, S.W., *Mol. Cell. Biol.* 1997, **17**: 2312-2325.
- [63] Tuynder, M., Susini, L., Prieur, S., Besse, S., Fiucci, G., Amson, R., Telerman, A., *Proc. Natl. Acad. Sci. USA* 2002, **99**: 14976-14981.
- [64] Guillame, E., Pineau, C., Evrad, B., Dupaix, A., Moertz, E., Sanchez, J.C., Hochstrasser, D.F., Jegou, B., *Proteomics* 2001, **1**: 880-889.
- [65] Chung, S., Kim, M., Choi, W., Chung, J., Lee, K., *Cancer Lett.* 2000, **156**: 185-190.
- [66] Liu, H., Wang, Y., Zhang, Y., Song, Q., Di, C., Chen, G., Tang, J., Ma, D., *Biochem. Biophys. Res. Commun.* 1999, **254**: 203-210.
- [67] Chen, Y., Sun, R., Han, W., Zhang, Y., Song, Q., Di, C., Ma, D., *FEBS Lett.* 2001, **509**: 191-196.
- [68] Rui, M., Chen, Y., Zhang, Y., Ma, D., *Life Sci.* 2002, **71**: 1771-1778.
- [69] Alli, E., Vash-Babula, J., Yang, J.M., Hait, W.N., *Cancer Res.* 2002, **62**: 6864-6869.
- [70] Cassimeris, L., *Curr. Opin. Cell Biol.* 2002, **14**: 18-24.
- [71] Mistry, S.J., Atweh, G.F., *Mt. Sinai J. Med.* 2002, **69**: 299-304.
- [72] Cecconi, D., Scarpa, A., Donadelli, M., Palmieri, M., Hamdan, M., Astner, H., Righetti, P.G., *Electrophoresis* 2003, **24**: 1871-1878.
- [73] Anderson, N.L., Hofmann, J.P., Gemmell, A., Taylor, J., *Clin. Chem.* 1984, **30**: 2031-2036.
- [74] Tarroux, P., Vincens, P., Rabilloud, T., *Electrophoresis* 1987, **8**: 187-199.
- [75] Leonard, R.B., Mayer, J., Sasser, M., Woods, M.L., Mooney, B.R., Brinton, B.G., Newcombgrayman, P.L., Carroll, K.C., *J. Clin. Microbiol.* 1995, **33**: 2723-2727.
- [76] Johansson, M.L., Quednau, M., Ahrne, S., Molin, G., *Int. J. Syst. Bacteriol.* 1995, **45**: 670-675.
- [77] 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.
- [78] Boon, N., De Windt, W., Verstraete, W., Top, E.M., *FEMS Microbiol. Ecol.* 2002, **39**: 101-112.
- [79] Alike, J.E., Akenova M.E., Fatokun C.A., *Genet. Res. Crop Evol.* 1995, **42**: 393-399.
- [80] Anderson, N.L., EsquerBlasco, R., Richardson, F., Foxworthy, P., Eacho, P., *Toxicol. Appl. Pharm.* 1996, **137**: 75-89.
- [81] Vohradsky, J., *Electrophoresis* 1997, **18**: 2749-2754.
- [82] Kovarova, H., Radzioch, D., Hajduch, M., Sirova, M., Blaha, V., Macela, A., Stulik, J., Hernychova, L., *Electrophoresis* 1998, **19**: 1325-1331.

- [83] 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.
- [84] Dewettinck, K., Dierckx, S., Eichwalder, P., Huyghebaert, A., *Lait* 1997, **77**: 77-89.
- [85] Kovarova, H., Hajduch, M., Korinkova, G., Halada, P., Krupickova, S., Gouldsworthy, A., Zhelev, N., Strnad, M., *Electrophoresis* 2000, **21**: 3757-3764.
- [86] Massart, D.L., Kaufman, L., in: Elving P.J., Winefordner J.D. (Editors), *The interpretation of Analytical Chemical Data by the Use of Cluster Analysis*, Wiley, New York, 1983.
- [87] Tuynder, M., Susini, L., Prieur, S., Besse, S., Fiucci, G., Amson, R., Telerman, A., *Proc. Natl. Acad. Sci. USA* 2002, **99**: 14976-14981.
- [88] Alli, E., Vash-Babula, J., Yang, J.M., Hait, W.N., *Cancer Res.* 2002, **62**: 6864-6869.
- [89] Cassimeris, L., *Curr. Opin. Cell Biol.* 2002, **14**: 18-24.
- [90] Mistry, S.J., Atweh, G.F., *Mt. Sinai J. Med.* 2002, **69**: 299-304.
- [91] Marengo, E., Robotti, E., Righetti, P.G., Cecconi, D., *Anal. Bioanal. Chemistry* 2004, **379**: 992-1003.
- [92] Wandschneider, S., Fehring, V., Jacobs-Emeis, S., Thiesen, H.J., Löhr, M., *Electrophoresis* 2001, **22**: 4383-4390.
- [93] Möller, A., Soldan, M., Volker, U., Maser, E., *Toxicology* 2001, **160**: 129-138.
- [94] Rountree, M.R., Bachman, K.E., Baylin, S.B., *Nat. Genet.* 2000, **25**: 269-277.
- [95] Robertson, K.D., Ait-Si-Ali, S., Yokochi, T., Wade, P.A., Jones, P.L., Wolffe, A.P., *Nat. Genet.* 2000, **25**: 338-342.
- [96] Jones, P.L., Veenstra, G.J., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N.M., Strouboulis, J., Wolffe, A.P., *Nat. Genet.* 1998, **19**: 187-191.
- [97] Ueki, T., Toyota, M., Sohn, T., Yeo, C.J., Issa, J.P., Hruban, R.H., Goggins, M., *Cancer Res.* 2000, **60**: 1835-1839.
- [98] Sorio, C., Bonora, A., Orlandini, S., Moore, P., Cristofori, P., Dal Negro, G., Gaviraghi, G., Falconi, F., Pederzoli, P., Scarpa, A., *Virchows Archiv.* 2001, **438**: 154-159.
- [99] Schutte, M., Hruban, R.H., Geradts, J., Maynard, R., Hilgers, W., Rabindran, S.K., Moskaluk, C.A., Hahn, S.A., Schwarte-Waldhoff, I., Schmiegel, W., Baylin, S.B., Kern, S.E., Herman, J.G., *Cancer Res.* 1997, **57**: 3126-3130.
- [100] Herbert, B., Galvani, M., Hamdan, M., Olivieri, E., McCarthy, J., Pedersen, S., Righetti, P.G., *Electrophoresis* 2001, **22**: 2046-2057.

- [101] James, P. (Ed.), *Proteome Research: Mass Spectrometry*, Springer, Berlin, 2001.
- [102] Kanai, Y., Ushijima, S., Saito, Y., Nakanishi, Y., Sakamoto, M., Hirohashi, S., *J. Cancer Res. Clin. Oncol.* 2001, **127**: 697-706.
- [103] Sinha, P., Hutter, G., Kottgen, E., Dietel, M., Schadendorf, D., Lage, H., *Electrophoresis* 1999, **20**: 2952-2960.
- [104] Hu, E., Chen, Z., Fredrickson, T., Zhu, Y., *Exptl. Nephrol.* 2001, **9**: 265-274.
- [105] Feldner, J.C., Brandt, B.H., *Exptl. Cell Res.* 2002, **272**: 93-108.
- [106] Provost, P., Doucet, J., Stock, A., Gerisch, G., Samuelsson, B., Radmark, O., *Biochem. J.* 2001, **359**: 255-263.
- [107] Provost, P., Doucet, J., Hammarberg, T., Gerisch, G., Samuelsson, B., Radmark, O., *J. Biol. Chem.* 2001, **276**: 16520-16527.
- [108] Nakatsura, T., Senju, S., Ito, M., Nishimura, Y., Itoh, K., *Eur. J. Immunol.* 2002, **32**: 826-836.
- [109] Meza-Zepeda, L.A., Forus, A., Lygren, B., Dahlberg, A.B., Godager, L.H.,k South, A.P., Marenholz, I., Lioumi, M., Florenes, V.A., Maelansmo, G.M., Serra, M., Mischke, D., Nizetic, D., Ragoussis, J., Tarkkanen, M., Nesland, J.M., Knuutila, S., Myklebost, O., *Oncogene* 2002, **21**: 2261-2269.
- [110] Lim, S.O., Park, S.J., Kim, W., Park, S.G., Kim, H.J., Kim, Y.I., Sohon, T.S., Noh, J.H., Jung, G., *Biochem. Biophys. Res. Commun.* 2002, **291**: 1031-1037.
- [111] Calkins, C.C., Sameni, M., Koblinski, J., Sloane, B.F., Moin, K., *J. Histochem. Cytochem.* 1998, **46**: 745-751.
- [112] Kos, J., Lah, T.T., *Oncol. Reports* 1998, **5**: 1349-1361.
- [113] Kos, J., Krasovec, M., Cimerman, N., Nielsen, H.J., Christensen, I.J., Brunner, N., *Clin. Cancer Res.* 2000, **6**: 505-511.
- [114] Ebert, E., Werle, B., Julke, B., Kopitaer-Jerala, N., Kos, J., Lah, T., Abrahamson, M., Spiess, E., Ebert, W., *Adv. Exptl. Med. Biol.* 1997, **421**: 259-265.
- [115] Seraj, M.J., Harding, M.A., Zgildea, J.J., Welch, D.R., Theodorescu, D., *Clin. Exp. Metastasis* 2000, **18**: 519-525.
- [116] Fritz, G., Brachetti, C., Bahlmann, F., Schmidt, M., Kaina, B., *Br. J. Cancer* 2002, **87**: 635-644.
- [117] Kovarova, H., Hajduch, M., Korinkova, G., Halada, P., Krupickova, S., Gouldsworth, A., Zhelev, N., Srnad, M., *Electrophoresis* 2000, **21**: 3757-3764.
- [118] Mistry, S.J., Atweh, G.F., *Mt. Sinai J. Md.* 2002, **69**: 299-304.

- [119] De Coupade, C., Gillet, R., Bennoun, M., Briand, P., Russo-Marie, F., Solito, E., *Hepatology* 2000, **31**: 371-380.
- [120] Wu, W., Tang, X., Hu, W., Lotan, R., Hong, W.K., Mao, L., *Clin. Exp. Matastasis* 2002, **19**: 319-326.
- [121] Rhee, H.J., Kim, G.Y., Huh, J.W., Kim, S.W., Na, D.S., *Eur. J. Biochem.* 2000, **267**: 3220-3225.
- [122] Pencil, S.D., Toth, M., *Clin. Exp. Matastasis* 1998, **16**: 113-121.
- [123] Weydert, C., Roling, B., Liu, J., Hinkhouse, M.M., Ritchie, J.M., Oberley, L.W., Cullen, J.J., *Mol. Cancer Therapy* 2003, **2**: 361-369.
- [124] Cullen, J.J., Weydert, C., Hinkhouse, M.M., Ritchie, J.M., Domann, F.E., Spitz, D., Oberley, L.W., *Cancer Res.* 2003, **63**: 1297-303.
- [125] Cecconi, D., Astner, H., Donadelli, M., Palmieri, M., Missiaglia, E., Hamdan, M., Scarpa, A., Righetti P.G. *Electrophoresis* 2003, **24**: 4291-4303.

CHAPTER 6

DRUG DEPENDENCE: PROTEOMIC CHANGES IN RAT SERUM AFTER CHRONIC NICOTINE TREATMENT

6.1 INTRODUCTION

Nicotine is widely accepted as the addictive constituent of tobacco, endowed with reinforcing properties, that is responsible for tobacco addiction (Fig. 1a) [1,2].

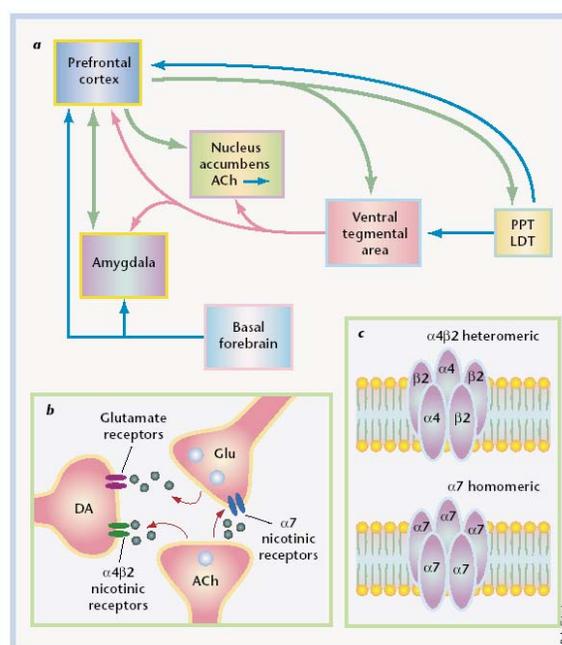


Figure 1. (a) Neural systems implicated in addiction and in the behavioral sensitization. Dopamine neurons (pink), located in the midbrain ventral tegmental area, project to limbic forebrain and cortical regions (such as the prefrontal cortex, amygdala and nucleus accumbens). These structures in turn are connected by glutamate pathways (green). The cholinergic pathways (blue) include the cell bodies in the pedunculo-pontine nucleus (PPT) and lateral dorsal tegmentum (LDT), located in the pons, that project to the ventral tegmental area and also up to cortical regions. A second major cholinergic pathway arises from the basal forebrain and reaches widespread cortical and limbic regions. A third source of forebrain acetylcholine constitutes the intrinsic neurons of the nucleus accumbens, a brain reward region. (b) Interactions between acetylcholine (ACh), glutamate (Glu) and dopaminergic (DA) cells in the midbrain. ACh release in this area could activate both presynaptic nicotinic $\alpha 7$ receptors (located on Glu terminals), and postsynaptic $\alpha 4 \beta 2$ nicotinic receptors (located on DA cell bodies). Activated ACh, released by drugs of abuse in the ventral tegmental area, could amplify the excitatory effects of Glu and DA and set in motion the addiction process. (c) Two different types of neuronal nicotinic receptors commonly found in the cholinergic pathways and proposed to be involved in the brain's response to drugs of abuse.

It acts via activation of nicotinic acetylcholine receptors (nAChRs) that are present in neuronal and in non-neuronal cells (Fig. 1c) [3-6].

However, the pharmacodynamics of nicotine are complex and results can be dependent on the duration of exposure to the drug; moreover tolerance to many effects of nicotine develops rapidly [7,8]. Due to its complexity, the mechanism of action of nicotine is still not fully understood, neither in the central nervous system nor in the peripheral system.

Particularly, in the periphery, several factors are involved in the pathological effects of smoke in humans, but there is less information about the pharmacological effects of nicotine itself. From the literature, we know that nicotine produces some sympathetic effects in the periphery, such as tachycardia, vasoconstriction, mydriasis, bronchiolar dilatation, decrease in gastrointestinal motility and glycogenolysis. These effects are mediated by the nicotine stimulation of postganglionic sympathetic neurons and adrenal medulla [9], where nAChRs are expressed in animals [10] as well as humans [11].

Recently, some studies in rodents and rabbit models and cell preparations showed that nicotine, when administered acutely, induced angiogenic and arteriogenic effects, which could be mediated through non-neural nAChRs [12].

It has been also shown that nicotine could stimulate peripheral blood monocytes by increasing their adhesion, chemotaxis and transendothelial migration [13,14], factors that are important for angiogenesis. These findings would suggest a role of nicotine in pathological (i.e. inflammation, tumor and atherosclerosis) as well as therapeutic angiogenesis (i.e. protection against vascular ischemia).

There is increasing evidence that nicotine can affect both humoral and cell-mediated branches of the immune system (Fig. 2).

Altered changes would be characterised by a decrease in inflammation, a decrease antibodies response and reduction in T cell-receptor-mediated signalling [14,15]. Few studies showed that chronic exposure to nicotine impaired immune response in rats [16,17]. It has been reported that continuous *sub cutaneous* treatment with nicotine for 3-4 weeks reduced in fact the antibody-forming cell response, producing an immunosuppression [17].

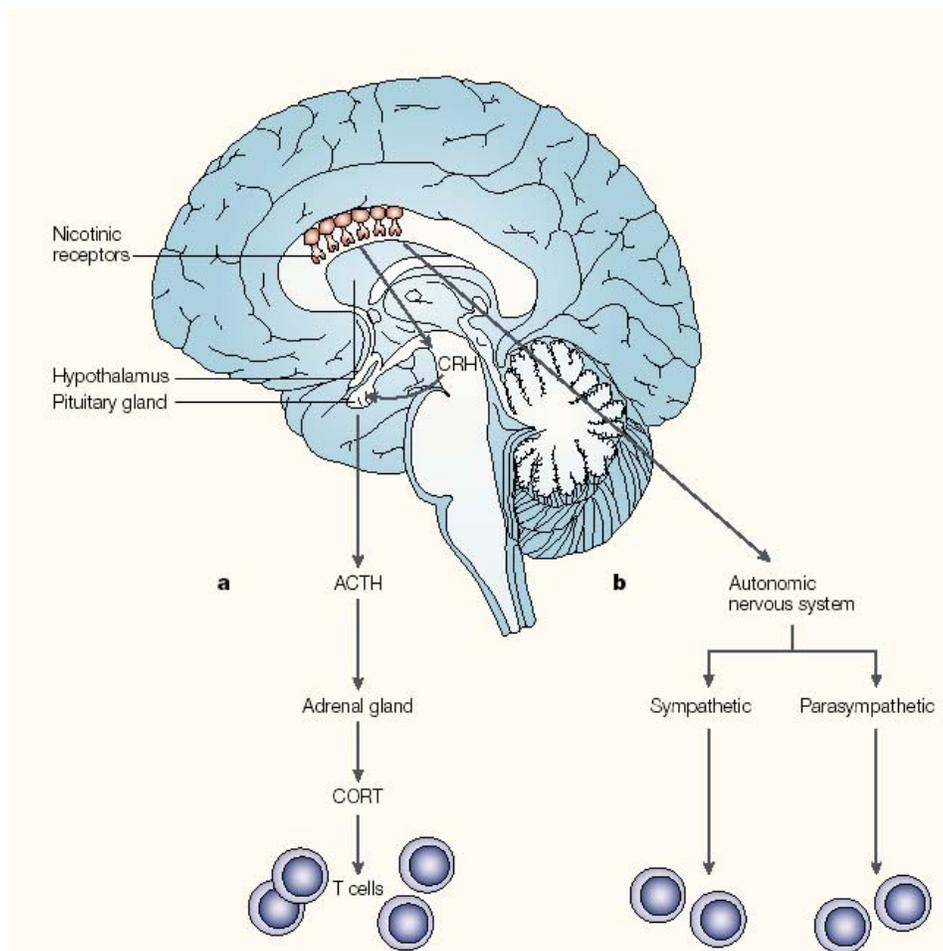


Figure 2. A simplified scheme of possible communication between the CNS and the immune system through nicotinic acetylcholine receptors. Nicotine from cigarette smoke enters the brain and interacts with nicotinic receptors in the brain. Activation of nicotinic receptors could modulate the immune response by either of two pathways: **a)** activation of the hypothalamus–pituitary–adrenal axis, whereby corticotropin-releasing hormone (CRH) from the hypothalamus stimulates the release of adrenocorticotropic hormone (ACTH) from the pituitary gland, which, in turn, stimulates the production of glucocorticoids (CORT) by the adrenal gland at the end of the chain increased levels of CORT suppress the immune system; and **b)** activation of the autonomic nervous system, which connects the brain directly to the visceral target tissues, including lymphoid tissues, through sympathetic and parasympathetic innervations. Noradrenaline from the sympathetic nerve terminals might modulate T-cell function through adreno-receptors that are present on T cells. The role of the parasympathetic nervous system in the regulation of T-cell function is not clear.

Furthermore, recent evidence suggests that non-neuronal nAChRs are also involved in the regulation of vital cell functions, such as mitosis, differentiation, organisation of the cytoskeleton, cell-cell contact, locomotion, and migration [3].

The aim of this study was to establish if chronic nicotine treatment, in rats, could induce proteomic changes in peripheral blood serum, providing information about pathways involved in the nicotine pharmacological action, for example investigating the involvement of nicotine in the pathogenesis of inflammatory and immune system diseases.

For detecting proteins differentially expressed after nicotine treatment, a differential proteomic technique was applied, based on a 2-D gel electrophoresis/mass spectrometry approach. Serum was analysed from rats treated with nicotine in comparison with control rats, treated with saline solution. The chronic nicotine treatment was performed according to regimens similar to those known to have a stimulatory effect on locomotor activity and to produce a sensitisation of the mesolimbic dopamine system, mechanism involved in addiction development [18].

The same nicotine treatment is also known to induce axonal distribution changes in rat ventral tegmental area neurofilaments [19,20]. After 2-D gel analysis and matching, a statistical test was performed for detecting up- and down-regulated proteins by nicotine treatment, thus eliminating sample and gel variability.

Differentially expressed proteins in serum were identified by MALDI TOF mass spectrometry peptide fingerprinting. Database query was performed by using three different software packages, by comparing different algorithms to confirm protein identifications.

6.2 EXPERIMENTAL PROCEDURES

6.2.1 ANIMALS AND NICOTINE TREATMENT

Nicotine treatment and blood collection was performed by the Department of Drug Dependency & Behavioural Neurochemistry Staff, at the Psychiatry Centre of Excellence for Drug Discovery, GlaxoSmithKline.

Male Wistar rats (200-300 g) were purchased from Charles River, Calco, Italy. Rats were housed in three or four per cage and maintained with food and water *ad libitum* in a temperature-controlled environment on a 12 h/12 h light-dark cycle with light on at 6:00 a.m. Rats were treated with nicotine (0.4 mg/Kg/d in saline) or with saline solution (n=5/group) via subcutaneous injection for 14 days. The animals were sacrificed under deep CO₂ anaesthesia, blood was collected intracardially and immediately heparinised for leukocyte isolation. The research complied with national legislation and with the company policy on the Care of Use of Animals and related codes of practice.

6.2.2 SERUM PREPARATION

On the 14th day of treatment, 500 μL blood samples were taken from the tail vein of rats kept on fasting since the night before in order to avoid lipid interference. Samples were collected in Microtainer tubes (BD Biosciences, Erembodegen, Belgium), incubated for 30 min on an orbital shaker and centrifuged for 15 min at $1000 \times g$ at room temperature. From each sample about 200 μL of serum (about 60 mg/mL of total protein) were recovered and stored at -20°C until analysis.

6.2.3 SERUM 2-D ELECTROPHORESIS

One hundred μL of rat serum (containing about 60 mg/mL total protein) were added with 0.4 mL of a 2-D solubilising solution, containing 7 M urea, 2 M thiourea, 40 mM Tris, 5 mM tributyl phosphine. Samples were incubated 90 minutes at room temperature to reduce protein disulphide bonds. Cysteine thiolic groups were then alkylated by adding 20 mM iodoacetamide (Sigma) and incubating for 90 minutes at room temperature. Samples were dialysed in a regenerated cellulose membrane with a molecular mass cut-off of 3.5 KDa; dialysis was performed overnight in distilled water. Then chaotropes and surfactant were added to a final concentration of 7 M urea, 2 M thiourea, 20 mM Tris and 2% CHAPS (Sigma). Eighteen-cm long, pH 3-10 non-linear immobilised pH gradient strips (Amersham Pharmacia Biotech) were rehydrated for 8 h with 450 μL of the sample solution (final total protein concentration of 6 mg/mL) containing traces of bromophenol blue to monitor the electrophoretic run. Isoelectric focusing was conducted at 20°C for 60 000 Vh using a Protean IEF Cell (Bio-Rad Laboratories, Hercules, CA, USA) with a low initial voltage and then by applying a voltage gradient up to 10000 V with a limiting current of 50 μA /strip. For the second dimension, the immobilised pH gradient 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). The immobilised pH gradient strips were then laid on a 7-20% T gradient SDS-PAGE with 0.5% agarose dissolved in cathodic buffer [192 mM glycine (Sigma), 15 mM Tris, 0.1% SDS, pH 8.3]. The anode buffer consisted of 375 mM Tris-HCl pH 8.8. Gels were run at 10°C and 2 mA/gel for 2 hour, 5 mA/gel for 1 hour and 10 mA/gel overnight. Gels were stained 20 hours with colloidal Coomassie blue G-250 [0.1% Blue G-250 (BDH Laboratory Supplies, Poole, England), 34% v:v methanol, 3% v:v o-phosphoric acid 17% w:v ammonium sulphate (Sigma)] and destained in 5% acetic acid (Sigma). Images were captured with a GS-710 imaging densitometer (Bio-Rad) selecting a high resolution ($42.3 \mu\text{m}^2$ for pixel) scanning.

6.2.4 PROTEIN PATTERN DIFFERENTIAL ANALYSIS

2-D gel analysis was performed by PDQuest software (Bio-Rad), version 6.2. Each gel was analysed for spot detection, background subtraction and protein spot OD intensity quantification (spot quantity definition). The gel image showing the higher number of spots and the best protein pattern was chosen as a reference template, and spots in a standard gel were then matched across all gels. Spot quantity values were normalised in each gel by dividing the raw quantity of each spot by the total quantity of all the spots included in the standard gel. Gels were divided in two separate groups (control and nicotine-treated samples) and, for each protein spot, the average spot quantity value and its variance coefficient in each group were determined. A Student's t-test was performed in order to compare the two groups and identify sets of proteins that showed a statistically significant difference with a confidence level of 0.05. Among the differentially expressed spots, when matched spots were missing in some gels, the statistical test analysis was performed by GraphPad Prism version 3.00 for Windows (GraphPad Software, San Diego, CA, USA), for checking the results.

6.2.5 MALDI-TOF MASS SPECTROMETRY

Each selected spot was carefully cut and destained with 2×10 min washing steps in 50% acetonitrile (Sigma) (v:v), 50% 5 mM Tris pH 8.5 followed by a third wash with 5 mM Tris pH 8.5 for 10 minutes. The gel discs were dried in a Speedvac sc110A device (Thermo Savant, NY, USA) for 1h at room temperature and then covered with 15 μ L of Sequencing Grade Modified Trypsin (Promega, Promega Corporation, Madison, WI, USA) (0.02 mg/mL) in NH_4HCO_3 buffer (40 mM, pH 8.5) and left at 37 °C overnight. The gel pieces were then crushed and peptides were extracted twice in 50 μ L of 50% acetonitrile, 50% H_2O with 1% formic acid (v:v) and a third time with 50 μ L of acetonitrile. The extractions were conducted in an ultrasonic bath for 15 min. The three extraction solutions were mixed and evaporated to dryness in the Speedvac device and the residues were dissolved in 10 μ L of H_2O with 0.1% of trifluoroacetic acid. For an additional purification, the samples were cleaned by using ZIP-TIP C18 (Millipore Bedford, MA, USA). Two μ L of the resulting solution were mixed with an equivalent volume of matrix solution, prepared fresh every day by dissolving 10 mg/mL α -cyano-4-hydroxycinnamic acid in acetonitrile:ethanol (1:1, v:v). One μ L of the resulting mixture was loaded onto the MALDI sample plate and allowed to dry. Measurements were

performed using a ToFSpec 2E MALDI-TOF instrument (Micromass, Manchester, UK), operated in reflector mode, with an accelerating voltage of 20 kV.

Two different analysis tools were used for protein identification by peptide masses in order to have always a double confirmation of the results. Mascot software (Matrix Science, UK) [21], that incorporates a probability-based scoring, was used to search SwissProt, trEMBL and NCBI non redundant databases with Mammalia (Mammals) as taxonomic category. Profound (Version 4.10.5) searches were performed by using the same databases with the same taxonomic category. Database queries were carried out for monoisotopic peptide masses, using a peptide mass tolerance of ± 150 ppm, 1 as maximum number of missed tryptic cleavages, molecular mass range depending on the apparent Mr of the protein spot in the selected 2-D gel and all pI ranges. Allowed modifications included oxidation of Met and alkylation of Cys residues by iodoacetamide.

The criteria for positive identification of proteins were set as follow: protein scores are considered significant when the probability of random events is less than 1 in 20 ($p < 0.05$).

For MASCOT analyses, this means that proteins with a Probability Based MOWSE Score (the reported Mascot score in the identification table 1) greater than the threshold calculated by the tool algorithm were considered significant identifications.

For Profound it roughly corresponds to a Z-Score greater than 1.65.

A third tool, ProteinProbe (Version 3.4, BioLynx, Micromass, UK), was used as an additional confirmation of the identified proteins and the resulting MOWSE score was considered significant if there was a difference of two orders of magnitude in the score value between the first and the second hit.

The result of a database search was significant if the protein was ranked as best hit and there was a sequence coverage of at least 20%. Positive identification of the protein was assigned only if the mass deviation of matched peptides was constant over the whole mass range. Matched peptide masses were evenly distributed throughout the complete protein amino acid sequence, and the identified protein molecular mass and pI values corresponded approximately to experimental values, with some exceptions.

6.3 RESULTS

6.3.1 DIFFERENTIAL PROTEOMIC ANALYSIS

In order to obtain good focusing of proteins in the first dimension, the serum samples were dialysed for desalting. Ten maps, one for each animal in the experiment (five controls, S1-S5, and five nicotine-treated rats, N1-N5), were performed. The separated protein spots were visualised on 2-D gels by colloidal Coomassie blue staining, which allows good reproducibility and protein spot quantification for comparison analysis. Figure 3 shows an example of a Coomassie-stained 2-D gel of the control group.

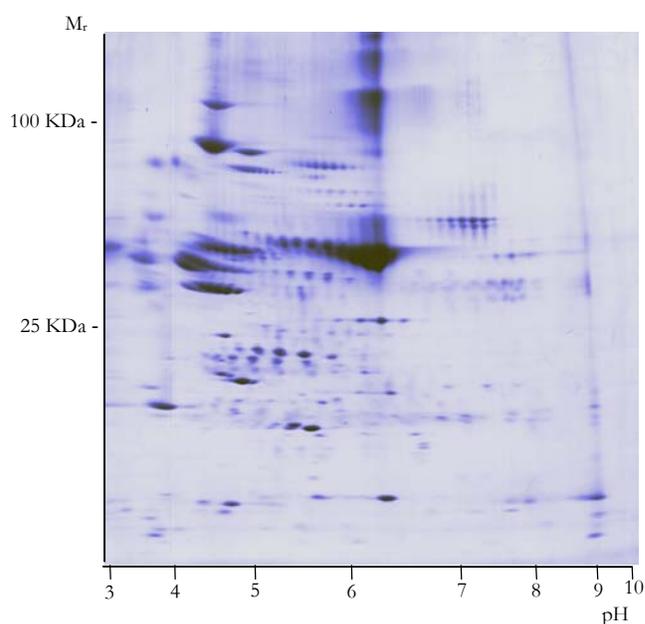


Figure 3. Examples of Coomassie-stained 2-D gel of serum samples.

The ten serum 2-D gels corresponding to the ten animals involved in the experiment were analysed and matched in the same matchset. Control and nicotine-treated samples were then divided in two groups (five gels per group) and statistically compared as reported in section 3.3. A student's T-Test was performed in order to identify sets of proteins that are differentially expressed with a confidence level of 0.05. The results were checked again by extracting the normalised quantity value data and performing the statistical analysis with Prism software.

Thirty-six protein spots resulted statistically significant by PDQuest analysis, after this revision: 7 up-regulated and 29 down-regulated in nicotine-treated respect to control samples (Figure 4).

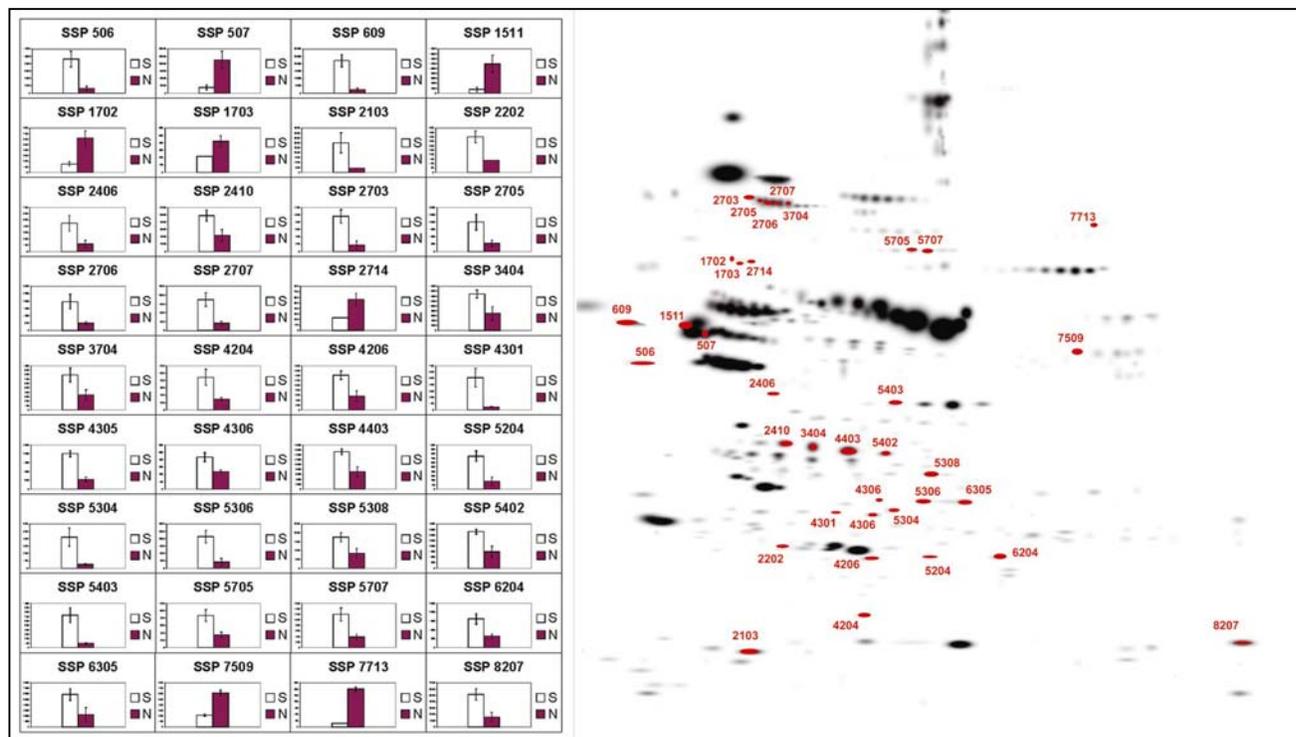


Figure 4. Reference image for serum matchset. Significant spots resulting from statistical analysis are numbered. Histograms represent the corresponding spot intensity (average value between replicates) in saline (left bar) and in nicotine (right bar purple) groups.

6.3.2 IDENTIFICATION OF PROTEINS

Spots selected from the differential analysis of serum were subjected to peptide fingerprinting analysis by MALDI TOF mass spectrometry for protein identification. In table 1 the identity of the successfully identified proteins corresponding to up- or down-regulated spots in the 2D-gel maps for serum are shown, together with the identification parameters and the indication of the main role(s) of the protein in cellular pathway and metabolism.

Spot	Exp. pI	Exp. Mr (Da)	ID	Name	Theor. pI, Mr.	MASCOT	MOWSE-Score	Z-score	No. of matched peptides	% sequence coverage	Function of the protein	Observed expression after nicotine administration
2 (507)	~ 4	~ 50000	CPI1_RAT	Contrapsin-like protease inhibitor1	5,4 46562	154	4,34 E15	2,34	23	52	inhibits kallikreins and trypsin but not chymotrypsin or elastase.	Upregulated
4 (1511)	~ 4	~ 50000	CPI1_RAT	Contrapsin-like protease inhibitor1	5,4 46562	169	3,10 E10	2,35	17	36	inhibits kallikreins and trypsin but not chymotrypsin or elastase.	Upregulated
5 (1702)	~ 5	~ 70000	AFAM_RAT	Afamin	5,9 69336	245	3,48 E21	2,37	37	52	possible role in the transport of yet unknown ligand.	Upregulated
6 (1703)	~ 5	~ 70000	AFAM_RAT	Afamin	6,1 69336	126	1,72 E14	2,38	25	31,6	possible role in the transport of yet unknown ligands.	Upregulated
11 (2703)	~ 5,5	~ 100000	O35802	Inter-alpha inhibitor H4 heavy chain	6,4 103607	254	4,05 E17	2,42	29	30	plasma proteinase inhibitor.	Downregulated
12 (2705)	~ 5,5	~ 100000	O35802	Inter-alpha inhibitor H4 heavy chain	6,4 103607	224	7,44 E20	2,41	34	35	plasma proteinase inhibitor.	Downregulated
14 (2707)	~ 5,5	~ 100000	O35802	Inter-alpha inhibitor H4 heavy chain	6,4 103607	285	7,46 E21	2,37	35	37	plasma proteinase inhibitor.	Downregulated
16 (3404)	~ 6	~ 40000	Q63332	Alpha-1-Macroglobulin	6,8 167160		1,14 E9	2,37	15	11	inactivation of proteinases released in response to tissue injury, necrosis or inflammation.	Downregulated
23 (4403)	~ 6,5	~ 40000	Q63332	Alpha-1-Macroglobulin	6,8 167160		2,29 E9	2,4	15	10	inactivation of proteinases released in response to tissue injury, necrosis or inflammation.	Downregulated
10 (2410)	~ 6	~ 40000	Q63332	Alpha-1-Macroglobulin	6,8 167160		3,91 E9	2,36	14	11	inactivation of proteinases released in response to tissue injury, necrosis or inflammation.	Downregulated
27 (5308)	~ 7,5	~ 38000	Q63332	Alpha-1-Macroglobulin	6,8 167160		8,46 E11	2,29	17	14	inactivation of proteinases released in response to tissue injury, necrosis or inflammation.	Downregulated
28 (5402)	~ 7	~ 40000	Q63332	Alpha-1-Macroglobulin	6,8 167160		1,25 E9	2,41	16	12	inactivation of proteinases released in response to tissue injury, necrosis or inflammation.	Downregulated
30 (5705)	~ 7	~ 80000	GELS_MOUSE	Gelsolin	5,7 80747	248	4,85 E10	2,34	27	23	controls cytoskeletal assembly and disassembly.	Downregulated
32 (6204)	~ 8	~ 25000	GSHP_RAT	Plasma glutathione peroxidase	8,4 25393	167	3,29 E8	2,35	14	45	protects cells and enzymes from oxidative damage, by catalyzing the reduction of hydrogen peroxide, lipid peroxides and organic hydroperoxide, by glutathione.	Downregulated
24 (5204)	~ 7,5	~ 25000	GSHP_RAT	Plasma glutathione peroxidase	8,4 25393	143	8,23 E7	2,41	13	45	protects cells and enzymes from oxidative damage, by catalyzing the reduction of hydrogen peroxide, lipid peroxides and organic hydroperoxide, by glutathione.	Downregulated

Table 1. Summary of the identified proteins from serum 2D-gels. For spot numbers, refer to Figure 4. Matching peptides vs. total number of peptides submitted to database search, sequence coverage and score resulted for peptide fingerprinting matching are listed. Experimental pI and Mr values, as derived from gel separations, are indicated in the "Exp. pI and Mr" columns. Protein ID, theoretical pI and Mr and protein functions were obtained by the SWISS-PROT and NCBI data bases. MASCOT, MOWSE and Z Scores (output of the identification softwares Mascot, ProteinProbe and ProFound respectively) are measures of the statistical significance of the identification hits.

Not all the differentially expressed spots were identified, because some of them did not give a good MS spectrum after the tryptic digest of the protein: 15 spots were identified in serum samples. A few

different spots in serum resulted in the same identified protein, thus indicating that the modulation of expression occurs in different isoforms of the same protein, often with the same Mr but different pI. All the reported identifications had significant score values resulting from the three selected identification software packages. Alpha-1-Macroglobulin was an exception as only a small fragment of the protein was identified by two of the three algorithms (Profound and ProteinProbe software). Mascot software did not recognise the fragment as a part of the protein, probably due to the low coverage percentage of the identified sequence with respect to the total protein.

Overall, the main impact of chronic treatment with nicotine can be observed in acute-phase proteins (indicating an overall effect on inflammatory cascades and more in general on the immune system), in proteins belonging to the oxidative stress metabolism. The indication arising from the analysis of the individual changes observed and their potential implications are discussed below.

6.4 DISCUSSION

This is the first investigation on a proteomic scale of the changes caused by nicotine in the peripheral blood of laboratory animals. In our approach, we have chosen to evaluate the change in the proteomic complement of the serum fraction by 2-D gel proteomic analyses. The aim of the present approach is to obtain a broader picture of the consequence of nicotine administration on the organism. The proteomic analysis of serum has highlighted a main impact of chronic nicotine treatment on the inflammatory response, in agreement with the known anti-inflammatory properties of nicotine [22-24]. Indeed, it is known that cigarette smokers have a lower incidence of several inflammatory diseases and it has been recognised that the psychoactive component of smoke, nicotine, is acting as an anti-inflammatory agent and moderates the severity of some inflammatory responses [22]. Accordingly, we have shown upon chronic nicotine administration an increase in the serum content of contrapsin-like protease inhibitor 1 (or kallikrein binding protein), a negative acute-phase protein that has a protective effect during acute phase inflammation and is able to produce vasodilatation effects [27]. The expression of additional proteins involved in the acute phase inflammatory response was also modified by the treatment. As an example, inter alpha Trypsin inhibitor, a type II acute phase protein in humans, was found to be significantly decreased in serum, whilst it is known that its concentration is increased in patients suffering from acute phase processes. Alpha-1 macroglobulin, that is required to bind and inactivate proteinases released in response to

tissue injury and inflammation, is decreased, indicating that the acute phase synthetic response of the liver is down-regulated by chronic exposure to nicotine.

In answering the question of whether nicotine is the major causal agent of oxidative damage triggered by cigarette smoke, a number of investigations have shown a strong relationship between nicotine and oxidative stress. In polymorphonuclear cells (PMN), nicotine has been shown to potentiate phorbol 12-myristate 13-acetate-induced superoxide anion generation [27]. In addition, *in vitro* strong evidence has been provided that nicotine produces oxidative stress at high concentrations, as indicated by a decrease in reduced glutathione levels (suggesting a disturbed antioxidant defence mechanism) [28]. *In vivo*, oral administration of nicotine to rat fed a high-fat diet was found to decrease the radical detoxifying enzyme superoxide dismutase (SOD), catalase and glutathione reductase [29].

In our context, changes observed in serum for gelsolin, plasma glutathione peroxidase and afamin, seem to indicate a clear disturbance of the oxidative stress balance and a series of potential compensatory mechanisms occurring in the organism.

Gelsolin is an actin modulating protein that exists both as cytoplasmic and plasma secreted form, which is known to be reduced in pulmonary failures. In a murine model of oxidant injury [30], the degree of hyperoxic injury was inversely related to gelsolin levels suggesting an association between gelsolin depletion and the pathogenesis of acute oxidant lung injury in patients. Decrease in the activity of blood and erythrocyte glutathione peroxidase (a key enzyme secreted in plasma that catalyses reduction of hydrogen peroxide) has been also observed in studies conducted on smokers, where oxidative stress is clearly exacerbated [31-33]. In two of the above studies [31, 32], a parallel reduction in the plasma level of vitamin E has been shown. Interestingly, in our study, we have observed a concomitant increase in the vitamin E binding protein afamin, that has been shown to be able to increase the survival of cortical neurons under oxidative stress *in vitro* that can be read as a compensatory mechanism.

All of the changes highlighted in the blood by our proteomic approach can originate from the primary interaction of nicotine with nicotinic acetylcholine receptors (nAChRs) that are present either in neuronal and non-neuronal cells. In other words, the effects that we have observed in the periphery could be a direct reflection of the interaction with non-neuronal binding sites present in peripheral tissues or an indirect impact on the periphery mediated through the central nervous system (CNS).

Indeed, there is a large body of evidence that indicates a bi-directional communication between the CNS and the immune system, and, in addition to the potential direct effects on nicotine on the leukocyte response, a CNS-mediated strong modulation of the neuroendocrine and immune systems can be expected.

A series of studies on morphine and nicotine have shown that drugs of abuse can modulate both immune cell activities and the hypothalamic-pituitary-adrenal (HPA) axis function by acting through specific receptors. Morphine, for instance, suppresses the immune system via activation of central opioid receptors, whilst nicotine and nicotinic agonists appear to act predominantly at peripheral receptors, presumably at the autonomic ganglia. As for morphine, acutely administered nicotine has been shown to increase the plasma corticosterone levels, and to decrease the peripheral blood leukocyte proliferative response to the T cell mitogen concanavalin A. Nicotine however, in contrast to morphine, does not alter blood lymphocyte response when administered centrally, suggesting a mechanism that is independent of its ability to activate the HPA axis [34].

Evidence of the involvement of nicotinic receptors in linking inflammation and leukocyte functions with oxidative stress arises from studies on the effects of nicotine agonists on microcirculation. As an example, the nicotine-induced leukocyte rolling and adhesion can be blocked in a murine model by mecamylamine, as well as by anti-P-selectin antibodies, and by SOD, suggesting the action through a ganglionic-type nicotinic receptor and via P-selectin and reactive oxygen radical-dependent mechanisms [26].

The overall effect of nicotine on the oxidative stress can be mediated through different mechanisms and results into opposite effects in the CNS and in the periphery. Nicotine has been shown to be able to display both pro-oxidant and antioxidant properties depending on the system being investigated [35], because one effect is probably compensating the other. Indeed, nicotine administration can result in oxidative stress upon induction of the generation of reactive oxygen species in the periphery and in CNS. In parallel, it has been shown *in vitro* that nicotine displays intracellular antioxidant properties through the activation of nicotinic receptors [36, 37] but also extracellularly by acting as a iron-binding radical scavenger [38], therefore suggesting at the CNS level a potential role for nicotine in neuroprotection.

In agreement with all the above indications, our results point out the existence of a number of potentially different mechanisms triggered by nicotine, both mediated by binding to specific central and peripheral receptors, or by additional chemical or pharmacological properties of nicotine, that can result in physiologically opposite actions.

With our global approach [39], we have chosen not to dissect out a single mechanism, or to focus on a specific target district (that could result in an more precise but less comprehensive observation), but to provide a wider picture of the overall effects of a “physiologically” relevant nicotine dose regimen in rodents. By selecting plasma we aimed at identify potential non-invasive markers for the CNS and non-CNS mediated activities of nicotine.

By a large-scale proteomic analysis of serum from rats subjected to chronic treatment with nicotine we have shown an overall impact on a series of proteins involved in a variety of cellular and metabolic pathways. This could suggest, as a main effect, a general reduction in the inflammatory response with a concomitantly increased unbalance of the oxidative stress metabolism in the periphery, both implying a potential impact on leukocyte activation in different directions. This approach has therefore highlighted a series of potential changes reflecting *in vivo* centrally and peripherally mediated nicotine effects that deserve to be further tested by using nicotine antagonists and in dose-response studies for better understanding the physiological mechanisms underlying the observed changes. In perspective, similar studies in humans can open up an unprecedented source of protein changes to be validated as non-invasive markers for the neuro-immunomodulatory activities exerted by the psychoactive component of smoke.

6.5 BIBLIOGRAPHY

- [1] Benowitz, N.L., *British J. Addiction* 1991, **86**: 495-499.
- [2] Stolerman, I.P., Jarvis, M.J., *Psychopharm.* 1995, **117**: 2-10.
- [3] Wessler, I., Kirkpatrick, C. J., Racke, K., *Clin. Exp. Pharm. Physiol.* 1999, **26**: 198-205.
- [4] Grando, S.A., Horton, R.M., Pereira, E.F., Diethelm-Okita, B.M., George, P.M., Albuquerque, E. X., Conti-Fine, B.M., *J. Invest. Dermatology* 1995, **105**: 774-781.
- [5] Maus, A.D., Pereira, E.F., Karachunski, P.I., Horton, R.M., Navaneetham, D., Macklin, K., Cortes, W.S., Albuquerque, E.X., Conti-Fine, B.M., *Mol. Pharm.* 1998, **54**: 779-788.
- [6] Macklin, K.D., Maus, A.D., Pereira, E.F., Albuquerque, E.X., Conti-Fine, B.M., *J. Pharm. Exp. Therap.* 1998, **287**: 435-439.
- [7] Benowitz, N.L., *N. Engl. J. Med.* 1988, **319**: 1318-1330.
- [8] Fattinger, K., Verotta, D., Benowitz, N.L., *J. Pharm. Exp. Therap.* 1997, **281**: 1238-1246.
- [9] Haass, M., Kubler, W., *Card. Drugs & Therapy* 1997, **10**: 657-665.
- [10] Rogers, K.R., Fernando, J.C., Thompson, R.G., Valdes, J.J., Eldefrawi, M.E., *Anal. Bioch.* 1992, **202**: 111-116.
- [11] Mousavi, M., Hellstrom-Lindahl, E., Guan, Z.Z., Bednar, I., Nordberg, A., *Life Sciences* 2001, **70**: 577-590.
- [12] Heeschen, C., Weis, M., Aicher, A., Dimmeler, S., Cooke, J.P., *J. Clin. Invest.* 2002, **110**: 527-536.
- [13] Heeschen, C., Weis, M., Cooke, J.P., *J. Am. Coll. Cardiology* 2003, **41**: 489-496.
- [14] Sopori, M. L., Kozak, W., Savage, S. M., Geng, Y., Kluger, M.J., *Adv. Exp. Med. Biology* 1998, **437**: 279-289.
- [15] Sopori, M.L., Kozak, W., *J. Neuroimmunology* 1998, **83**: 148-156.
- [16] Kalra, R., Singh, S.P., Kracko, D., Matta, S. G., Sharp, B.M., Sopori, M.L., *J. Pharm. Exp. Therap.* 2002, **302**: 935-939.
- [17] Geng, Y., Savage, S.M., Razani-Boroujerdi, S., Sopori, M.L., *J. Immunology* 1996, **156**: 2384-2390.
- [18] Balfour, D.J., Benwell, M.E., Birrell, C.E., Kelly, R.J., Al Aloul, M., *Pharm. Bioch. Behavior* 1998, **59**: 1021-1030.
- [19] Bunnemann, B., Terron, A., Zantedeschi, V., Merlo, P. E., Chiamulera, C., *Europ. J. Pharm.* 2000, **393**: 249-253.
- [20] Sbarbati, A., Bunnemann, B., Cristofori, P., Terron, A., Chiamulera, C., Merigo, F., Benati, D., Bernardi, P., Osculati, F., *Europ. J. Neuroscience* 2002, **16**: 877-882.

- [21] Perkins, D.N., Pappin, D.J., Creasy, D.M., Cottrell, J.S., *Electrophoresis* 1999, **20**: 3551-3567.
- [22] Sopori, M., *Nature Reviews Immunology* 2002, **2**: 372-377.
- [23] Singh, S.P., Kalra, R., Puttfarcken, P., Kozak, A., Tesfaigzi, J., Sopori, M.L., *Tox. Appl. Pharm.* 2000, **164**: 65-72.
- [24] Sopori, M. in: Americ S. and Brioni J.D. (eds) *Neuronal Nicotinic Receptors : Pharmacology and Therapeutic Opportunities*, John Wiley & Sons Press, New York, pp 197-209.
- [25] Chao, J., Miao, R.Q., Chen, V., Chen, L.M., Chao, L., *Biol. Chem.* 2001, **382**: 15-21.
- [26] Yong, T., Zheng, M.Q., Linthicum, D.S., *J. Neuroimmunology* 1997, **80**: 158-164.
- [27] Gillespie, M.N., Owasoyo, J.O., Kojima, S., Jay, M., *Toxicology* 1987, **45**: 45-52.
- [28] Yildiz, D., Ercal, N., Armstrong, D.W., *Toxicology* 1998, **130**: 155-165.
- [29] Ashakumary, L., Vijayammal, P.L., *Pharmacology* 1996, **52**: 153-158.
- [30] Christofidou-Solomidou, M., Scherpereel, A., Solomides, C.C., Muzykantov, V.R., Machtay, M., Albelda, S.M., DiNubile, M.J., *Lung* 2002, **180**: 91-104.
- [31] Zhou, J.F., Yan, X.F., Guo, F.Z., Sun, N.Y., Qian, Z.J., Ding, D.Y., *Biom. Envir. Sciences* 2000, **13**: 44-55.
- [32] Abou-Seif, M.A., *J. Biochem. Toxicol.* 1996, **11**: 133-138.
- [33] Hulea, S.A., Olinescu, R., Nita, S., Crocnan, D., Kummerow, F.A., *J. Envir. Path. Toxicol. Oncology* 1995, **14**: 173-180.
- [34] Mellon, R.D., Bayer, B.M., *J. Pharm. Exper. Therapeutics* 1999, **288**: 635-642.
- [35] Newman, M.B., Arendash, G.W., Shytle, R.D., Bickford, P.C., Tighe, T., Sanberg, P.R., *Life Sciences* 2002, **71**: 2807-2820.
- [36] Guan, Z.Z., Yu, W.F., Nordberg, A., *Neurochem. Intern.* 2003, **43**: 243-249.
- [37] Liu, Q., Zhao, B., *Br. J. Pharm.* 2004, **141**: 746-754.
- [38] Feger, B., Spratt, C., Earl, C. D., Teismann, P., Oertel, W.H., Kuschinsky, K., *Naunyn-Schmiedeberg's Arch. Pharm.* 1998, **358**: 351-359.
- [39] Piubelli, C., Cecconi, D., Astner, H., Caldara, F., Tessari, M., Carboni, L., Righetti, P.G., Domenici, E., *Proteomics* 2005 (in press).

CHAPTER 7:

PSYCHIATRIC DISORDERS: PROTEOMIC ANALYSIS OF FLUOXETINE- INDUCED MOLECULAR CHANGES IN RAT CORTICAL NEURONAL CULTURES

7.1 DEPRESSION AND ANTIDEPRESSANTS

Depression is a complex, heterogeneous disorder and several neurotransmitter and neurohormonal pathways have been implicated in the pathophysiology of depression [1]. The mechanisms underlying the pathogenesis of depression are not well understood. The serendipitous discovery of antidepressant treatments in the 1950s provided the first evidence of an inherent biochemical abnormality underlying the disorder. In the following 30-40 years research efforts concentrated on studying the mechanisms underlying antidepressant action thus attempting to gain insight into the dysfunction that originate in depression. These studies have led to the establishment of several theories not only for the mechanism of action of antidepressant drugs but also for the pathophysiology of depression. Amongst these has been the monoamine theory of depression. Both preclinical and clinical studies have clearly implicated the serotonin and norepinephrine neurotransmitter systems in antidepressant action. Although these studies have fuelled research efforts in the field and guided the development of novel therapeutic agents, they have not been able to provide a clear model for the pathogenesis of depression. The monoaminergic hypothesis does not provide an adequate explanation for the number of patients who do not respond to current therapeutic agents and also does not solve the puzzle of the lag period in the therapeutic actions of antidepressants. Observations that monoamine depletion does not produce depressive symptoms in healthy individuals and that rapid elevation in monoamines is not correlated with quick antidepressant action has led to the need of revising the framework of existing theories [2, 3].

An emerging hypothesis suggests that the pathogenesis and treatment of depression is likely to involve a plasticity of neuronal pathways [4, 5]. Depression may arise when neuronal systems do not exhibit appropriate, adaptive plasticity in response to external stimuli such as stress. The dysfunction of adaptive pathways that control neuronal plasticity could contribute to the pathogenesis of

depression. Antidepressant treatments may exert their therapeutic effects by either reversing this dysfunction or by independently stimulating an adaptive neuronal plasticity within the system.

Antidepressants such as selective serotonin reuptake inhibitors (SSRIs) are usually the first line of therapy, although the mechanisms by which they exert their therapeutic effects are poorly understood [6].

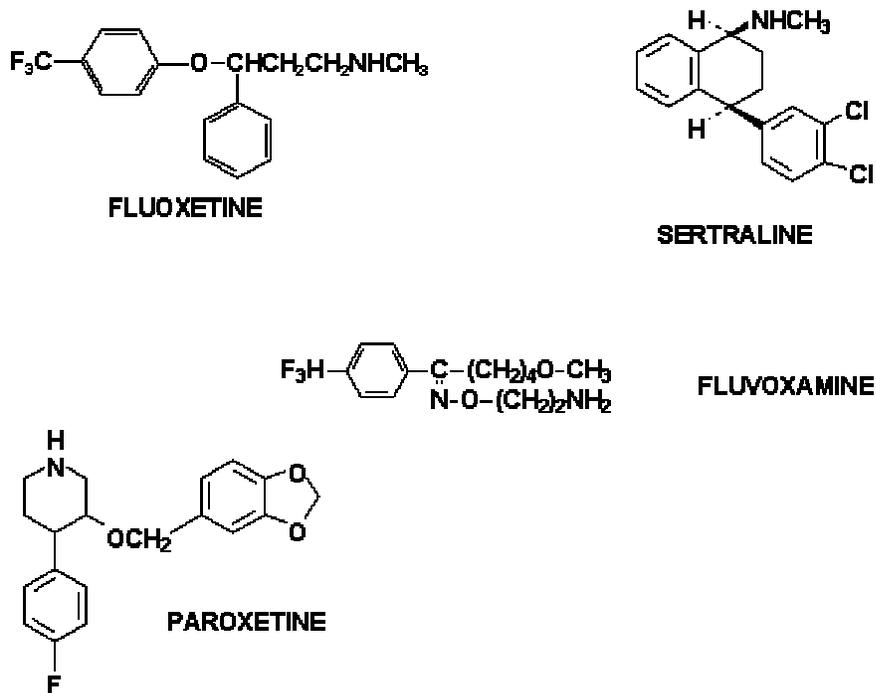


Figure 1. Selective Serotonin Reuptake Inhibitor (SSRI) Antidepressants.

For example, it is not known why several weeks of antidepressant treatment are required before the onset of clinical efficacy [7]. Recent theories to explain this lag include the hypothesis that antidepressants act to restore neurogenesis and presumably alter synaptic connectivity in the hippocampus [8]. However, because the majority of clinically used antidepressants are monoamine uptake blockers, it is difficult to determine whether changes in brain function after chronic administration of these drugs reflect their psychotherapeutic properties, or simply result from targeting specific neurotransmitter systems [9].

In the last 10 years, among selective serotonin reuptake inhibitors (SSRIs), fluoxetine (Fig. 1) has become a first-line drug for the treatment of mood disorders [10] and has expanded considerably its therapeutic potential which, up to the present time, encompasses antidepressive [11], antiobsessive

[12], analgesic [13], anticonvulsant [14], antiphobic [15], antidysphoric [16], anorectic [17] and antibulimic [18], antiepileptic [19], anti-alcohol [20] and cocaine [21] craving effects; fluoxetine has proven to be beneficial also for patients suffering from autism [22] or recovering from stroke [23]. In preclinical pharmacology, fluoxetine is regarded as a selective serotonin 5-HT reuptake inhibitor SSRI because it blocks 5-HT but not catecholamine reuptake and is devoid of significant affinity for a variety of neuroreceptors including serotonergic 5-HT_{1A-2}, β -adrenergic₁₋₂, dopaminergic D₂ and histaminergic H₁ [11].

However, abnormalities in the 5-HT neurotransmitter systems have been convincingly associated with only some of the various disease states that benefit from fluoxetine treatment such as mood [24-26] and obsessive-compulsive disorders [12]. Further, in order to produce therapeutic activities, fluoxetine has to be administered for weeks, which suggests the intervention of common adaptive changes.

In order to explore the effects of sub-chronic administration of the fluoxetine and to improve the understanding of the mechanism of action of this antidepressant drug, we have carried out comparative proteomic analysis of cerebral cortical neuron cells from primary cultures (obtained from rats) treated with the SSRI fluoxetine or vehicle. We have used 2D-PAGE comparative proteomics analysis and identified differentially expressed proteins by MALDI-TOF mass spectrometry. We show that this antidepressant alters the expression of marker proteins involved in axonal transport, synaptic vesicle assembly and in neuroprotection and may contribute to the mechanisms mediating long-term effects of this antidepressant drug.

7.2 EXPERIMENTAL PROCEDURES

The rat cortical neuron primary cultures and immunocytochemistry analysis were performed by the Department of Molecular Medicine Staff, at the Psychiatry Centre of Excellence for Drug Discovery, GlaxoSmithKline.

7.2.1 RAT CORTICAL NEURON PRIMARY CULTURE

Rat embryonic cortical neurons were obtained from timed pregnant Sprague-Dawley rats at 18 days of gestation using a method similar to the one described in Xie et al. [27]. Cerebral cortices were dissected under a stereomicroscope and kept in ice-cold Hank's Balanced Salt Solution (HBSS) without calcium and magnesium. A single cell suspension of cortical neurons was obtained by both enzymatic and mechanical dissociation of cortices. Briefly, cortices were incubated for 10 min at

37°C in a solution containing 0.1% trypsin in HBSS, then cells were dissociated by trituration through the narrowed bore of a fired-polished Pasteur pipette. Cortical neurons were plated in 140 cm² poly-D-Lysine/Laminin (PDL)-coated Petri dishes at a density of 85.000cells/cm² (for proteomic studies) or in PDL treated 19 mm² diameter glass coverslips at a density of 25.000cells/cm² (for immunocytochemistry analysis). Neuronal cultures were maintained in a humidified atmosphere at 5% CO₂ in Neurobasal Medium+B27 Supplement +Penicillin/Streptomycin (all from Gibco™) and the medium changed no more than once a week (50% liquid replacement).

7.2.2 TREATMENT WITH FLUOXETINE

Cortical neurons at day 7 in vitro (DIV-7) were treated for 3 days with 1µM fluoxetine or vehicle. At DIV-10 neurons were scraped from substrate and centrifuged (at 2000 rpm for 4 min at 4°C), pellets were added with 1 ml TBS (Tris- buffered saline) plus protease inhibitor cocktail (Complete Mini, Roche). The cells were then pulse-sonicated on ice two times for 10 sec each then frozen at – 80°C for proteomics analysis, whereas cover slips were processed for immunocytochemistry

7.2.3 IMMUNOCYTOCHEMISTRY

Cortical neurons at DIV10 were fixed in PBS + 4% paraformaldehyde/4% sucrose and processed for immunocytochemistry. Briefly cells were washed in Tris Buffer, permeabilized in 0.25% Triton-X100 Tris Buffer, then incubated overnight at 4°C with primary antibodies diluted in Tris Buffer + 0.1% BSA: mouse monoclonal anti neuron specific b-Tubulin 1:2000 (Covance) and rabbit polyclonal anti-Glial Fibrillar Acidic Protein (GFAP) 1: 4000 (Chemicon). The day after the coverslips were washed, incubated for 2 hours at RT with anti-mouse biotinylated secondary antibody, then developed with *Elite*ABC Kit (Horseradish Peroxidase Development Kit; Vector Laboratories), following manufacturer instructions and utilising as chromogenic substrate DAB-Nickel Enhanced (Vector Laboratories). The procedure was repeated with anti-rabbit biotinylated secondary antibodies utilising as chromogenic substrate VectorNovaRed. The resulted staining allows to distinguish black-stained neurons from red-stained glial cells. The ratio neurons/astrocytes was evaluated by microscopy manual count on immunostained coverslips.

7.2.4 TWO-DIMENSIONAL GEL ELECTROPHORESIS

Control and fluoxetine-treated cortical neuron samples (containing about 1 mg/mL total protein) were added with 0.45 mL of a 2-D solubilising solution, containing 7 M urea, 2 M thiourea, 20 mM Tris, 3% CHAPS, 1% carrier ampholytes, protease inhibitor cocktail and 5 mM tributyl phosphine. The cell samples were then pulse-sonicated on ice five times for 30 sec each, and centrifuged (at 10000 rpm for 10 min at 4°C) to remove nucleic acids. Samples were incubated 90 minutes at room temperature with 5 mM tributyl phosphine to reduce protein disulphide bonds. Cysteine thiolic groups were then alkylated by adding 20 mM iodoacetamide and incubating for 90 minutes at room temperature.

The protein solutions were used to rehydrate 17 cm long IPG strips (pH 3-10, BioRad Laboratories) for 8 h at room temperature. Isoelectric focusing was conducted at 20°C for 60 000 Vh using a Protean IEF Cell (Bio-Rad Laboratories, Hercules, CA, USA) with a low initial voltage and then by applying a voltage gradient up to 10000 V with a limiting current of 50 μ A/strip. The IPG strips were equilibrated for 27 min in a solution containing 375 mM Tris-HCl (pH 8.8), 6 M urea, 20% glycerol (Sigma), 2% SDS (Fluka). The IPG strips were then sealed with 0,5% agarose in SDS running buffer [192 mM glycine (Sigma), 15 mM Tris, 0.1% SDS, pH 8.3] at the top of 8-18% acrylamide gels and second dimension electrophoresis was performed in a Protean II XL tanks (BioRad) at 10 °C and 2 mA/gel for 2 hour, 5 mA/gel for 1 hour and 10 mA/gel overnight. Gels were stained overnight with Sypro Ruby dye, the gels images were scanned with VersaDoc (BioRad) and then analyzed by using the PDQuest software package (BioRad). Protein spots which were altered by more than 50% with a statistical significance of $P < 0.05$ (Student's T-test) were accepted as being differentially expressed between the samples under comparison.

7.2.5 MALDI-TOF MASS SPECTROMETRY

The differentially expressed spots were carefully excised from the gel with a razor blade, placed in Eppendorf tubes, and destained by washing three times for 20 min in 50% v/v acetonitrile, 2.5 mM Tris, pH 8.5. The gel pieces were dehydrated at room temperature and covered with 10 μ L of trypsin (0.04 mg/mL) in Tris buffer (2.5 mM, pH 8.5) and left at 37°C overnight. The spots were crushed and peptides were extracted in 15 μ L of 50% acetonitrile, 1% v/v formic acid. The extraction was conducted in an ultrasonic bath for 15 min. The sample was centrifuged at 8000 x g for 2 min, and the supernatant was collected.

The extracted peptides were loaded onto the 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 by using a ToFSpec 2E MALDI-TOF instrument (Micromass, Manchester, UK), operated in reflectron mode, with an accelerating voltage of 20 kV. Peptide masses were searched against SWISS-PROT and NCBI nr databases by utilizing the identification softwares Mascot, ProteinProbe and ProFound respectively.

7.2.5 SEMI-QUANTITATIVE WESTERN BLOT ANALYSIS

Samples were diluted 1:1 with Laemmli's sample buffer, boiled for 3 min and subjected to SDS/PAGE on 13% T acrylamide gels in Tris/glycine/SDS buffer. The electrophoresed proteins were subjected to electrophoretic transfer onto PVDF membranes at 400 mA for 3h at 4°C. The blocking was done in Tris-buffered saline (TBS) with 5% non-fat dry milk and 0.05% Tween-20 overnight at 4°C. The different primary antibodies, at the appropriate dilutions (see Table 1), were bound to the membranes in TBS containing 1% non-fat dry milk and 0.05% Tween-20.

Primary Antibody	Dilution in TBS, 1% non-fat dry milk, 0.05% Tween-20
Goat anti-UCH-L1 (Chemicon)	1:300
Rabbit anti-14-3-3 ζ (Santa Cruz)	1:3000
Goat anti-SF2/ASF (Santa Cruz)	1:200
Rabbit anti-CypA (Upstate)	1:1000
Rabbit anti-GRP78 (Santa Cruz)	1:1000
Rabbit anti-14-3-3 ϵ (Santa Cruz)	1:500
Rabbit anti-H2B (Santa Cruz)	1:300
Rabbit anti-enolase (Santa Cruz)	1:1000

Table 1. The different primary antibodies used to validate the proteomics results, and the correspondent dilution.

Immunoreactive protein was detected by incubating blots with horseradish peroxidase-conjugated secondary antibody for 45 min followed by chemiluminescent substrate for 1 min (Amersham, Uppsala, Sweden) and visualized on photographic film. For purposes of normalization, the blots were also stained with a monoclonal anti- β actin antibody (Sigma) in a dilution of 1:5000.

The intensity of the ECL response was measured by scanning the photographic films and processing the image using Quantity One software Version 4.4 (BioRad, Hercules, CA, USA).

7.3 RESULTS AND DISCUSSION

7.3.1 IMMUNOCYTOCHEMISTRY ANALYSIS

Neurons were fixed and stained with mouse monoclonal anti neuron specific β -Tubulin 1:2000 and rabbit polyclonal anti-Glial Fibrillar Acidic Protein (GFAP) 1:4000 and then were developed with *EliteABC* Kit (Horseradish Peroxidase Development Kit; Vector Laboratories) as described in section 7.2.3. The ratio neurons/astrocytes evaluated by microscopy manual count on immunostained coverslips showed that the cultures of cortical neurons of the two preparations contain minimum percentage of astrocytes (about 20%, see Table 2).

Furthermore, as shown in Table 2, the two preparations obtained were similar (78.32% and 80.36% of neurons for the control samples of the first and second preparation, 77.80% and 78.32% of neurons for the fluoxetine-treated samples of the first and the second preparation respectively) and so suitable for comparative analysis by 2D PAGE electrophoresis.

First Preparation						
	<i>CTRL</i>			<i>Fluoxetine</i>		
	<i>Neurons</i>	<i>Astrocytes</i>	<i>Total</i>	<i>Neurons</i>	<i>Astrocytes</i>	<i>Total</i>
<i>Sum</i>	1066	295	1361	890	254	1144
<i>Percentage</i>	78,32%	21,68%		77,80%	22,20%	
Second Preparation						
	<i>CTRL</i>			<i>Fluoxetine</i>		
	<i>Neurons</i>	<i>Astrocytes</i>	<i>Total</i>	<i>Neurons</i>	<i>Astrocytes</i>	<i>Total</i>
<i>Sum</i>	847	207	1054	907	251	1158
<i>Percentage</i>	80,36%	19,64%		78,32%	21,68%	

Table 2. Percentage of neurons and astrocytes of the two preparations obtained by immunocytochemistry analysis.

7.3.2 DIFFERENTIAL PROTEOMICS ANALYSIS

Five replicate maps were obtained for vehicle and fluoxetine-treated cortical neurons, for both individual preparations, in Fig. 2 is shown a 2-D map of the control sample of cortical neurons primary culture, stained with Sypro Ruby. About 400 polypeptide spots could be revealed in the pH 3-10 interval with this high sensitivity stain.

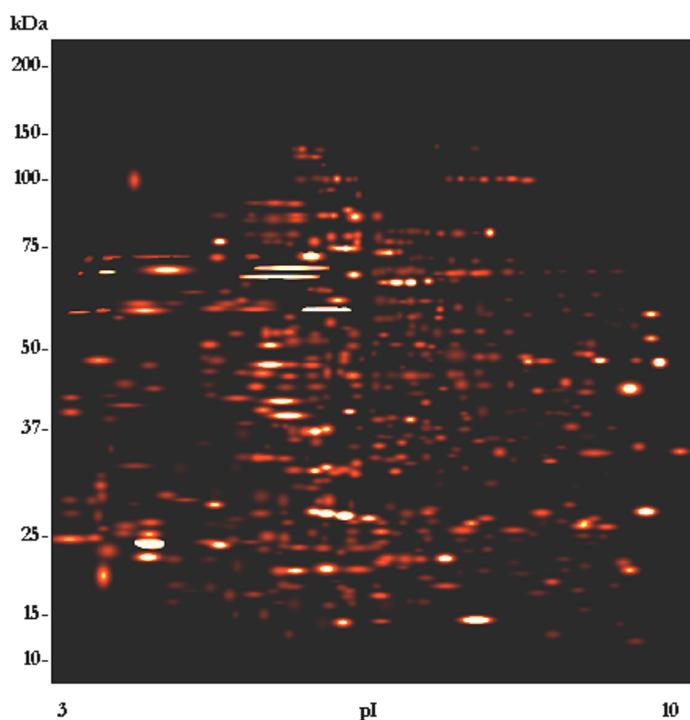


Figure 2. Example of Sypro Ruby-stained 2-D gel of rat cortical neurons primary culture.

After matching the replicated maps of the cortical neurons treated with fluoxetine and vehicle, six proteins turned out to be differently expressed by more than 100% (with $P < 0.05$) while seven proteins were differently expressed by more than 50% (with $P < 0.05$) after treatment with this antidepressant drug. All these spots were eluted and treated as described in section 7.2.4. The identification, together with experimental and theoretically predicted pI and Mr values and the indication of the main role(s) of the protein in cellular pathway and metabolism are given in Table 3.

Spot number	MW [Da]	pI	MOWSE-Score	MASCOT	Z-Score	Accession Number	coverage [%]	No. of peptides	PROTEIN NAME	GENE NAME	variation	Function
2812	72347	5,1	1.23 E12	123	2,31	P06761	33	20	78 kDa glucose-regulated protein precursor	GRP78	up 2	plays a role in facilitating the assembly of multimeric protein complexes inside the ER.
3310	24782	5,2	2.74 E6	69	1,94	Q00981	42	9	Ubiquitin carboxyl-terminal Hydrolase	UCHL1	up 1,8	processing of ubiquitin precursors and of ubiquitinated proteins.
3308	27771	4,8	7.23 E6	88	2	P35215	40	12	14-3-3 Protein zeta/delta	YWHAZ	down 6	activates hydroxylases and protein kinase C, is a regulator of the cell signaling.
7401	27613	10,5	5.07 E5	68	2,11	Q07955	34	9	Splicing factor, arginine/serine-rich	SFRS1	down 2,5	prevents exon skipping, ensuring the accuracy of splicing and regulating alternative splicing.
2105	13775	10,6	5.15 E5	164 mixture	2,35	P02278	67	10	Histone H2B	H2BFG	down 2,5	belongs to the histone H2B family.
8106	17743	8,6	4.16 E7	84	2,34	P10111	64	10	Peptidyl-prolyl cis-trans isomerase A	PPIA	down 2,2	accelerate the folding of proteins; it catalyzes the cis-trans isomerization of proline imidic peptide bonds in oligopeptides.
7205	16605	7,6	4.37 E4	73	2,29	O89086	37	6	Putative RNA-binding protein 3	RBM3	down 1,5	nuclear ribonucleoproteins.
3406	29174	4,7	1.77 E10	150	2,34	P42655	62	18	14-3-3 protein epsilon	YWHAE	down 1,4	activates hydroxylases and protein kinase C, is a regulator of the cell signaling.
6706	46985	6,5	3.88 E11	155	2,34	P04764	47	18	Alpha Enolase	ENO1	down 1,4	belongs to the enolase family.

Table 3. Summary of the 9 identified proteins from rat primary cortical neurons 2D-gels. Matching peptides vs. total number of peptides submitted to database search, sequence coverage and score resulted for peptide fingerprinting matching are listed. Protein Accession Number, theoretical pI and Mr and protein functions were obtained by the SWISS-PROT and NCBI data bases. MASCOT, MOWSE and Z-Scores (output of the identification softwares Mascot, ProteinProbe and ProFound respectively) are measures of the statistical significance of the identification hits.

Among the identified proteins some are of particular interest and may contribute to explain the mechanism of action of fluoxetine.

For example peptidyl prolyl cis-trans isomerase A, also called cyclophilin A (CypA), (down regulated in cortical neurons treated with fluoxetine) is a soluble cytoplasmic immunophilin, that acts as a peptidyl-prolyl isomerase (PPIase). It binds the immunosuppressive drug cyclosporine and the resulting complex blocks T cell function by inhibiting the calcium-dependent phosphatase

calcineurin [28]. Although CyPA has a pivotal role in the immune response, it is most highly concentrated in brain, where its functions are largely unknown [29]. It is classically considered a housekeeping gene, although CypA is a gene with a brain plasticity-related expression in the central nervous system [30]. Furthermore CypA, with other immunophilins, participates in axonal transport, synaptic vesicle assembly and may play a role in neuroprotection against abnormal protein aggregation [31].

Another protein found to be down-regulated in cortical neurons after fluoxetine treatment is 14-3-3 protein zeta/delta. It plays important roles in a wide range of vital regulatory processes, including signal transduction, apoptosis, cell cycle progression and DNA replication [32]. 14-3-3 zeta/delta protein, a brain-specific protein, is an activator of tyrosine and tryptophan hydroxylases, key enzymes for biosynthesis of dopamine and serotonin [33]. It is involved in the regulation of serotonin and catecholamine biosynthesis in neurons and other monoamine-synthesizing cells [34]. The interaction of 14-3-3 zeta/delta protein with Rim1 (a component of the presynaptic active zone and a modulator of exocytosis) may be important in the dynamic function of central nervous system neurons [35].

Also the alternative splicing factor/splicing factor 2 (ASF/SF2) is down regulated after fluoxetine treatment of cortical neurons primary cultures. ASF/SF2 is a member of the family of SR proteins so called because of their C-terminal arginine- and serine-rich domains (RS domains). SR proteins are critical components of the spliceosome, influencing both constitutive and alternative splicing of pre-mRNA [36]. Pre-mRNA processing is an important mechanism for globally modifying cellular protein composition and can be regulated in a highly cell- and tissue-specific or developmentally specific manner. In neurons, the functions of many gene products, such as those of *trk* genes, are regulated by alternative splicing [37]. The *trk* family of receptor tyrosine kinases is crucial for neuronal survival in the vertebrate nervous system [38]. Furthermore the brain-derived neurotrophic factor (BDNF) and its receptor, tyrosine kinase receptor B (*trkB*), play a critical role in activity-dependent synaptic plasticity and have been implicated as mediators of hippocampal-dependent learning and memory and in amygdala-dependent fear conditioning [39].

A protein whose expression resulted instead to be upregulated after fluoxetine treatment is glucose-regulated protein (GRP) 78. It is a member of the heat shock protein (hsp) 70 family and also an endoplasmic reticulum (ER) lumen protein whose expression is induced during ER stress. GRP78 is involved in polypeptide translocation across the ER membrane, and acts as an apoptotic regulator by protecting the host cell against ER stress-induced cell death [40]. GRP78 is not only resident in the

ER, but also exists on the cell surface, where it associates with MHC class I. In the absence of MHC class I the cell surface expression of GRP78 is upregulated [41].

In another study it has been also observed that exposure to valproate (VPA), a drug often prescribed as a long-term therapeutic mood stabilizing agent for individuals with bipolar disorders, increased the expression of endoplasmic reticulum stress protein GRP78 in rat cerebral cortical cells [42, 43]. Regulation of ER stress proteins by VPA may prove to be an important mechanism of action of the drug, and the neuroprotective role of GRP78 may also be involved in the pathophysiology of bipolar disorder (manic depression) [44].

As already described above, chaperone proteins were modified by fluoxetine-treatment: the GRP78 resulted up-regulated, while CypA was down-regulated, indicating that the protein folding is affected by antidepressant treatment. These results are consistent with a proteomic study on social defeat stress model where hippocampi of rats under stress conditions displayed down-regulation of GRP78 and up-regulation of cyclophilin A [Piubelli's PhD thesis, personal communication].

7.3.3 WESTERN BLOT ANALYSIS

We used Western blots for validating protein identification and expression patterns from two-dimensional gels by using antibodies specific to the candidate proteins and anti-beta actin antibody for the normalization of the optical density values.

We performed the analysis with three preparations of cortical neurons (fluoxetine-treated and untreated samples), and each Western blot was replicated twice in two different experiments.

In Fig. 3 an example of three different Western blots (for one of the preparations) of the first experiment are shown; on the right the histograms represent the corresponding 2D-PAGE spot intensity (average value between 10 replicated 2D-maps) in control (left bar) and in fluoxetine-treated (right bar) groups.

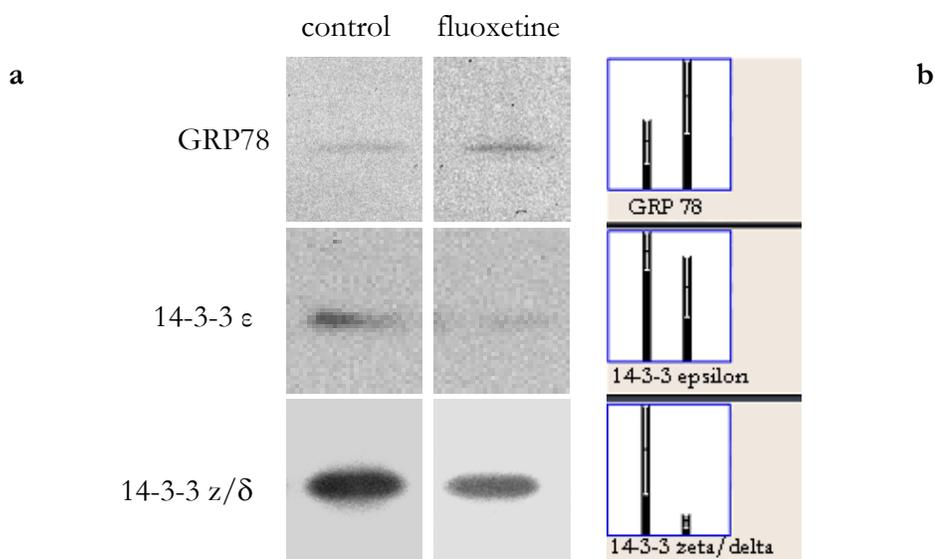


Figure 3. (a) Western blot analyses for GRP78, 14-3-3 ϵ and 14-3-3 z/δ proteins. Results seem to confirm the increased expression of GRP78 and the decreased expression of 14-3-3 ϵ and 14-3-3 z/δ indicated by 2D gel electrophoresis. (b) Gel density measurements show a 2-fold increase in protein for GRP78 and 1.4 and 6-fold decreased for 14-3-3 ϵ and 14-3-3 z/δ proteins respectively. The results represent the average of 5 gels \pm SD. $P < 0.05$.

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 expression of the different isoforms of one protein (separated and identified with 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 replicated 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.

Nevertheless the significantly different expressed proteins that we have identified from proteomics and mass spectrometry analysis may contribute to improve the understanding of fluoxetine mechanism of action.

7.4 CONCLUDING REMARKS

In conclusion in this study a comparison between primary cultures of cortical neurons treated with vehicle or fluoxetine was made. Both preparations were found to be similar for morphological and percentage of neuronal cells. This indicates that the primary culture of cortical neurons is a suitable system for studying the effects of fluoxetine action.

Our findings seem to suggest that the antidepressant fluoxetine alters the expression of marker proteins (involved in neuroprotection, in serotonin biosynthesis, and in the axonal transport) that may contribute to improve the understanding of the fluoxetine psychotherapeutic action and the mechanisms mediating the long-term effects of this antidepressant drug.

The results obtained suggest also new targets for the development of new therapeutic strategies for the treatment of depression.

7.5 BIBLIOGRAPHY

- [1] Vaidya, V.A., Duman, R.S., *British Medical Bulletin* 2001, **57**: 61–79.
- [2] Schildkraut, J.J., *Am J Psychiatry* 1965; **122**: 509–22.
- [3] Heninger, G.R., Delgado, P.L., Charney, D.S., *Pharmacopsychiatry* 1996; **29**: 2–11.
- [4] Duman, R.S., Malberg, J., Thome, J., *Biol. Psychiatry* 1999; **46**: 1181–91.
- [5] Duman, R.S., Heninger, G.R., Nestler, E.J., *Arch Gen Psychiatry* 1997; **54**: 597–606.
- [6] Vaswani, M., Linda, F.K., Ramesh, S., *Prog. Neuro-psychopharmacol. Biol. Psychiatry* 2003, **1**:154-219.
- [7] Blier, P., *Eur. Neuropsychopharmacol.* 2003, **13**: 57–66.
- [8] Duman, R.S., Malberg, J., Nakagawa, S., D'Sa, C., *Biol. Psychiatry* 2000, **48**: 732–739.
- [9] Guesta, P.C., Knowlesa, M.R., Molon-Noblotb, S., Salima, K., Smitha, D., Murraya, F., Laroqueb, P., Huntc, S.P., de Feliped, C., Rupniaka, N.M., McAllistera, G., *Brain Research* 2004, **1002**: 1– 10.
- [10] Freo, U., Ori, C., Dam, M., Merico, A., Pizzolato, G., *Brain Research* 2000, **854**: 35–41.
- [11] Tulloch, I.F., *Eur. J. Psychopharmacol.* 1993, **8**: 5–8.
- [12] Baxter, L., *J. Clin. Psychiatry* 1994, **55S**: 54–59.
- [13] Gatch, M.B., Negus, S.S., Mello, N.K., *Psychopharmacology* 1998, **1**: 99–106.
- [14] Frey, J., Darbonne, C., *Neurology* 1994, **44**: 707–709.
- [15] Abene, M.V., Hamilton, J.D., *Anxiety Disord.* 1998, **12**: 599–603.
- [16] Steiner, M., Korzekwa, M., Lamont, J., Wilkins, A., *Psychopharmacol. Bull.* 1997, **33**: 771–774.
- [17] Caccia, S., Anelli, M., Codegoni, A.M., Fracasso, C., Garattini, S., *Br. J. Pharmacol.* 1993, **110**: 355–359.
- [18] Kornreich, C., Dan, B., Verbanck, P., Fontaine, E., Pelc, I., *Biol. Psychiatry* 1998, **43**: 310–311.
- [19] Favale, E., Rubino, V., Mainardi, P., Lunardi, G., Albano, C., *Neurology* 1995, **45**: 1926–1927.
- [20] Lejoyeux, M., *Alcohol Alcohol.* 1996, **1**: 69–75.
- [21] Batki, S.L., Washburn, A.M., Delucchi, K., Jones, R.T., *Drug Alcohol Depend.* 1996, **41**: 137–142.
- [22] DeLong, G.R., Teague, L.A., McSwain Kamran, M., *Dev. Med. Child Neurol.* 1998, **40**: 551–562.
- [23] Dam, M., Tonin, P., De Boni, A., Pizzolato, G., Casson, S., Ermani, M., Freo, U., Piron, L., Battistin, L., *Stroke* 1996, **27**: 1211–1214.
- [24] Delgado, P.L., Price, L.H., Miller, H.L., Salomon, R.M., Aghaja-nian, G.K., Heninger, G.R., Charney, D.S., *Arch. Gen. Psychiatry* 1994, **51**: 865–874.

- [25] Maes, M., Meltzer, H.Y., The serotonin hypothesis of major depression, in: F.E. Bloom, D.J. Kupfer (Eds.), *Psychopharmacology. The Fourth Generation of Progress*, Raven Press, New York, USA, pp. 933–944.
- [26] Mann, J.J., Malone, K.M., Diehl, D.J., Perel, J., Cooper, T.B., Mintun, M., *Am. J. Psychiatry* 1996, **153**: 174–182.
- [27] Xie, C., Markesbery, W.R., Lovell, M.A., *Free Radical Biology & Medicine* 2000, **28**: 665-672.
- [28] Colgan, J., Asmal, M., Neagu, M., Yu, B., Schneidkraut, J., Lee, Y., Sokolskaja, E., Andreotti, A., Luban, J., *Immunity* 2004, **21**: 189-201.
- [29] Nahreini, P., Hovland, A.R., Kumar, B., Andreatta, C., Edwards-Prasad, J., Prasad, K.N., *Cell Mol. Neurobiol.* 2001, **21**: 65-79.
- [30] Arckens, L., Van der Gucht, E., Van den Bergh, G., Massie, A., Leysen, I., Vandenbussche, E., Eysel, U.T., Huybrechts, R., Vandesande, F., *Eur. J. Neurosci.* 2003, **18**: 61-75.
- [31] Avramut, M., Achim, C.L., *Curr. Top. Med. Chem.* 2003, **3**: 1376-82.
- [32] Qi, W., Liu, X., Qiao, D., Martinez, J.D., *Int. J. Cancer.* 2005, **113**: 359-63.
- [33] Muratake, T., Hayashi, S., Ichimura, Y., Morii, K., Kuwano, R., Ichikawa, T., Kumanishi, T., Isobe, T., Watanabe, M., Kondo, H., *et al.*, *Mol. Neurobiol.* 1995, **11**: 223-30.
- [34] Isobe, T., Ichimura, T., Okuyama, T., *Arch. Histol. Cytol.* 1989, **52**:25-32.
- [35] Sun, L., Bittner, M.A., Holz, R.W., *J. Biol. Chem.* 2003, **278**: 38301-9.
- [36] Aubol, B.E., Chakrabarti, S., Ngo, J., Shaffer, J., Nolen, B., Fu, X.D., Ghosh, G., Adams, J.A., *Proc. Natl. Acad. Sci. USA* 2003, **100**: 12601-6.
- [37] Shinozaki, A., Arahata, K., Tsukahara, T., *Int. J. Biochem. Cell Biol.* 1999, **31**:1279-87.
- [38] Beck, G., Munno, D.W., Levy, Z., Dissel, H.M., Van-Minnen, J., Syed, N.I., Fainzilber, M., *J. Neurobiol.* 2004, **60**: 12-20.
- [39] Rattiner, L.M., Davis, M., French, C.T., Ressler, K.J., *J Neurosci.* 2004, **24**: 4796-806.
- [40] Rao, R.V., Peel, A., Logvinova, A., del Rio, G., Hermel, E., Yokota, T., Goldsmith, P.C., Ellerby, L.M., Ellerby, H.M., Bredesen, D.E., *FEBS Lett.* 2002, **514**: 122-8.
- [41] Triantafilou, M., Fradelizi, D., Triantafilou, K., *Hum Immunol.* 2001, **62**: 764-70.
- [42] Kim, A.J., Shi, Y., Austin, R.C., Werstuck, G.H., *J Cell Sci.* 2004, *in press*.
- [43] Wang, J.F., Azzam, J.E., Young, L.T., *Neuroscience* 2003, **116**:485-9.
- [44] Bown, C.D., Wang, J.F., Chen, B., Young, L.T., *Bipolar Disord.* 2002, **4**: 145-51.

PAPERS PUBLISHED DURING THE PhD STUDIES

Quantitative analysis of two-dimensional gel-separated proteins using isotopically marked alkylating agents and matrix-assisted laser desorption/ionization mass spectrometry

Sylvie Gehanne¹, Daniela Cecconi³, Lucia Carboni², Pier Giorgio Righetti³, Enrico Domenici² and Mahmoud Hamdan^{1*}

¹Computational, Analytical and Structural Sciences, Discovery Research, GlaxoSmithKline, Verona, Italy

²Center of Excellence for Drug Discovery in Psychiatry, GlaxoSmithKline, Verona, Italy

³Department of Agricultural and Industrial Biotechnologies, University of Verona, Verona, Italy

Received 27 May 2002; Revised 3 July 2002; Accepted 3 July 2002

We describe a simple approach for the relative quantification of individual proteins within a mixture. The method is based on the differential labelling of the mixtures by use of a commercially available acrylamide and deuterium-labelled [2,3,3'-d₃]-acrylamide to alkylate proteins prior to two-dimensional (2-D) gel electrophoresis. The tryptic digests of the separated proteins were subjected to reflector matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) analysis and the relative peak heights of cysteine-containing peptides were used to quantify their precursor proteins. This approach was tested for the relative quantification of proteins within an artificial mixture of standard proteins and for proteins observed in a 2-D map of rat serum. A good correlation was found between the measured ratios derived from MALDI-TOF data and those theoretically calculated prior to 2-D analysis via known mixing ratios of the two alkylating reagents. The described procedure has proved to be effective for comparative measurements of protein abundances within the investigated mixtures. Copyright © 2002 John Wiley & Sons, Ltd.

With the completion of an increasing number of genomic sequences, there is a growing need for qualitative and quantitative analysis of proteins which would accelerate the development of new diagnostic tools, new therapeutic agents and lead to a better understanding of the molecular logic that governs cell behavior.¹⁻⁵ A number of existing approaches to protein quantification rely on combining protein separations, most commonly by high-resolution two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), with mass spectrometry.⁶⁻⁹ Despite the efficacy of this approach, the last three years have witnessed different strategies based on reactions between the whole protein mixture with chemical agents, proteolysis, and analysis by liquid chromatography/mass spectrometry (LC/MS) to address the need for reliable measurements of differences in individual protein levels within two different samples.

One of these strategies was first described by Gygi *et al.*,¹⁰ the method uses a new class of chemical reagents termed

isotope-coded affinity tags (ICATs) and tandem mass spectrometry (MS/MS). In this method, two protein mixtures representing two different cell states are treated with the isotopically light and deuterated reagents, respectively. The resulting mixtures are then combined, proteolyzed and the ICAT-labelled peptides are separated by affinity chromatography coupled to MS or MS/MS. The relative quantification is determined by the ratio of two sister peptides which have reacted with the light/heavy versions of the ICAT, where the *m/z* values of the singly charged species differ by 8 Da. A similar strategy for quantitative profiling of protein mixtures has recently been reported by Cagney and Emili,¹¹ their approach, termed mass-coded abundance tagging (MCAT), is based on differential guanidination of C-terminal lysine residues of tryptic peptides followed by capillary LC coupled to electrospray tandem mass spectrometric analysis. Peptides terminating with lysine are modified, which adds 42 Da to their mass but does not otherwise alter their biophysical properties during LC/MS analysis. As in the ICAT approach, protein samples derived from two different cell states are subjected to tryptic digestion, alternatively modified with O-methylisourea, combined, and analyzed by full-scan LC/MS, from which the relative abundances of sister peptides are obtained by comparing peak intensities of their reconstructed single-ion chromatograms. Goodlett *et al.*¹² have recently demonstrated the use of per-methyl

*Correspondence to: M. Hamdan, Computational, Analytical and Structural Sciences, Discovery Research GlaxoSmithKline, Verona, Italy.

E-mail: mahmoud.h.hamdan@gsk.com

Contract/grant sponsor: GlaxoSmithKline; Contract/grant number: GEC 556-2000.

Contract/grant sponsor: MURST.

Contract/grant sponsor: The European Community; Contract/grant number: QLG2-CT-2001-01903.

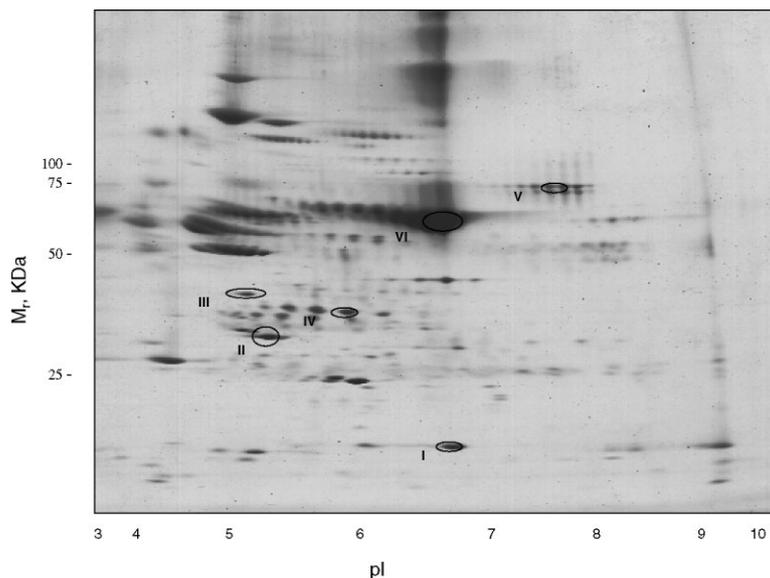
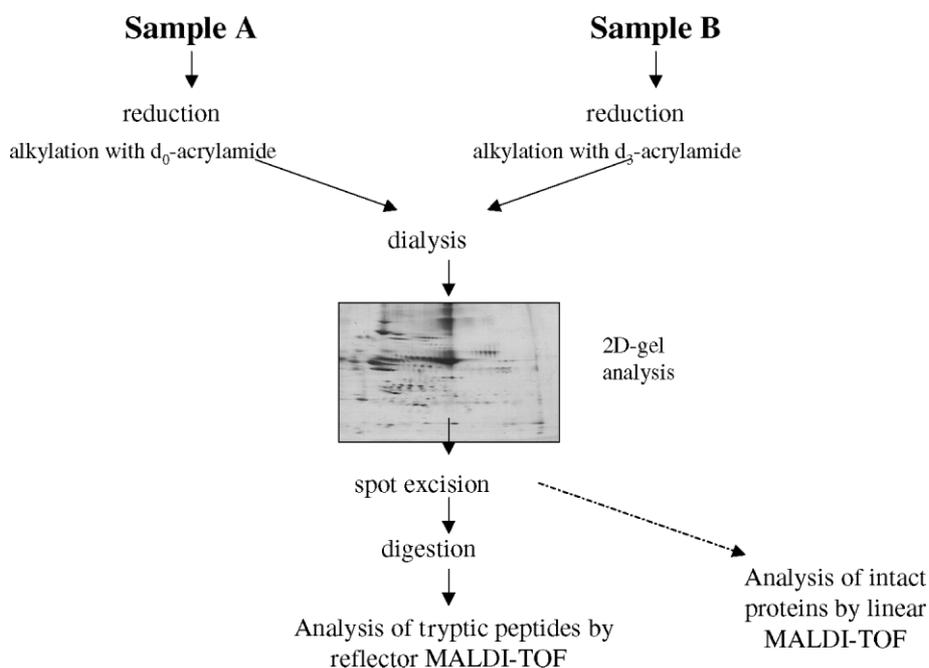


Figure 1. Two-dimensional map in the 3–10 IPG interval of rat plasma composed of two fractions, the first (30%) was alkylated with d_0 -acrylamide, while the second (70%) was alkylated with d_3 -acrylamide. The circled spots refer to: transthyretin (I), apo-lipoprotein (II), apo-lipoprotein A4 (III), α_1 -macroglobulin (IV), sero-transferrin (V) and albumin(VI).

esterification of peptides for relative quantification of proteins. In this approach protein mixtures for comparison are digested and methylated using either d_0 - or d_3 -methanol. The resulting methylated peptide mixtures are then combined and examined by microcapillary LC/MS/MS.

Similar approaches have been reported by Munchbach *et al.*¹³ who described a chemical approach for labelling the N-termini of peptides with an isotopically marked reagent, and by Yao *et al.*¹⁴ who described a proteolysis-catalyzed labelling of peptides at their C-termini using ^{18}O .

This report describes the use of light/heavy acrylamide to alkylate proteins prior to their separation by 2-D gel electrophoresis and their subsequent analysis by MALDI-TOF-MS. It is worth noting that the use of deuterium-labelled acrylamide was used by Sechi and Chait¹⁵ to improve the informational content of MS experiments. The present data demonstrate that the alkylation efficiency is not influenced by the isotopic composition of the acrylamide, which makes the present approach fairly simple and reliable for the relative quantification of proteins derived from two



Scheme 1.

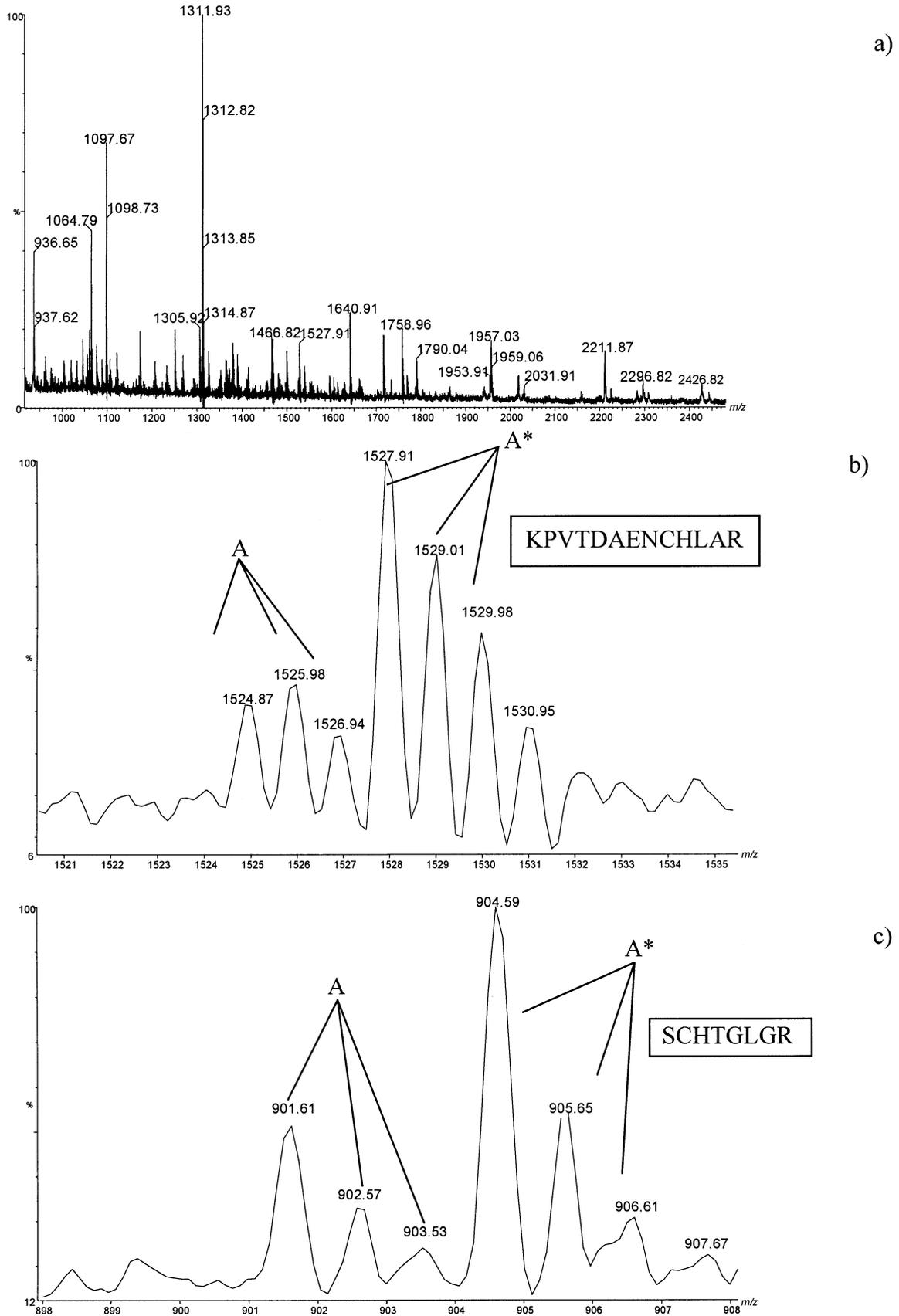


Figure 2. (a) Reflector MALDI mass spectrum of an *in situ* digest of apo-transferrin taken from a 2-D map prepared using a mixture of standard proteins alkylated with d_0 -acrylamide/ d_3 -acrylamide in the ratio 30:70. (b) and (c) are two short intervals taken from the spectrum in (a) and are associated with the two indicated peptide sequences.

Table 1. Peptide sequences identified in the digests of the two listed proteins, together with the theoretical and measured ratios of peptides marked with d_0 - and d_3 -acrylamide

Protein	Acrylamide (d_0/d_3)		Identified sequences
	Theoretical	Measured	
Apo-transferrin	30:70	33:67	DGTRKPVTDAENCHLAR
		34:66	KPVTDAENCHLAR
	50:50	34:66	KNYELLCGDNTR
		35:65	SCHTGGR
		47:53	DGTRKPVTDAENCHLAR
		49.5:50.5	KPVTDAENCHLAR
α_1 Acid glycoprotein	30:70	47:53	SCHTGGR
	50/50	33:67	CEPPLLEKQHEK
		53/47	CEPPLLEKQHEK

different cell states or other experiments to be compared, when treated with the light and heavy forms of this commonly used alkylating agent.

EXPERIMENTAL

Standard proteins mixture

The standard proteins used in this study were: chicken egg lysozyme (accession no. P00698 in Swiss-Prot database), human α_1 acid glycoprotein (accession no. P02763 in Swiss-Prot database) and bovine apo-transferrin (accession no. Q29443 in Swiss-Prot database) purchased from Sigma-Aldrich (St. Louis, MO, USA). Proteins were dissolved (50 μ L) in 40 mM Tris, pH 8.5, and fully unfolded and reduced by treatment with 5 mM tributyl phosphine (TBP), 2 M thiourea and 7 M urea for 90 min at room temperature (RT). The resulting reaction mixture was divided into two parts; the first was alkylated with 100 mM d_0 -acrylamide, while the second was alkylated with 100 mM d_3 -acrylamide. After 5 h, the two fractions were mixed in five different ratios, 90:10–10:90 and subjected to overnight dialysis at 4 °C to stop the alkylation and to remove excess reagents and salts.

Rat serum

Rat serum (100 μ L) containing 6 mg total protein were added to 40 mM Tris buffer, pH 8.5, containing 5 mM TBP, 1.5 M thiourea, and 6 M urea and left for 2 h at RT. The resulting mixture was divided in two parts; the first was alkylated with 50 μ mol d_0 -acrylamide, while the second fraction was alkylated with 50 μ mol d_3 -acrylamide. The resulting two fractions were then mixed in the ratios 50:50 and 30:70, dialyzed overnight and subjected to 2-D gel analysis.

2D-PAGE

Immobilized pH gradient (IPG) strips (7 cm long, pH 3–10; Bio-Rad Laboratories, Hercules, CA, USA) were rehydrated for 4 h with 150 μ L of 2-D solubilizing solution containing 450 μ g protein. Isoelectric focusing (IEF) was conducted at 20 °C for 25000 Vh at a maximum of 5000 V using a Protean IEF cell (Bio-Rad). For the second dimension, the IPG strips were equilibrated in a solution of 6 M urea, 2% SDS, 20% glycerol, Tris-HCl pH 8.8 and were subjected to 7–20% T gradient SDS-PAGE. Gels were run at 10 °C and 5 mA/gel

for 1 h, 10 mA/gel for 1 h, and 20 mA/gel for 2 h. Gels were stained overnight with colloidal Coomassie Blue, and destained in 5% acetic acid.

In situ digestion and peptides extraction

The gel bands of interest were carefully excised with a razor blade, placed in Eppendorf tubes, and destained by washing twice with 50% 5 mM Tris/50% acetonitrile solution. The gel pieces were dehydrated by addition of acetonitrile. Excess solvent was removed, followed by drying for 20 min on a SpeedVac centrifuge apparatus. The gels were rehydrated by adding 15–30 μ L of a solution containing 2.5 mM Tris and modified trypsin from Promega. The rehydrated gels were kept for 4 h at 37 °C. Peptides were extracted by adding 30 μ L of a solution containing 50% acetonitrile and 50% water with 1% of formic acid.

MALDI-TOF-MS

The extracted peptides were loaded onto the target plate by mixing 1 μ L of each solution with the same volume of a matrix solution (10 mg/mL α -cyano-4-hydroxycinnamic acid in 50% ethanol and 50% acetonitrile) and left to dry at RT. The MS analyses were performed using a ToFSpec-2E MALDI-TOF instrument (Micromass, Manchester, UK) equipped with a pulsed nitrogen laser (337 nm, pulse width 4 ns) and operated in reflectron mode with an accelerating voltage of 20 kV.

RESULTS

The main steps of the present approach are depicted in Scheme 1. It should be pointed out that the analysis of the intact protein is an optional step that may be used as an early screening for possible gel-induced and/or post-translational modifications.

To probe the capability of the present approach in reflecting the pre-separation labelling, a mixture of three standard proteins, lysozyme, α_1 acid glycoprotein and apo-transferrin, was divided into two parts, the first being alkylated with d_0 -acrylamide, while the second with its deuterated version. The resulting reaction mixtures were then combined in five different ratios (90:10–10:90) and subjected to 2-D separation which yielded five different maps; protein spots were excised, digested and examined by

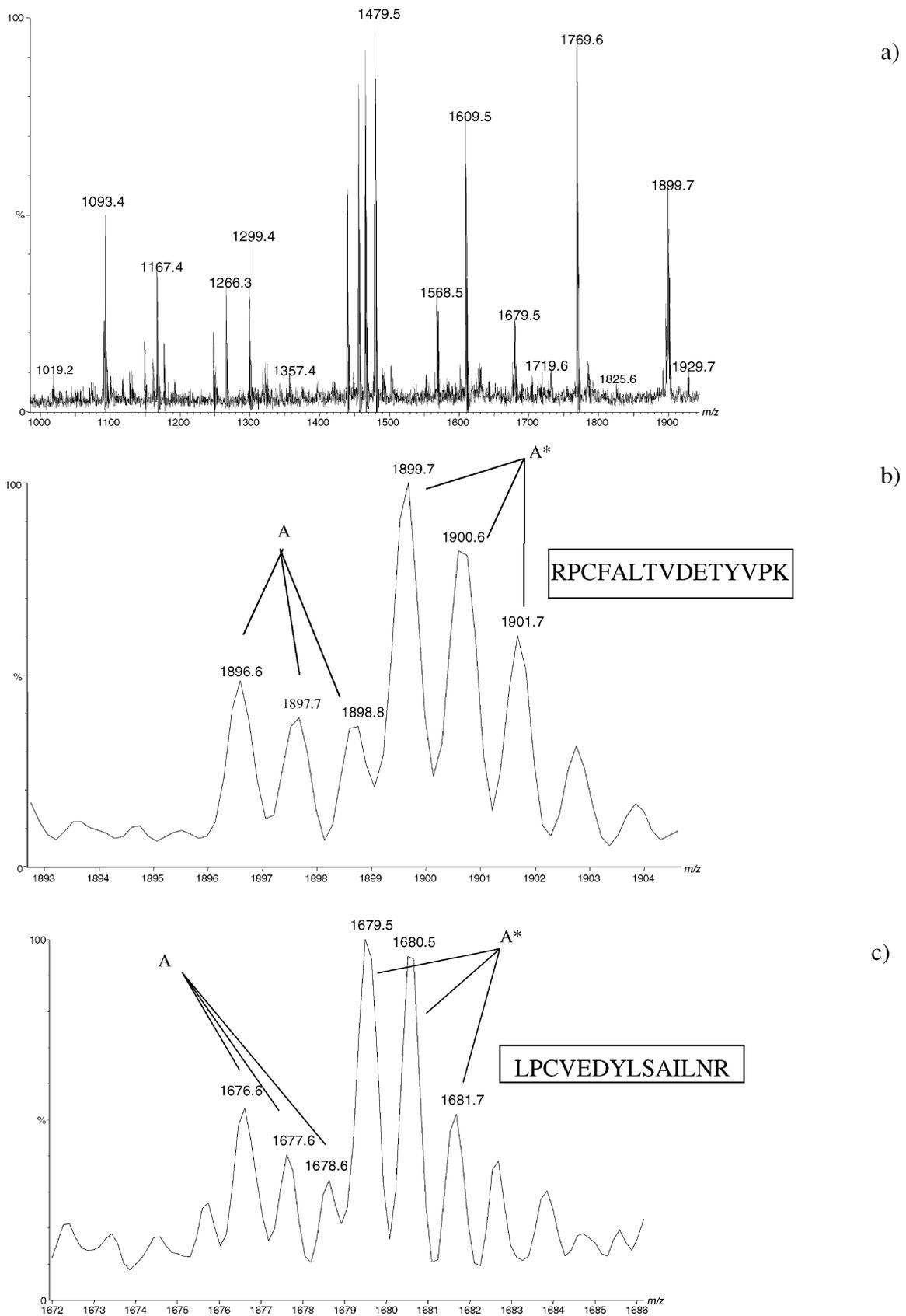


Figure 3. (a) Reflector MALDI mass spectrum of an *in situ* digest of albumin (spot VI) in Fig. 1. (b) and (c) are two short intervals taken from the spectrum in (a) associated with the indicated peptide sequences.

reflector MALDI-TOF. A representative example, pertaining to apo-transferrin, of the latter measurements is given in Figs 2(a)–(c). Figure 2(a) gives the MALDI spectrum of the entire digest, while Figs 2(b) and 2(c) give two short intervals taken from the same spectrum. These intervals refer to the indicated peptide sequences each of which has a single cysteine residue. In each of the two spectra there are two isotopic distributions marked A and A* for which a difference of 3 Th in the m/z values of the corresponding peaks within the two distributions is clearly evident.

This observation can be easily interpreted by the following considerations. It is well documented^{16–18} that the free-SH group in the side chain of cysteine is the favorite site for interaction with acrylamide and other commonly used alkylating agents. The two isotopic distributions in Figs 2(b) and 2(c) refer to two sister peptides derived from two protein fractions, the one treated with d_0 - and the other one with d_3 -acrylamide. The same spectra show that the relative peak heights associated with (A*) are ~66% compared with ~34% associated with the corresponding peaks in distribution (A), which is in reasonable agreement with the labelling ratio (70:30) effected prior to the electrophoretic separation. Similar spectra were used to construct Table 1, which compares the measured and calculated protein ratios marked with d_0 - and d_3 -acrylamide.

After testing this procedure on a standard mixture, we evaluated the same method on rat serum, which yielded the 2-D map in Fig. 1. The alkylation was achieved using the ratio 30:70 of d_0/d_3 -acrylamide; a number of separated proteins were digested *in situ* and subjected to reflector MALDI-TOF analysis. A representative spectrum pertaining to albumin is given in Fig. 3(a), while short scan intervals taken from the same spectrum are presented in Figs 3(b) and 3(c). The observed isotopic distributions in both spectra are attributed to the indicated sequences, each of which contains a single cysteine residue. Furthermore, the more intense peaks in both spectra are displaced by 3 Th from their weaker counterparts. Such displacement coincides with a single alkylation channel associated with the d_3 -acrylamide. Considering the relative peak heights in both isotopic distributions we have calculated a ratio of 34:66, which is in reasonable agreement with the labelling ratio 30:70 prior to 2-D separation.

DISCUSSION

Cysteine residues are found in a wide variety of storage and structural proteins, enzymes, hormones, receptors, and microbial toxins.¹⁹ Disulfide bonds associated with Cys residues are structural units essential in most proteins and are largely responsible for the conformations and tertiary structures that proteins assume both in solution and in the solid state. Because -SH groups are among the most reactive functionalities in chemical and biochemical systems (e.g., they undergo sulfhydryl-disulfide exchange, oxidation-reduction, nucleophilic addition, displacement conjugation reactions²⁰), it is only natural that a variety of methods have been described to stabilize and/or protect them, in order to facilitate protein analysis. Among the most popular methods reported have been alkylation with iodoacetamide (IAA)

and *N*-ethylmaleimide (NEM). The use of IAA is not ideal because it is difficult to drive the reaction to completion,²¹ moreover, the reaction is not specific for -SH groups, since it has been reported to modify also α -NH₂ groups of N-terminal chains and ϵ -NH₂ groups of Lys. Modification with NEM can result in a multiplicity of labelled peptides owing to the formation of diastereoisomers and hydrolysis of the succinimide ring of the Cys adducts.²²

In the light of the above considerations, it is worth critically evaluating the results presented here. The data in Figs 2 and 3 give an initial indication that isotopic labelling of the acrylamide alkylating agent together with MALDI-TOF-MS can be considered a valid approach for relative quantification of gel-separated proteins. The difference of 3 Th in the measured m/z values of sister peptides was enough to distinguish between them; however, there is no reason why other alkylating agents with a higher number of deuterium atoms cannot be used in a similar fashion.

Having said this it is fair to look closely at this approach compared with existing strategies for identifying the strengths and the weaknesses of each. Both the ICAT and the present approaches assume specific and complete alkylation of all free -SH groups within the investigated sequence(s), an assumption which is not fully supported by recent work.^{16,17,23} These authors have demonstrated that alkylation times as long as 6 h were not sufficient to produce more than 80% alkylation; the efficiency is likely to be less in the case of the ICAT procedure which has to accommodate a biotin group and an ethylene glycol linker group. Not all proteins of interest contain cysteine, so some proteomes will have much lower total protein coverage than the 92% commonly cited for yeast.¹² Some cysteine residues may contain post-translational modifications prior to isolation making them unavailable for alkylation with cysteine-specific ICAT or acrylamide. To address the limitation of *in vitro* labelling of non-cysteine-containing peptides for protein expression analysis, a method complementary to the use of ICAT reagents has been described by Goodlett *et al.*¹² The method uses differential stable isotope esterification of carboxylic acids in peptides to provide an internal standard for relative quantification of peptides in a mixture regardless of their cysteine content. Essentially, protein digests derived from two different biological states are esterified with either d_0 - or d_3 -methanol; the two samples are combined and analyzed by LC/MS/MS.

The MCAT strategy cannot be considered an alternative to its ICAT counterpart. It is more appropriate to describe the two strategies as complementary approaches to protein quantification. There are differences in details, yet the outcome of either approach gives a partial answer to the question of protein quantification in complex mixtures. This statement can be supported by the following considerations: Both strategies are based on the modification of a specific amino acid, Cys in the case of ICAT and C-terminal Lys in the case of MCAT, to allow discrimination between two sister peptides derived from two different biological samples. In the first approach, the alkylation of Cys results in a difference of 8 Da, while, in the second case, guanidination of Lys results in an increase of 42 Da. The use of MCAT in conjunction with ICAT would overcome two main

drawbacks of the latter approach, i.e., those associated with analysis of Cys-free proteins and the detection of peptides that happen to contain post-translationally modified Cys.

CONCLUSIONS

Preliminary data on the use of 2-D gel electrophoresis, combined with MALDI-TOFMS of light and heavy acrylamide-labelled polypeptides, as a tool for relative quantification of proteins derived from standard protein mixtures and from rat serum have been reported. The present method is not an alternative to existing strategies that rely on LC/MS/MS; however, it offers a number of advantages which can be exploited in the field of protein quantification. (1) This method uses relatively cheap and commercially available reagents that are commonly exploited in the solubilization cocktail prior to any electrophoretic step, meaning that no additional steps are needed in the phase of sample preparation. (2) The use of 2-D gel separations in the present method requires more time compared with LC/MS of an entire digest; on the other hand, the use of high-resolution separation protocols offers distinct advantages, i.e., it gives a more comprehensive picture of the protein content of the sample, simpler mass spectral data to interpret, and the possibility to analyze intact proteins based on which a number of protein modifications can be recognized at an early stage of analysis. The ability to elute and analyze intact proteins from a 2-D map is, in fact, a unique advantage for those scientists seeking to attribute a post-synthetic modification to a specific protein zone. Such a possibility is in reality lost (or greatly hampered) in fractionation strategies based on massive protein digestion prior to the separation process, such as those adopted in the ICAT and other protocols.

While this paper was in the late stages of the review process, a very similar approach was reported by Sechi.²⁴ Although the present approach is basically similar to that of Sechi, some information obtained in the present work is complementary to his work. The fact that two groups working independently have reached the same general conclusions concerning the suitability of alkylation of Cys by acrylamide for quantitative comparisons of protein levels indicates that this is likely to be a useful approach.

Acknowledgements

This work was supported in part by a GlaxoSmithKline grant (No. GEC 556-2000), by MURST (Coordinated Project 40%, Proteome analysis, 2000) and by the European Community (Grant No. QLG2-CT-2001-01903) to PGR.

REFERENCES

1. Matthias M, Hendrickson RC, Pandey A. *Annu. Rev. Biochem.* 2001; **70**: 437.
2. Blackstock WP, Weir MP. *Trends Biotech.* 1999; **17**: 121.
3. Griffin TJ, Han DKM, Gygi SP, Rist B, Lee H, Aebersold R, Parker CP. *J. Am. Soc. Mass Spectrom.* 2001; **12**: 1238.
4. Godovac-Zimmermann J, Brown LR. *Mass Spectrom. Rev.* 2001; **20**: 1.
5. Washburn MP, Walters D, Yates JR. *Nat. Biotechnol.* 2001; **19**: 242.
6. Herbert B, Galvani M, Hamdan M, Olivieri E, MacCarthy J, Pedersen S, Righetti PG. *Electrophoresis* 2001; **22**: 2046.
7. Chang H-T, Yergey AL, Chrambach A. *Electrophoresis* 2001; **22**: 394.
8. Larsen MR, Larsen PM, Fey SJ, Roepstorff P. *Electrophoresis* 2001; **22**: 566.
9. Galvani M, Hamdan M, Righetti PG. *Rapid Commun. Mass Spectrom.* 2001; **15**: 258.
10. Gygi SV, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. *Nat. Biotechnol.* 1999; **17**: 994.
11. Cagney G, Emili A. *Nat. Biotechnol.* 2002; **20**: 163.
12. Goodlett DR, Keller A, Watts JD, Newitt R, Yi EC, Purvine S, Eng JK, Von Haller P, Aebersold R, Kolker E. *Rapid Commun. Mass Spectrom.* 2001; **15**: 1214.
13. Munchbach M, Quadroni M, Miotto G, James P. *Anal. Chem.* 2000; **72**: 4047.
14. Yao X, Freas A, Ramirez J, Demirev PA, Fenselau C. *Anal. Chem.* 2001; **73**: 2836.
15. Sechi S, Chait BT. *Anal. Chem.* 1998; **70**: 5150.
16. Galvani M, Rovatti L, Hamdan M, Herbert B, Righetti PG. *Electrophoresis* 2002; **22**: 2066.
17. Hamdan M, Bordini E, Galvani M, Righetti PG. *Electrophoresis* 2001; **22**: 1633.
18. Galvani M, Hamdan M, Herbert B, Righetti PG. *Electrophoresis* 2001; **22**: 2058.
19. Thannhauser TW, Sherwood RW, Scheraga HA. *J. Protein Chem.* 1998; **17**: 37.
20. Friedman M. *J. Agric. Food Chem.* 1994; **42**: 3.
21. Came AF, Sally U. *Methods Mol. Biol.* 1997; **64**: 271.
22. Bures EJ, Hui JO, Young Y, Chow DT, Katta V, Rohde MF, Zeni L, Rosenfeld RD, Stark KL, Haniu M. *Biochemistry* 1998; **37**: 12172.
23. Hamdan M, Galvani M, Righetti PG. *Mass Spectrom. Rev.* 2001; **20**: 121.
24. Sechi S. *Rapid Commun. Mass Spectrom.* 2002; **16**: 1416.

Emilio Marengo¹
Elisa Robotti¹
Valentina Gianotti¹
Pier Giorgio Righetti²
Daniela Cecconi²
Enrico Domenici³

¹Dipartimento di Scienze e Tecnologie Avanzate, Università del Piemonte Orientale, Alessandria, Italy

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

³Molecular Medicine, Psychiatry Centre of Excellence for Drug Discovery, GlaxoSmithKline Medicines Research Centre, Verona, Italy

A new integrated statistical approach to the diagnostic use of two-dimensional maps

Two-dimensional (2-D) electrophoresis is a very useful technique for the analysis of proteins in biological tissues. The complexity of the 2-D maps obtained causes many difficulties in the comparison of different samples. A new method is proposed for comparing different 2-D maps, based on five steps: (i) the digitalisation of the image; (ii) the transformation of the digitalised map in a fuzzy entity, in order to consider the variability of the 2-D electrophoretic separation; (iii) the calculation of a similarity index for each pair of maps; (iv) the analysis by multidimensional scaling of the previously obtained similarity matrix; (v) the analysis by classification or cluster analysis techniques of the resulting map co-ordinates. The method adopted was first tested on some simulated samples in order to evaluate its sensitivity to small changes in the spots position and size. The optimal setting of the method parameters was also investigated. Finally, the method was successfully applied to a series of real samples corresponding to the electrophoretic bidimensional analysis of sera from normal and nicotine-treated rats. Multidimensional scaling allowed the separation of the two classes of samples without any misclassification.

Keywords: Fuzzy logic / Multidimensional scaling / Proteomics / Two-dimensional maps

EL 5230

1 Introduction

1.1 General aspects

Electrophoretic separations are widely used for identifying proteins in different chemical and biological fields, namely botanic and animal physiology and above all in pharmaceutical and clinical research [1–8]. These last applications are certainly the most important for the recent progresses in the understanding of many human diseases [9, 10], such as cancer [11–13], encephalopathies [14], and Alzheimer's disease [15, 16]. All these diseases can cause changes in the protein content of biological fluids (blood, urine, plasma, serum) and/or in the proteins produced by different types of cells. So it is fundamental to separate and possibly identify the proteins contained in biological samples with efficacy and accuracy. Two-dimensional (2-D) electrophoresis is the most widely used method, in which the first dimension (pH gradient) consists of a separation of the proteins according to their isoelectric points, whereas the second dimension

(SDS-PAGE) is based on molecular mass discrimination. This technique produces a 2-D map that consists of a series of spots (the proteins) spread over the surface of the final SDS-PAGE slab. This method is used to investigate the quantitative expression of different proteins in the samples. The comparison of 2-D maps from control and pathological (or treated) samples in order to identify proteins connected with the disease is however a complex process [17–23]. This complexity arises from the complexity of the maps, which may contain thousands of spots, from the low reproducibility of the 2-D maps, caused by the effect of several experimental factors on the final result (polymerisation conditions, sample pretreatment, 2-D run conditions) [24], and from the small differences which often characterize samples from healthy individuals and diseased patients. These differences often might be due not simply to the absence or presence of some proteins, but to even subtle changes in their relative amounts.

Today, the comparison of the 2-D maps is performed by different methods [25–27], all based on similar approaches. The first step of the analysis consists of the alignment of the maps, following a warping and matching process by shrinking or widening the spots with respect to the two dimensions. This step is based on the choice of some landmark spots defined by the user. Then each map is “cleaned”, by rejecting the spots which do not show a typical optical density pattern. This step allows the elim-

Correspondence: Prof. E. Marengo, Dipartimento di Scienze e Tecnologie Avanzate, Università del Piemonte Orientale, Spalto Marengo 33, I-15100 Alessandria, Italy
E-mail: marengo@tin.it
Fax: +39-(0)-131-287416

Abbreviation: MDS, multidimensional scaling

ination of several spots which do not contain biochemical information, whose presence is caused by the effect of external experimental factors (*e.g.*, staining artefacts). Once the alignment is obtained and the maps have been cleaned, they are matched to one another and the common information is identified. In spite of all the problems connected with their comparison, 2-D maps are still very powerful for proteome analysis. It is evident that it is necessary to work with replicates of the same experiment, in order to take into account the experimental variability.

In the present paper, the comparison of different 2-D maps was approached with a different method: the comparison is performed on each replicate, not on a synthetic map; the spots in the maps are allowed to become fuzzy, with a broadening of the spots themselves, to simulate the experimental uncertainty in their position, shape and size; similarity indexes between the fuzzy maps are calculated; the matrix of similarities thus obtained is used as input in a multidimensional scaling (MDS) calculation; the coordinates of each map obtained from the (MDS) are employed in a pattern recognition analysis which enables the clustering of different classes of samples, as belonging to control and diseased, or treated, tissues. The proposed procedure is first analyzed using simulated maps in order to show its potential applications. Finally, the results of its application to real data are presented.

1.2 Theory

The method proposed for analyzing 2-D maps obtained from different patients requires the availability of some maps of control and diseased (or treated) subjects. The procedure is based on two steps: (i) the calibration of the method, and (ii) its use for the classification of unknown samples. The calibration produces a model that must be evaluated to check its ability to distinguish between the different classes of subjects. If the model performs satisfactorily the classification of the known samples (supervised analysis), it can be used for the assignment of new samples to the existing classes (usually diseased or treated and control patients).

The calibration phase consists of five steps:

1. Digitalization of each map with the creation of a rectangular grid of cells. The grid is matched on the 2-D map and a signal varying from 0 (empty cell) to 1 (black cell) is assigned to each cell, depending on the degree of black of the corresponding area of the 2-D map.
2. Transformation of the maps in fuzzy entities (hereafter dubbed “fuzzification”, with a neologism), *i.e.* a step where the spots are allowed to smear, so that the cells occupied by the spots can influence the signal of the

near cells [28]. In this step, a probability function is matched to each occupied cell of the map, extending on the neighbor cells.

3. The comparison of all the pairs of transformed 2-D maps with the calculation of a similarity matrix where each similarity is obtained from the matching of the corresponding maps.
4. A multidimensional scaling analysis of the similarity matrix.
5. The statistical analysis of the coordinates of the 2-D maps in the virtual space obtained from MDS by cluster analysis and/or classification methods.

1.2.1 Digitalization

A 2-D map can be treated as a 2-D surface corresponding to a grid of a given step. The digitalization of the map is obtained by a grid with a pass of 0.001×0.001 m, that produces a map of 200×200 cells. Each 2-D map is thus transformed into a matrix, corresponding to the grid. A 0 value is assigned to the cell corresponding to empty zones of the 2-D-maps, a unit value to the area occupied by the spots. In the first application the percentage of black area of the cells is not taken into account, only 0 or 1 is used to distinguish between empty and occupied zones. An example of digitalization is presented in Fig. 1.

1.2.2 Fuzzification

The spot fuzzification is very important since, as it has already been remarked, the position, size and shape of the spots on the maps may change in different experimental runs performed on the same biological sample. Moreover, different factors (such as temperature, contact time with development solutions, *etc.*) can contribute to the appearance of spurious spots not directly attributable to the protein content of the original sample.

The influence of the presence of a spot present in cell x_i, y_i on the neighbor cell x_k, y_k is calculated by the following 2-D gaussian function:

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

where σ_x and σ_y are two constants corresponding to the standard deviation of the gaussian function along each of the two dimensions and ρ is the correlation between the two dimensions x and y . The parameter ρ is fixed at 0, corresponding to the expected complete independence of the two electrophoretic runs.

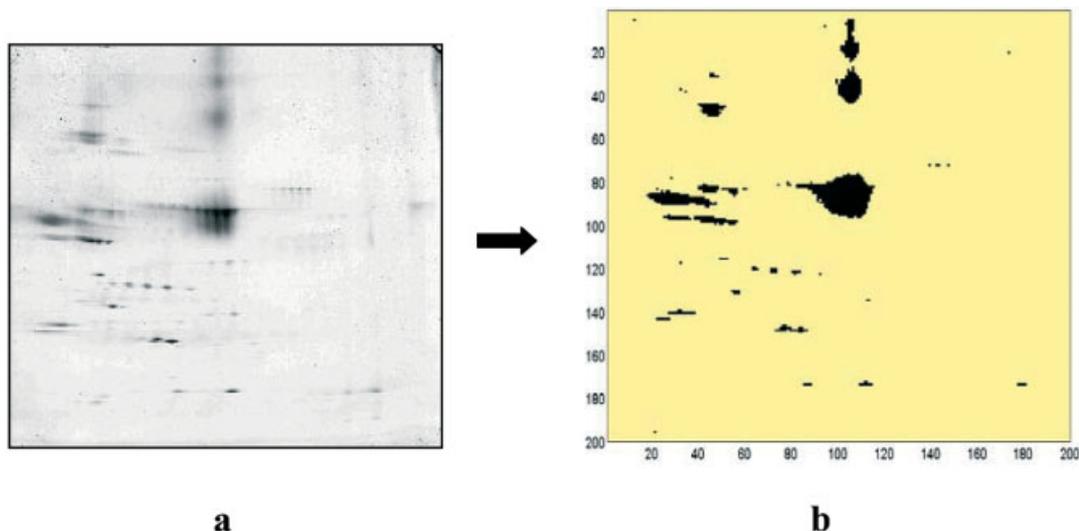


Figure 1. (a) 2-D map of sample HEA5 and (b) the correspondent digitalized image.

In this paper, the two parameters σ_x and σ_y are maintained identical, so that the gaussian function presents the same uncertainty with respect to both dimensions. This is based on the assumption of a similar variability of the two electrophoretic runs. Following this statement, the only parameter which shall be analyzed for its effect on the final result is $\sigma = \sigma_x = \sigma_y$. Obviously, a large value for σ shall extend the effect of each spot at a larger distance, making the whole virtual map more fuzzy. A gaussian function was chosen since the spots can be properly described by a function with the highest intensity in the center of the spot itself and decreasing values for increasing distance from the spot.

The value of the signal S_k in each cell x_i, y_j of the virtual grid is given by the sum of the effect of all neighbor cells containing spots:

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

The summation runs on all cells of the grid, but in dependence on the value of the parameter σ , only the neighbor cells produce a significant effect. Each sample is then transformed in a virtual map containing in each cell the sum of the influence of all the spots of the original map. These virtual maps represent a sort of probability of the presence of a spot in each given cell of the grid, and are used for calculating the similarity matrix between every pair of samples.

The strategy is in principle similar to the traditional one, which makes use of virtual maps obtained from the matching of several experimental 2-D maps. In our case, the virtual map is obtained by allowing a fuzzy change of

the spots, whose position and size thus become blurred, according to a gaussian dependence. In principle this method eliminates operator intervention to match spots or clean the 2-D maps.

1.2.3 Similarity index

The similarity between each pair of samples was calculated after performing a match of the potential matrices. The match of the two virtual grids k and l allows the computation of the common signal (SC_{kl}), namely the sum of all the signals present in both maps, and the total signal (ST_{kl}):

$$SC_{kl} = \sum_{i=1,n} \min(S_i^k, S_i^l) \quad (3)$$

$$ST_{kl} = \sum_{i=1,n} \max(S_i^k, S_i^l) \quad (4)$$

The similarity index can then be computed as the ratio between SC_{kl} and ST_{kl} :

$$S_{kl} = \frac{SC_{kl}}{ST_{kl}} \quad (5)$$

A similarity equal to 1 indicates that the two virtual maps are identical, since the common and total signals are identical. The lower the similarity index, the greater the difference between common and total signals. A value of 0 indicates that there is no match between the signals of the two maps, which corresponds to the minimum possible similarity. The similarity matrices were calculated by an algorithm programmed in Visual Basic 6.0 (Microsoft) and Matlab 6.1 (The Mathworks Inc.).

1.2.4 MDS and statistical analysis

MDS is a branch of multivariate data analysis geared towards dimensional reduction and graphical representation of data. Given a set of n objects and a measure of their similarity S_{ij} , MDS consists in the search for a low dimensional space in which the objects are represented by points in the space and such that the distances between the points match as much as possible with the original similarities of the objects [29]. The space is usually Euclidean, but this is not a constraint. There are several different approaches to MDS depending on the notion of matching of the similarities, on the metrics, on the method used to compute the similarities and on the way the point configuration is obtained [30].

The method for finding the original Euclidean coordinates from the derived Euclidean distances was first given by Shoenberg [31] and Young and Householder [32], but it was Gower [33] who provided a clear formulation and stated the importance of the classical scaling technique with the introduction of the principal coordinate analysis (PCO). Shepard [34, 35] and Kruskal [36, 37] provided an extension of classical MDS to the study of nonparametric similarities introducing iterative algorithms for the search of the best solution. The calculation of MDS from the 2-D map similarities was performed by using the Kruskal iterative method. The configuration of points obtained at convergence of the iterative process, *i.e.*, when changes of the points coordinates do not introduce any further decrements of the overall matching between real similarities and calculated distances, is considered as the final solution. The search for the coordinates is based on the steepest descent minimization algorithm, where the target function is the so-called stress (S), which is proportional to the sum of squares of differences between calculated and real distances, *i.e.*, a measure of the ability of the configuration of points to simulate the distance matrix. MDS calculations were performed by the STATISTICA software (Statsoft Inc., Ver. 5.1). Once the coordinates have been obtained, they can be used for a multivariate statistical analysis. The results are often self-evident, so that a visual inspection is sufficient for obtaining the final results.

2 Materials and methods

The method proposed has been applied on simulated and real 2-D maps in order to investigate its ability to discriminate between different samples and its response to several experimental issues (for example, changes of size, position and number of spots). The effect of parameter σ was investigated at five levels: 0.50, 0.75, 1.00, 1.25, 1.50.

2.1 Simulated samples

The simulated 2-D maps were used for evaluating the effect on the similarity indexes of the following parameters: changes in size and shape of a single spot (10 samples, Fig. 2a, where DIM = dimension); changes in position of a single spot (10 samples, Fig. 2b, where POS = position); the effect of two spots, with the largest that changes its size and shape (12 samples, Fig. 2c, where DIM = dimension, plus two reference spots); the effect of two spots, with the smallest that changes its position (12 samples, Fig. 2d, where POS = position, plus two reference spots). This set of 44 images was used to obtain the corresponding fuzzy maps for (40×40) grids. In this case we used (40×40) grids instead of the 200×200 grids adopted for the real samples, in order to decrease the computational effort.

2.2 Real samples

The 2-D maps used to test the computational method were divided into two groups: (i) five serum samples from five different control Wistar rats (untreated or healthy, HEA1 to HEA5); (ii) five serum samples from five different Wistar rats treated with nicotine (ILL1 to ILL 5). Figure 3 shows the ten 2-D maps. As it can be noticed, it is very difficult to distinguish the control animals from the ones treated with nicotine by a visual inspection of the image. Moreover, a large variability of the spots position, shape and size is evident along the 2-D maps belonging to the same group, probably enhanced by differences in the amount of loaded material. Looking at Fig. 3, it appears clear that sample HEA2 is shifted with respect to the other four healthy samples; in order to correct this problem, sample HEA2 was shifted and matched to the other healthy samples. This correction does not consist in a data manipulation but it is simply a shift of the image, without any distortion of the image itself, preserving the original distances between the spots. The 2-D maps presented in Fig. 3 were obtained with a method developed in six steps:

(i) Rat treatment: five Wistar rats were treated for 14 days with a saline solution (control samples) and the other five were treated for the same 14 days with nicotine. The nicotine was administered subcutaneously by injecting 1 mL/kg of a 0.4 mg/mL nicotine solution.

(ii) Sample treatment: blood samples were collected on the 14th day (when it is known that nicotine administration begins to induce dependence on treated rats) on rats which were fasted for 12 h prior to collection in order to avoid interferences due to high concentrations of lipids in the blood. All samples were centrifuged at 4°C to separate from each clot the serum samples

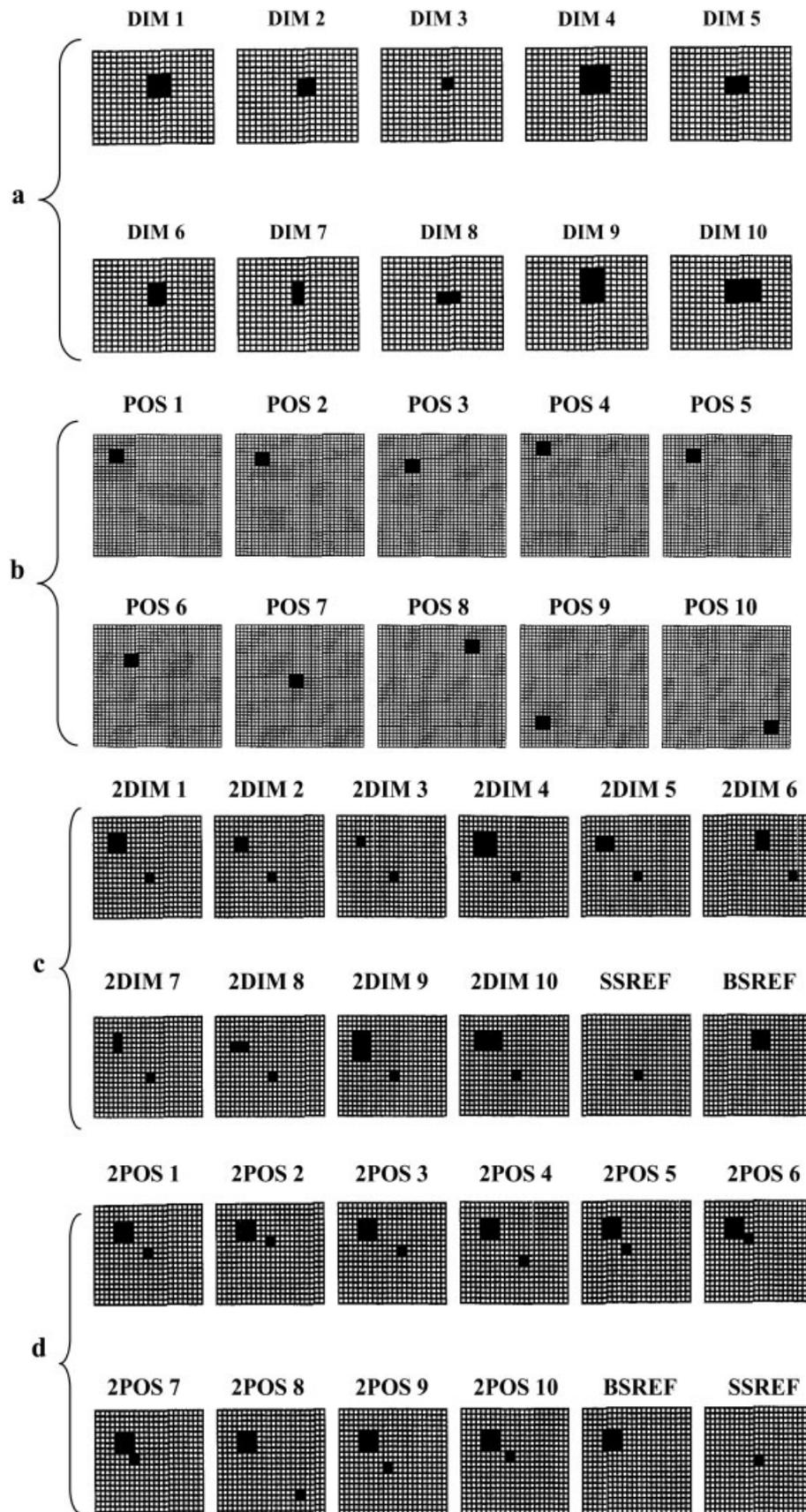


Figure 2. Simulated samples with (a) changes in shape and size of a spot; (b) changes in position of a spot; (c) changes in shape and size of the largest of two spots; (d) changes in position of the smallest of two spots.

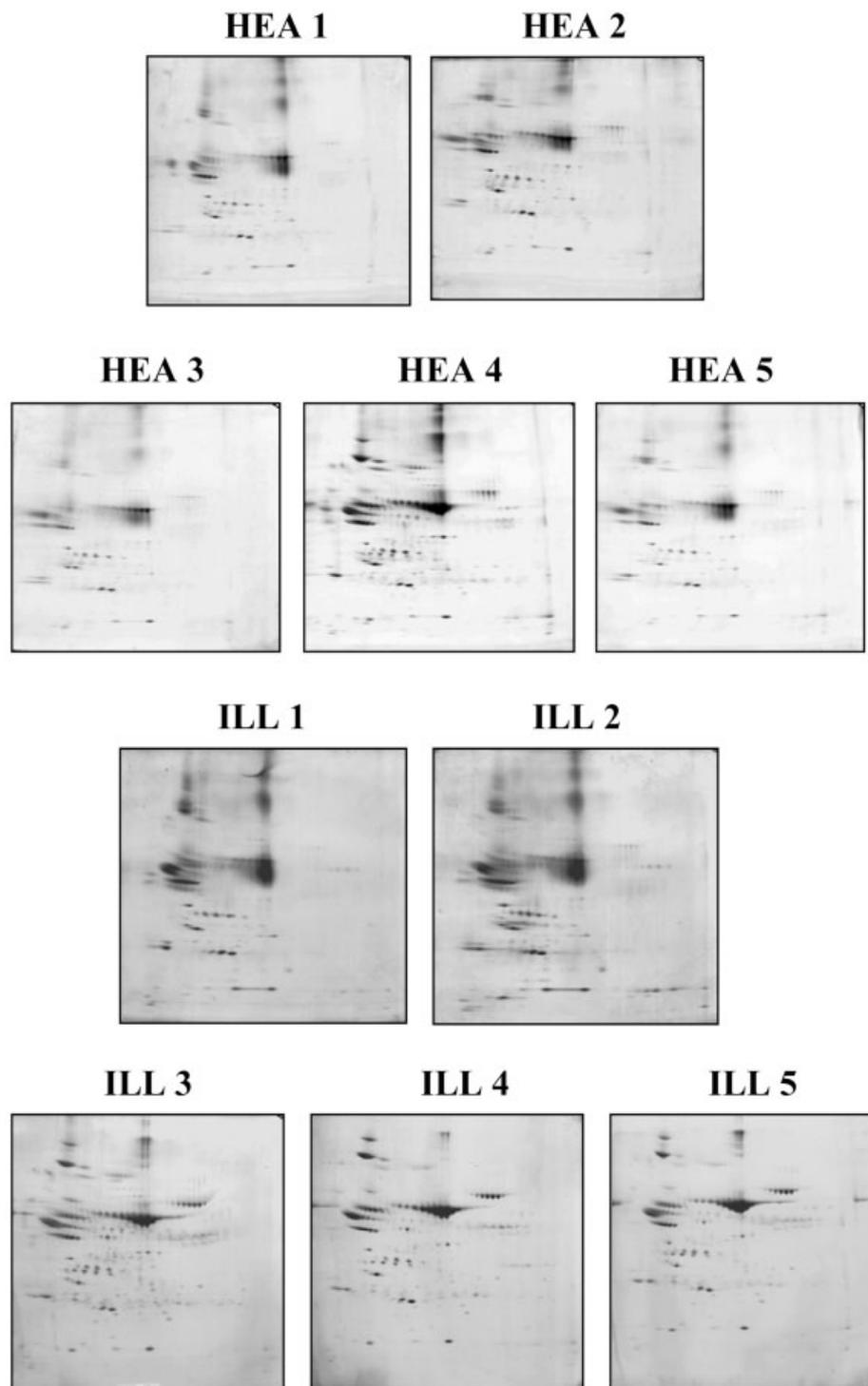


Figure 3. The ten experimental 2-D maps belonging to untreated (healthy, *i.e.*, HEA 1–HEA 5) and nicotine-treated (ILL 1–ILL 5) rat sera.

(about 200 μ L for each blood sample) and they were preserved at -20°C until the analysis was performed. One hundred μ L of serum was added with 0.4 mL of a denaturing solution containing 7 M urea, 2 M thiourea, 5 mM tributylphosphine (TBP) and 40 mM Tris. 20 mM iodoacetamide (IDD) was then added and alkylation

was continued for 1 h. The samples were then submitted to dialysis in order to eliminate the salts present in sera and then the reagents eliminated by the dialysis process were added (7 M urea, 2 M thiourea and 20 mM Tris); 2% CHAPS was added as a surfactant.

(iii) Electrophoretic separation (first dimension): the first-dimensional run was performed on strips (18 cm length, 0.5 mm thickness) with a nonlinear pH gradient from pH 3 to pH 10 (Amersham Pharmacia Biotech, Uppsala, Sweden). The IPG strips were rehydrated with 450 μ L of the sample solution obtained from the previous step 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, Hercules, CA, USA), 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 60 000 Vh for each strip and the temperature was set at 20°C.

(iv) Interfacing the IPG strips with the denaturing SDS solution: each strip was equilibrated with an SDS denaturing solution containing: 0.375 M Tris-HCl (pH 8.8), 6 M urea, 20% glycerol, 4 mM TBP, 4% w/v acrylamide, 2% SDS and traces of bromophenol blue. The contact took place for 30 min in tubes containing 20 mL each of the equilibrating solution. Each strip was then interfaced with an SDS gel slab using 0.5% agarose solubilized in the cathodic buffer (192 mM glycine, 0.1% SDS and Tris to pH 8.3). All the SDS gel slabs were cast with a two-vessel gradient mixer; the volumes of solutions used for polymerizing six SDS gel slabs of 1.5 mm in thickness and with a pore gradient from 7%T to 20%T, are given in Table 1. Polymerization took place overnight.

(v) Electrophoretic separation (second dimension): the second-dimensional 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 2 h, then 5 mA/gel for 1 h, 10 mA/gel for 20 h and 20 mA/gel until the end of the run. During the whole run the temperature was set at 11°C.

Table 1. SDS-PAGE composition with a porosity gradient 7–20% T

Compound	Light solution T = 7%, C = 2.5%	Dense solution T = 20%, C = 2.5%
Glycerol (87%)	–	49.7 mL
Acrylamide (40%T, 2,5%C)	35 mL	100 mL
SDS 10%	2 mL	2 mL
Tris-HCl, 1.875 M, pH 8.8	40 mL	40 mL
Water (Milli-Q)	122.7 mL	8 mL
TEMED 100%	80 μ L	80 μ L
APS 40%	200 μ L	200 μ L

(vi) Coomassie blue staining: the solution of Coomassie blue used to reveal the protein spots present on the 2-D maps contained 0.1% Coomassie Brilliant Blue G-250, 34% methanol, 3% phosphoric acid and 17% ammonium sulfate. All gels were left in the revealing solution for 1 h under gentle rocking before changing the solution with a fresh one; staining was continued overnight. Destaining was performed with a solution of 5% acetic acid till a clear background was achieved. All the samples were then scanned with a densitometer (Bio-Rad).

3 Results and discussion

The study of the effect of changes of the fuzzification parameters σ_x , σ_y was performed both on simulated and real samples. The values of the two parameters define the area of influence of each spot, *i.e.*, the variability of the spot position, shape and size. The larger the standard deviation along one of the electrophoretic directions, the larger the uncertainty of the spot position, shape and size along the corresponding direction. Figure 4 shows one of the real samples (HEA5) for all the levels chosen for σ_x and σ_y . It is evident that the spots increase their area of influence as the standard deviation increases. In this way it is possible to simulate the low reproducibility of the experimental results, taking into account the different reproducibility for each elution direction. The larger each standard deviation, the more fuzzy the resulting map.

In the real samples series, it is evident how the increase of the two σ values tends to mask the differences between different samples, with the nearest spots merging one into the other. The matching of the fuzzy maps was performed for each group of simulated and real samples and the similarity indexes obtained were analyzed separately by MDS. The final configuration provided by MDS for the first group of simulated samples is reported in Fig. 5a. The result depends heavily on the size of the spot for all σ values. In this case, the final coordinates are not very different in dependence on the different σ values, apart from a swap of DIM2 and DIM7 with σ greater than 0.75. The reference spot is the DIM1, which is in a central position in all the maps. All the other spots have been obtained by modifying the spot DIM1. The final configuration respects the observable differences of size and shape between the samples. So DIM5 and DIM6 are the nearest to DIM1, they have both been obtained from DIM1 by eliminating an entire row or column. In all instances, the farthest is DIM3, which represents the smallest spot. In the opposite side of the graph the largest spots are located, obtained from DIM1 by adding some rows and/or columns. The analysis of single spots seems

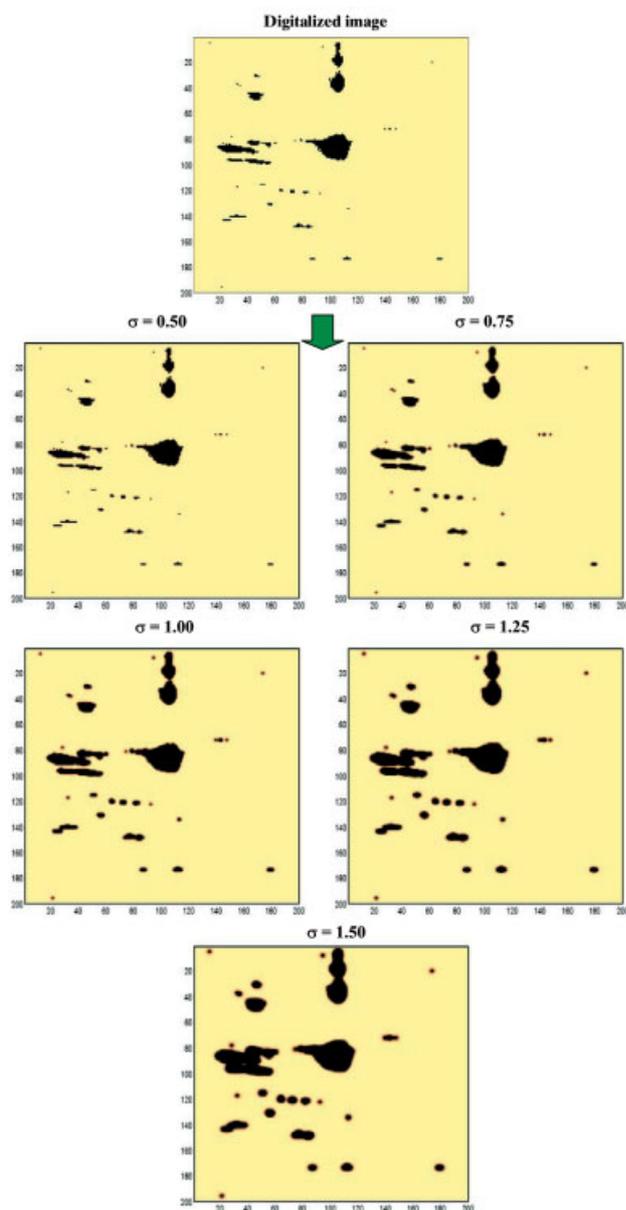


Figure 4. Fuzzy matrices of sample HEA5 at the five levels chosen for σ_x and σ_y .

to provide useful information, with a hierarchy of similarity based on the shape and size of the spots, which is reflected on their position in the scatter plots.

The result of MDS for the second group, where a square spot changes its position, is reported in Figure 5b. The lowest value of σ shows that three types of maps do exist: the POS9 and POS10 and a group which includes all other maps. The former two samples are completely different from the reference (POS1). In this case what we would expect is to obtain a representation of the spot centroids, since they do not change in size and shape,

but only in their position. Therefore, the MDS results should represent their shift along the space of the 2-D map. For low values of σ the method does not recognize that also sample POS8 is an outlier. This probably depends on the presence of some more diverse samples which obscure POS8 features. These highly diverse samples also affect the position of the other ones. Better results are obtained for larger values of σ , whereas the three extreme samples appear well separated one from another. Moreover, sample POS7 assumes a central position, which is in accordance with its equidistance from all other samples in the original map. By enlarging the zone where all other samples are concentrated, POS4 becomes correctly placed on the opposite side of the reference sample with respect to POS3 and POS6, while POS2 and POS5 remain very close to the reference sample, in agreement with their relative position in the original map. Again the method has been able to provide the correct result. The lowest value of σ does not provide optimal results, which are instead obtained for the other values tested.

The third group of simulated samples contains two spots of different size where the larger spot changes size and shape. There is a slight change in the configuration of the points obtained from the MDS treatment (Fig. 6a). The group of samples showing higher similarity becomes more compact in the plot as the σ value increases. All the scatter plots show the presence of two outliers, represented by the two reference samples which contain only one of the two spots. The major distance is correctly shown by the sample where the larger spot is missing. The other samples stay apart and their position can be easily analyzed zooming in the area. This allows the identification of the rules which determine the samples position: (i) sample 2DIM3, where the largest spot is reduced in size, is the most similar to the sample containing only the small spot; (ii) 2DIM10 sample, containing the largest large spot is the most similar to the sample containing only the large spot; (iii) the central position is occupied by sample 2DIM1, which is central since all other samples have been derived from it by changing the size and shape of the largest spot; (iv) the larger the distance from this sample, the larger the change in size of the spot. Also in this case it is possible to sort the samples with respect to their similarity, by simply analyzing their coordinates in the scatter plot. The low value of σ provides a less defined result. Probably the other configurations ($\sigma = 0.75, 1.0, 1.25, 1.5$) represent better solutions since the difference between the two “single spot” references and the others are better described.

In the case of the fourth group, where the position of the small spot is changed along the samples, there is not a large difference by varying the σ values (Fig. 6b). The

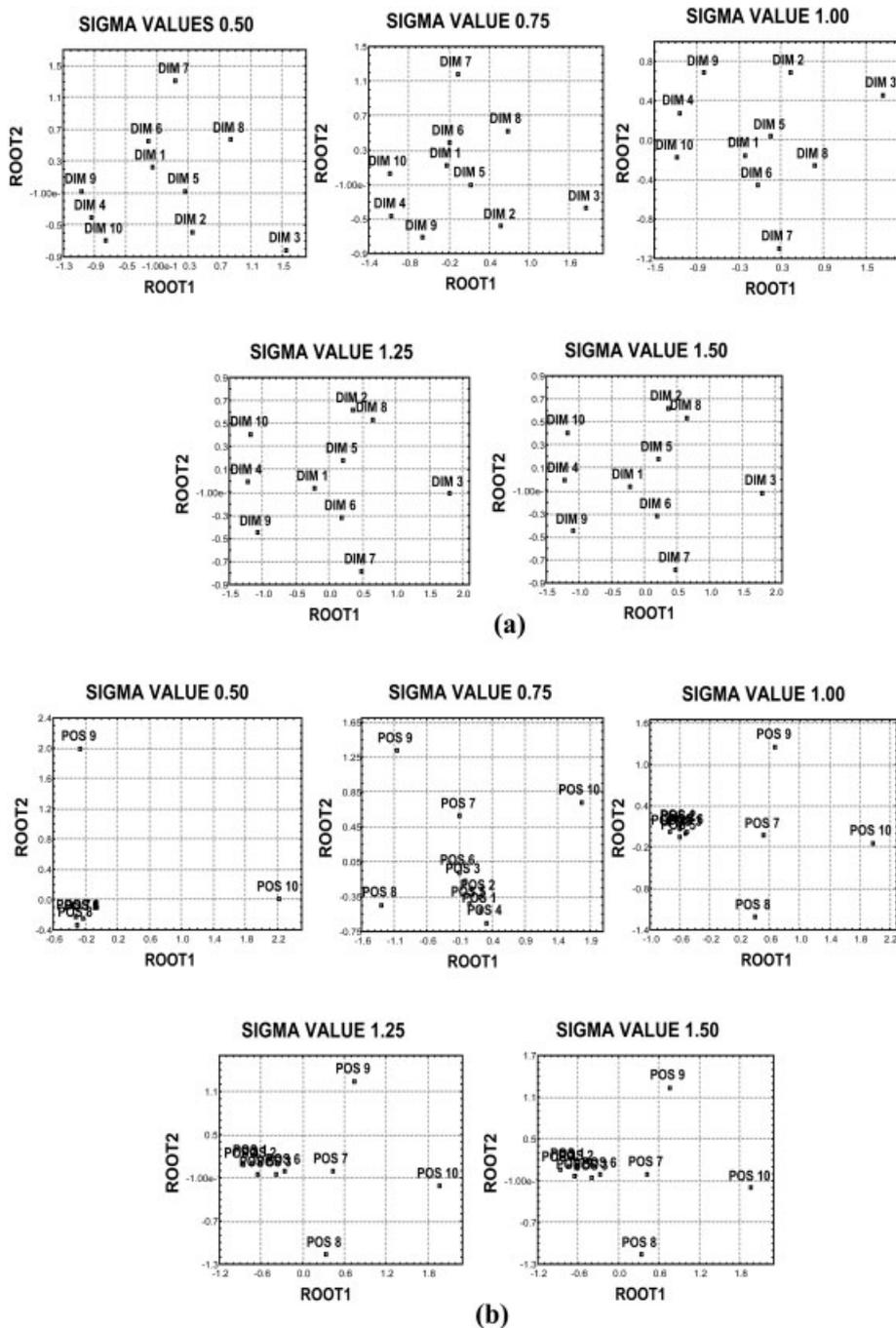


Figure 5. Final MDS configuration (a) for the first and (b) the second group of simulated samples for all σ values.

sample containing only the small spot stays aside and represents an outlier. The other reference, where there is only the large spot, is contained among the others, probably because of the higher “weight” of the large spot in the calculation. In all the scatter plots, the x-axis (root 1) takes into account the change in position of the small spot. The sample 2POS1 is in a central position and represents the starting sample for all position changes. The most distant samples are 2POS4 and 2POS3 on one

side; these are indeed the most different samples with respect to 2POS1. On the other side, the most distant samples are 2POS5 and 2POS7 where the small spot has moved in the opposite direction with respect to 2POS4 and 2POS3.

The last calculation was performed on the real samples (Fig. 7). From Fig. 3 it is evident that the 2-D maps are very different one from each other and it would be very

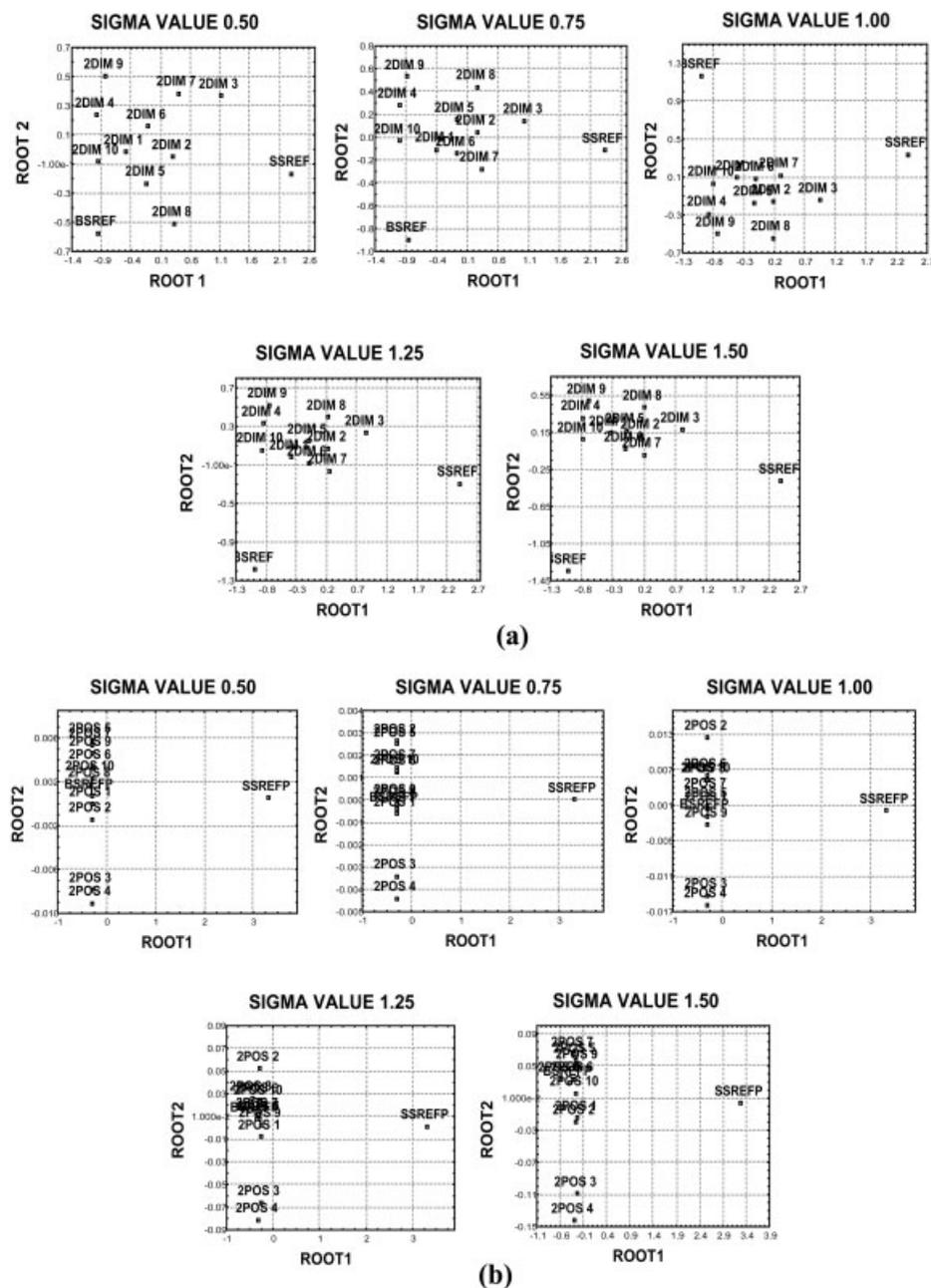


Figure 6. Final MDS configuration (a) for the third and (b) the fourth group of simulated samples for all σ values.

difficult, from a simple visual inspection, to state their similarity pattern. MDS provided a clear scheme which changes slightly passing from $\sigma = 0.5$ to the other values. In this case, our aim, *i.e.*, the separation of the two classes of rats, is obtained with all the σ values. In all the cases, it can be observed that it is possible to separate the two classes by means of only one dimension. Even if the separation is successful for all the σ values considered, it is more effective for the large ones. The best value of the fuzzification parameter seems to be $\sigma = 1.25$, which allows an optimal separation of the two classes of samples.

4 Concluding remarks

A new method for the statistical analysis of sets of 2-D maps in proteome research, has been developed. The method involves several steps, map digitalization, fuzzification, calculation of a similarity matrix and an MDS analysis, and has been applied satisfactorily to the analysis of both simulated and real samples. The analysis of the simulated maps allows an exploration of the effect of the σ parameter, *i.e.*, the effect of changing the uncertainty along each electrophoretic direction. The appli-

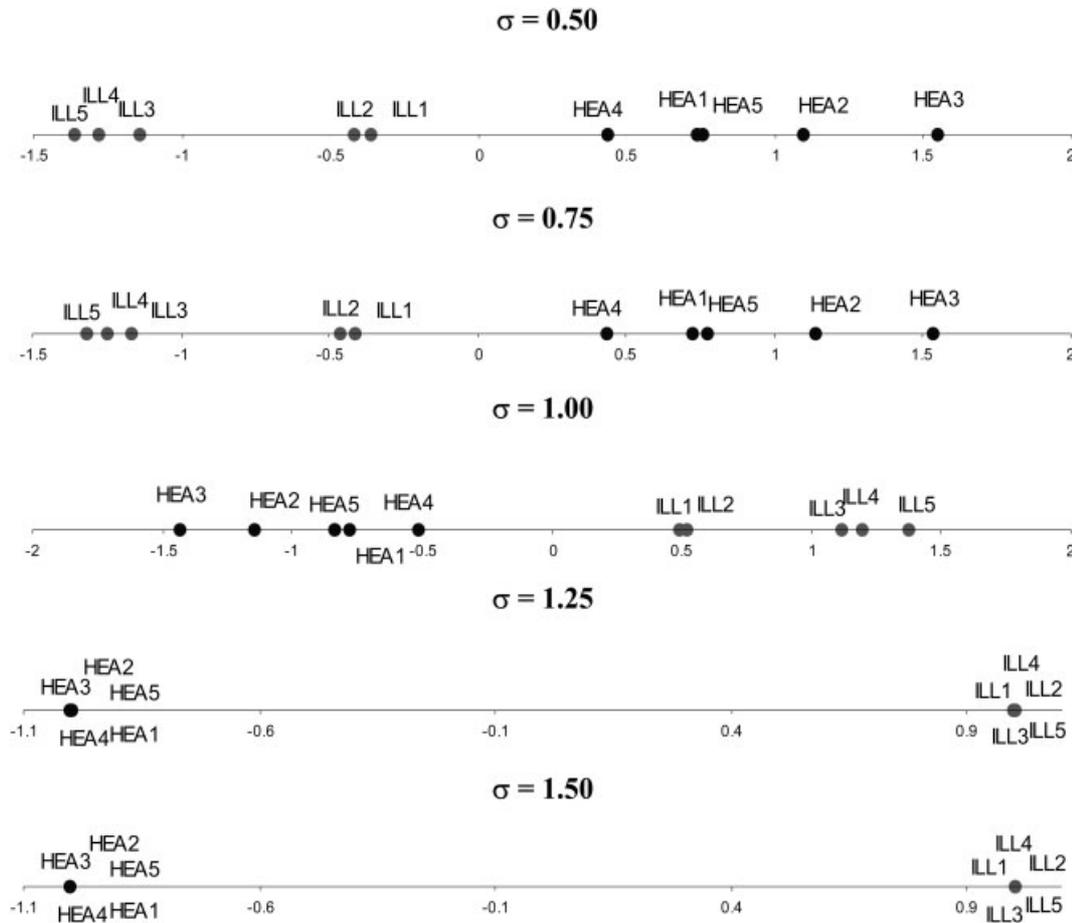


Figure 7. Final MDS configuration for the real samples for all σ values.

cation of the method to a complex dataset constituted by several 2-D maps of sera from rats treated with nicotine (ill) and controls has shown that this method allows the discrimination between the two classes.

The final aim of the method based on MDS is the diagnosis of the patient condition from a global analysis of a 2-D map, so no information is being acquired on the spots responsible for its classification. In order to extend the approach here described to pinpoint the differences in the protein content occurring between healthy and diseased (treated) subjects, the applications of other multivariate techniques (such as principal component analysis and classification methods) to the fuzzy maps are presently under development in our research group.

The authors gratefully acknowledge financial support from MIUR (Ministero dell'Istruzione, dell'Università e della Ricerca, Rome, Italy; COFIN 2000). Professor P. G. Righetti is also supported by grant No. GEC 556-2000

from GSK. The authors also thank N. Busti (GSK) for technical assistance in animal treatment and serum sample collection.

Received August 13, 2002

5 References

- [1] Banks, R. E., Dunn, M. J., Hochstrasser, D. F., Sanchez, J. C., Blackstock, W., Pappin, D. J., Selby, P. J., *The Lancet* 2000, 356, 1749–1756.
- [2] Blackstock, W. P., Weir, M. P., *TIBTECH* 1999, 17, 121–127.
- [3] De Vienne, D., Bost, B., Fievet, J., Zivy, M., Dillmann, C., *Plant Physiol. Biochem.* 2001, 39, 271–283.
- [4] Westergren-Thorsson, G., Malmstrom, J., Marko-Varga, G., *J. Pharmaceut. Biomed.* 2001, 24, 815–824.
- [5] Tsonis, P. A., Del Rio-Tsonis, K., *J. Chromatogr. A* 1995, 698, 361–367.
- [6] Toda, T., *Exp. Gerontol.* 2000, 35, 803–810.
- [7] Jungblut, P., Wittmann-Liebold, B., *J. Biotechnol.* 1995, 41, 111–120.

- [8] Anderson, N. L., Matheson, A. D., Steiner, S., *Curr. Opin. Biotechnol.* 2000, 11, 408–412.
- [9] Toda, T., Satoh, M., Sugimoto, M., Goto, M., Furuichi, Y., Rimura, N., *Mech. Ageing Dev.* 1998, 100, 133–143.
- [10] Macri, J., Rapundalo, S. T., *TCM* 2001, 11, 66–75.
- [11] Lawrie, L. C., Fothergill, J. E., Murray, G. I., *Lancet Oncol.* 2001, 2, 270–277.
- [12] Lin, J. D., Huang, C. C., Wenig, H. F., Chen, S. C., Jeng, L. B., *J. Chromatogr. B* 1995, 667, 153–160.
- [13] Simpson, R. J., Dorow, D. S., *Trends Biotechnol.* 2001, 19, S40–S48.
- [14] Zerr, I., Bodemer, M., Otto, M., Poser, S., Windl, O., Kretzschmar, H. A., Gefeller, O., Weber, T., *The Lancet* 1996, 348, 846–849.
- [15] Tsuji, T., Shimohama, S., Kamiya, S., Sazuka, T., Ohara, O., *J. Neurol. Sci.* 1999, 166, 100–106.
- [16] Aksenov, M. Y., Aksenova, M. V., Butterfield, D. A., Geddes, J. W., Markesbery, W. R., *Neuroscience* 2001, 103, 373–383.
- [17] Frey, S. J., Larsen, P. M., *Curr. Opin. Biotechnol.* 2001, 5, 26–33.
- [18] Patton, W. F., *J. Chromatogr. A* 1995, 698, 55–87.
- [19] Yamaoka, T., *Anal. Chim. Acta* 1998, 372, 91–98.
- [20] Celis, J. E., *et al.*, *FEBS Lett.* 1996, 398, 129–134.
- [21] Wilkins, M., *TIBTECH* 2000, 18, 91–92.
- [22] Naaby-Hansen, S., Waterfield, M. D., Cramer, R., *Trends Pharmacol. Sci.* 2001, 22, 376–384.
- [23] Patterson, S. D., *Curr. Opin. Biotechnol.* 2000, 11, 413–418.
- [24] Gianazza, E., Astrua-Testori, S., Caccia, P., Giacon, P., Quaglia, L., Righetti, P. G., *Electrophoresis* 1986, 7, 76–83.
- [25] Hoffmann, F., Kriegel, K., Wenk, C., *Discrete Appl. Math.* 1999, 93, 75–88.
- [26] Lee, K. H., *Trends Biotechnol.* 2001, 19, 217–222.
- [27] Vihinen, M., *Biomol. Eng.* 2001, 18, 241–248.
- [28] Marengo, E., Robotti, E., Gianotti, V., Righetti, P. G., *Ann. Chim.* 2003, 93, in press.
- [29] Tulp, A., Verwoerd, D., Neefjes, J., *J. Chromatogr. B* 1999, 722, 141–151.
- [30] Young, G., Householder, A. S., *Psychometrica* 1930, 3, 19–22.
- [31] Cox, T. F., Cox, M. A. A., *Multidimensional Scaling*, Chapman & Hall, London 1994.
- [32] Schoenberg, I. J., *Ann. Math.* 1935, 36, 724–732.
- [33] Young, G., Householder, A. S., *Psychometrika* 1938, 3, 19–22.
- [34] Gower, J. C., *Biometrika* 1966, 53, 325–338.
- [35] Shepard, R. N., *Psychometrika* 1962, 27, 125–140.
- [36] Shepard, R. N., *Psychometrika* 1962, 27, 219–246.
- [37] Kruskal, J. B., *Psychometrika* 1964, 29, 1–27.

Application of Three-Way Principal Component Analysis to the Evaluation of Two-Dimensional Maps in Proteomics

Emilio Marengo,^{*,†} Riccardo Leardi,[‡] Elisa Robotti,[†] Pier Giorgio Righetti,[§]
Francesca Antonucci,[§] and Daniela Cecconi[§]

Dipartimento di Scienze e Tecnologie Avanzate, Università del Piemonte Orientale, Spalto Marengo 33, 15100 Alessandria, Italy, Dipartimento Chimica e Tecnologie Farmaceutiche e Alimentari, Via Brigata Salerno, 16147 Genova, Italy, and Dipartimento Scientifico e Tecnologico, Università di Verona, Strada Le Grazie 15, 37134 Verona, Italy

Received January 2, 2003

Three-way PCA has been applied to proteomic pattern images to identify the classes of samples present in the dataset. The developed method has been applied to two different datasets: a rat sera dataset, constituted by five samples of healthy Wistar rat sera and five samples of nicotine-treated Wistar rat sera; a human lymph-node dataset constituted by four healthy lymph-nodes and four lymph-nodes affected by a non-Hodgkin's lymphoma. The method proved to be successful in the identification of the classes of samples present in both of the groups of 2D-PAGE images, and it allowed us to identify the regions of the two-dimensional maps responsible for the differences occurring between the classes for both rat sera and human lymph-nodes datasets.

Keywords: three-way principal component analysis • proteomics • multivariate analysis • 2D-maps

Introduction

Two-dimensional electrophoresis is perhaps the most widespread technique for the separation of proteins in the field of *proteomics*.^{1,2} Each cell or biological fluid has a rich protein content that can be constituted by thousands of proteins. These proteins can show great differences in structure and size, thus complicating the separation process. Two-dimensional gel electrophoresis is a very powerful tool in this perspective; it consists of two successive electrophoretic runs: the first run (through a pH gradient) separates the proteins with respect to their isoelectric point, whereas the second run (through a porosity gradient) separates them according to their molecular mass. Each two-dimensional map (called "2D-PAGE" from Poly Acrylamide Gel Electrophoresis) thus appears as a "snapshot" of the protein content of the investigated cell, with the proteins represented as spots spread all over the gel matrix.

The study of the protein content of different cell types has become fundamental in the past few years, with the development of *genomics* and *proteomics*.^{1,2} The proteins contained in a particular tissue, in fact, are related to its particular physiological state: many differences can occur in the protein content of a cell due to the onset of a given disease.^{3–6} These differences can consist of changes in the relative abundance

or even in the appearance/disappearance of some proteins. The 2D-PAGE maps can thus be used for both diagnostic and prognostic purposes by comparing maps belonging to healthy subjects with samples belonging to individuals affected by different pathologies.

Unfortunately, the problem concerning the comparison of different groups of 2D-PAGE maps is not a trivial one; maps belonging to replicates of the same sample may, in fact, present numerous differences from one another. The high variability (and the subsequent low reproducibility) of 2D-PAGE maps is to be ascribed to some experimental factors and some problems due to the specimen investigated, which can be listed as follows: (1) a complex sample (generally biological tissues) which can produce maps with thousands of spots; (2) a complex sample pretreatment, with many purification/extraction steps, which greatly contributes to the overall experimental error; and (3) a number of experimental factors that may influence the final result (temperature, polymerization and running conditions) and which have to be closely controlled during the experimental step.

The comparison of 2D-PAGE maps belonging to different classes of subjects (e.g., healthy and pathological) is made even more difficult by the small differences that often occur between healthy and ill subjects and which are much more difficult to recognize in a complex map.

Usually, this comparison is performed with the help of some specific softwares (i.e., Melanie III or PD-Quest);^{7–10} these

* To whom correspondence should be addressed. E-mail: marengo@tin.it.

[†] Dipartimento di Scienze e Tecnologie Avanzate, Università del Piemonte Orientale.

[‡] Dipartimento Chimica e Tecnologie Farmaceutiche e Alimentari.

[§] Dipartimento Scientifico e Tecnologico, Università di Verona.

methods are all based on a similar approach constituted by three main steps:

(1) The 2D-PAGE images to be compared are aligned so that each image is reduced to the same size as the others. This step requires the choice of at least two spots of sure identification in all of the maps; the maps are then matched one to the other on the basis of the position of these two spots.

(2) The spots present on each map are independently revealed; only the spots for which the optic density distribution fits a Gaussian function are maintained.

(3) The maps are matched one to the another to identify the common information (spots present in all the maps) and the differences (spots detected only on some of the samples). If the comparison is performed on a set of replicate maps, then this step produces a “synthetic” map which summarizes the common information and contains only the spots present in all of the compared maps.

The approach just described, even if very powerful, performs a comparison among the “synthetic” maps produced by the analysis of the different replicates obtained for each class; the comparison is no longer performed on the real samples but on “synthetic” ones, not considering the variability which characterizes each class.

Some studies already present in the literature are concerned with the development of methods for the classification of images based on pixel-data;¹¹ principal component analysis (PCA)^{12–15} has been applied to the study of DNA and RNA fragments of several biological systems^{16–19} and to the characterization of proteomic patterns of different classes of tissues.^{20–25} PCA has recently been applied also to the characterization of the anticancer activity of bohemine, a new omoleucine-derived synthetic cyclin-dependent kinase inhibitor, by Kovarova et al.²⁶ These multivariate methods require the previous analysis of the proteomic pattern images by standard softwares (i.e., Melanie III or PD-Quest) to identify the spots, so they present the disadvantage of being submitted to human choice for maps alignment. To avoid this disadvantage, we have considered the problem from another point of view: all of the maps belonging to the same type of sample are maintained during the analysis, and the comparison is performed on each sample. In this way, the information about the variability and the reproducibility of the 2D-PAGE maps is maintained and the comparison is performed on all of the real samples together and not on “synthetic” maps. A similar approach has already been applied by our research group to develop methods for the comparison of different classes of 2D-maps of healthy and pathological individuals.^{27,28} In the present paper, a new method based on three-way principal component analysis (three-way PCA)^{29–35} is presented. The 2D-PAGE images are first digitalized, transforming each image into a grid containing the value of the average optical density in each cell, revealed in the corresponding area of the map. By doing so, a three-mode dataset is obtained, with the three modes being the isoelectric point, the molecular mass, and the samples, respectively. The digitalized images obtained are then analyzed by three-way PCA.

In this paper, three-way PCA was applied to two different datasets:

(1) a set of 10 2D-PAGE maps: five belonging to control rat serum and five belonging to nicotine treated rat serum;

(2) a set of 8 2D-PAGE maps: four belonging to healthy human lymph-nodes and four belonging to human lymph-nodes affected by a non-Hodgkin's lymphoma.

The approach just described was applied to discriminate the two classes of samples present in both the datasets under investigation and to identify, in both cases, the zones of the maps responsible for the differences occurring between the two classes.

Theory

Applied Method. The method applied for the comparison of 2D-PAGE maps belonging to different classes consists of four steps:

(1) Digitalization of the Image. Each image is digitalized, producing a grid of (50 × 50) cells: each cell of the grid contains a value ranging from 0 to 1, according to the intensity of the signal in the correspondent position.

(2) Data Transformation. Data transformations have to be applied to the dataset to scale all of the samples and make them comparable. In the present study, maximum scaling was chosen as the most suitable for the dataset under investigation.

(3) Three-Way PCA. This multivariate technique is applied to the digitalized and scaled images to identify the classes of samples present in the dataset and to identify the zones of the maps responsible of the differences occurring between the classes.

(4) Differences Analysis. The maps rebuilt using the relevant factors are compared to identify the differences occurring between the centroids of the two classes of healthy and pathological protein patterns.

Digitalization. A 2D-PAGE appears, as just pointed out in the Introduction, as a transparent polymeric matter with the separated proteins spread all over it as colored spots. The revelation of the spots is performed, in general, with a solution of an organic coloring agent (e.g., Coomassie Blue) or by the deposition of silver onto the protein surface: the use of these staining solutions generates a coloring intensity which is proportional (within given limits) to the protein concentration. Each 2D-PAGE map is previously scanned with a GS-710 densitometer (Bio-Rad Labs, Richmond, CA), which transforms the 2D-PAGE into an image (200 × 200 pixels) in which each pixel corresponds to the value of the average optical density in the correspondent area.

The scanned images are then transformed into a grid of 50 × 50 cells, in which each cell contains a value ranging from 0 to 1. Each numerical value corresponds to the color intensity of the image calculated by averaging the intensities of the pixels which are contained in the correspondent cell. The values smaller than 0.4 were cut off and substituted with null values to eliminate the information about the color intensity of the background.

The actual choice of a 50 × 50 grid is not a constraint but it was suggested by computational and memory requirements.

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

$$x_k(i, j) = \frac{x_k(i, j)}{\max(x_k)} \quad (1)$$

where $x_k(i,j)$ is the value corresponding to the cell in (i,j) position in the k -th 2D-PAGE map and $\max(x_k)$ is the maximum value in all the cells of the k -th 2D-PAGE map.

By applying such a transformation to each two-dimensional map, the maximum signal intensity value of every 2D-PAGE map becomes a unit value; all of 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 the variation.

Three-Mode Principal Component Analysis. Three-way Principal Component Analysis (three-way PCA), based on Tucker3 model,^{29–35} has been used for the identification of the classes of samples present in the two datasets. The interest of three-way PCA is that it allows us to take into account the three-way structure of the data set 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., pH), J is the number of columns of the grids (the y coordinates, i.e., molecular mass), and K is the number of samples. The three-way PCA is based on the fact that the observed modes I, J, 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

$$x_{ijk} = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R a_{ip} b_{jq} c_{kr} g_{pqr} + e_{ijk} \quad (2)$$

where x_{ijk} denotes the elements of the initial matrix \mathbf{X} , a_{ip} , b_{jq} , and c_{kr} denote reduced elements of the component matrices \mathbf{A} , \mathbf{B} and \mathbf{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 \mathbf{G} , and e_{ijk} denotes the error term for element x_{ijk} and is an element of the $I \times J \times K$ array \mathbf{E} .^{29–35}

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 as much as possible body-diagonal. If this condition is sufficiently achieved, i.e., if the elements g_{111} , g_{222} ... are the only elements of the core matrix being significantly different from 0, then the rotated sets of loadings can also be interpreted jointly by overlapping them.

The datasets were analyzed with a program developed by the authors in the MATLAB 6.1 (The mathworks, Natick, MA) environment.

Experimental Section

Chemicals and Materials. Urea, thiourea, CHAPS, iodoacetamide (IAA), tributyl-phosphine (TBP), and sodium dodecyl sulfate (SDS) were obtained from Fluka Chemie (Buchs, Switzerland). Glutaraldehyde, sodium acetate trihydrate and formaldehyde were from Sigma (St. Louis, MO). Ampholines,

bromophenol blue and agarose were from Pharmacia-LKB (Uppsala, Sweden). Ethanol, methanol, acetone, acetic acid, silver nitrate, and citric acid monohydrate were from Merck (Darmstadt, Germany). Acrylamide, N,N -methylenebisacrylamide, ammonium persulfate, TEMED, the Protean IEF Cell, Protean II xi cell, GS-710 Densitometer, Mini Trans Blot Electrophoretic Transfer Cell, linear Immobiline dry strips pH gradient 3–10 (17 cm) were from Bio-Rad Labs (Richmond, CA).

Rat Serum and Human Lymph-Nodes. Rat Serum. The first investigated dataset consists of 10 samples belonging to two different groups: 5 samples of Wistar rat serum belonging to healthy individuals and 5 samples of Wistar rat serum belonging to nicotine treated individuals

The 10 2D-PAGE maps obtained are represented in Figure 1(a). Looking at the 2D-PAGE images, it is not very easy to distinguish the healthy individual from the nicotine treated one by a visual inspection of the image. Moreover, the 2D-maps belonging to the same group show a large variability of the spots number, position, shape, and size.

Sample Treatment. Five Wistar rats were treated for 14 days with a saline solution (control samples), and the other five were treated for the same 14 days with nicotine. The nicotine was administered subcutaneously by injecting 1 mL/Kg of a 0.4 mg/mL nicotine solution.

Blood samples were collected on the 14th day (when it is known that nicotine administration begins to induce dependence on treated rats) on rats which were fasted for 12 h prior to collection to avoid interferences due to high concentrations of lipids in the blood. All samples were centrifuged at 4°C to separate from each clot the serum samples (about 200 μ L for each blood sample) and they were preserved at –20 °C until the analysis was performed. One hundred μ L of serum was added with 0.4 mL of a denaturing solution containing 7 M urea, 2 M thiourea, 5 mM TBP (tributylphosphine), and 40 mM Tris. A 20-mM IAA (iodoacetamide) portion was then added, and alkylation was continued for an hour. The samples were then submitted to dialysis to eliminate the salts present in sera, and then the reagents eliminated by the dialysis process were added (7 M urea, 2 M thiourea and 20 mM Tris); 2% CHAPS (3-[(cholamidopropyl)-dimethylammonium]-1-propane-sulfonate) was added as a surfactant.

Human Lymph-Nodes. The proposed method was applied to a set of eight real samples, divided in two classes: four samples belonging to healthy human lymph-nodes, and four samples belonging to human lymph-nodes affected by Mantle Cell Lymphoma. Figure 1(b) represents the 8 experimental 2D-PAGE maps obtained.

Sample Treatment. Sample preparation and solubilization for biopsies was performed essentially as described by Sanchez et al.³⁶ for human lymphoma biopsies; the same approach was used also for human healthy lymph-node biopsies (control biopsies). Briefly, 10 frozen slices (about 20 μ m 5 mm \times 10 mm) of a human lymphoma biopsy were mixed with 100 μ L of 2D-solubilizing solution containing 7 M urea, 2 M thiourea, 3% CHAPS, 40 mM Tris, 5 mM TBP, 1% Ampholines, protease inhibitor, pH ca. 9. After centrifugation, for removal of particulate material³⁷, 20 mM IAA was added to perform complete alkylation of proteins.^{38,39} Salts, which can interfere with the 2D separation process and visualization of 2-D result, were removed by dialysis. Protein estimation for each sample was carried out with the Bio-Rad DC Protein Assay in order to load

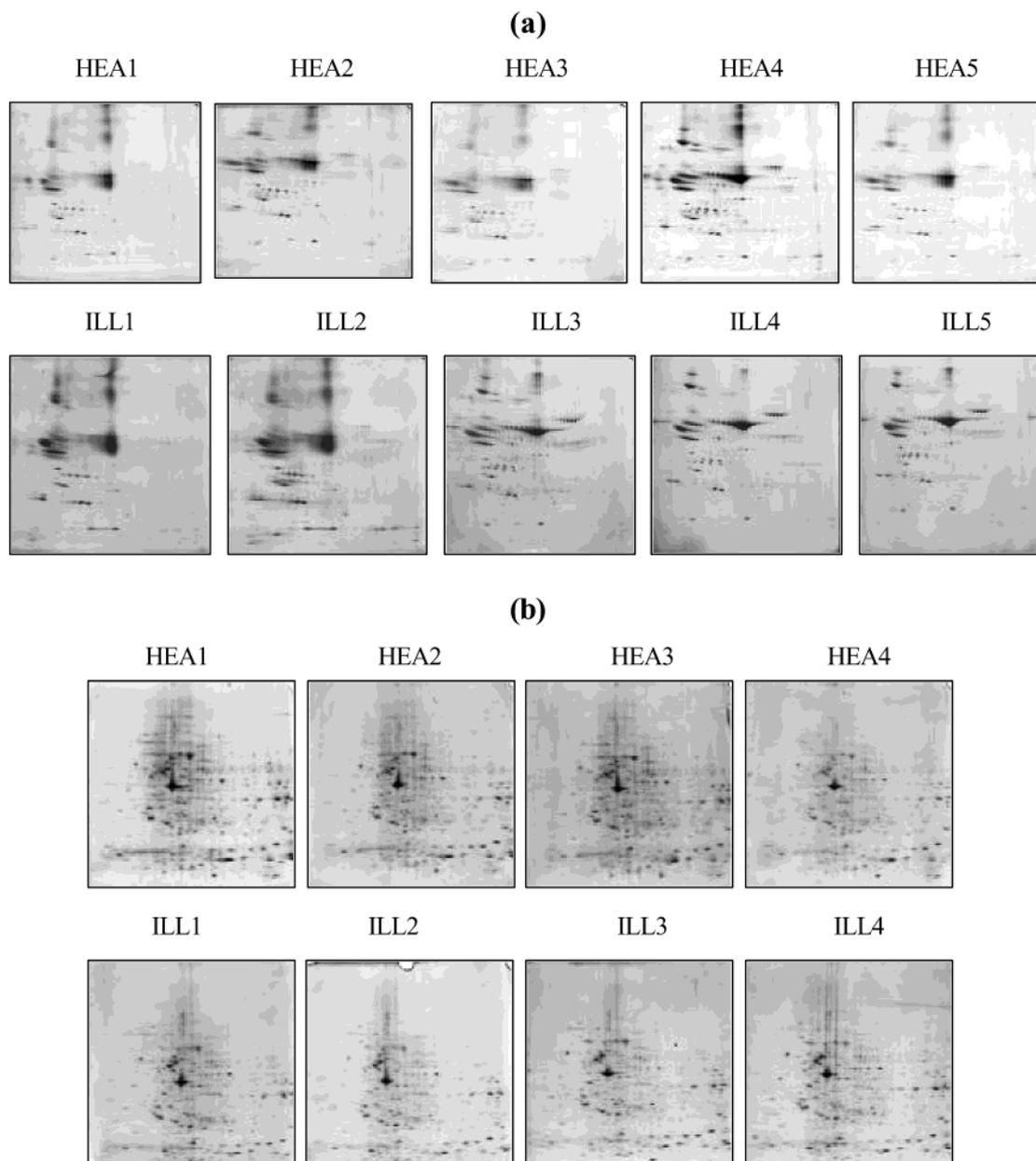


Figure 1. 2D-PAGE maps of the two investigated datasets: 10 rat serum samples (a) and 8 human lymph-node samples (b).

on the IPG strips always the same amount. Samples were stored at $-20\text{ }^{\circ}\text{C}$ until used. Samples from three control subjects (1 mg/ml of protein from each sample) were mixed to obtain a representative sample (pool), used to generate four control maps.

2D-PAGE Analysis. The first dimensional run was performed for rat serum samples on nonlinear IPG strips pH 3–10 (18 cm, Amersham Pharmacia Biotech, Uppsala, Sweden), and for human lymph-nodes samples on linear IPG strips pH 3–10 (17 cm, Biorad, Hercules, CA).

The IPG strips were re-hydrated with $450\text{ }\mu\text{L}$ of the sample solution. The passive gel re-hydration was allowed to continue for 8 h before the focusing step. Isoelectric focusing (IEF) was carried out with a Protean IEF Cell (Biorad, Hercules, CA). The total product time \times voltage applied was 75.000 Vh for each strip.

Each focused strip was equilibrated, for 25 min, with an SDS denaturing solution containing: 0.375 M Tris HCl (pH 8.8), 6 M urea, 20% glycerol, 2% SDS. In the second dimension, homemade vertical acrylamide-bisacrylamide gradient gels (8–18%; dimensions: $182 \times 190 \times 1.5\text{ mm}$) were used. Electrophoresis was performed by using a PROTEAN II xl Multi-Cell (Biorad). After separation in SDS-PAGE gels, the proteins were visualized by Colloidal Coomassie stain. The images of stained gels were scanned with GS-710 Calibrated Imaging Densitometer (Biorad, Hercules, CA).

Software. The software “MATLAB” (The Mathworks Inc, ver. 6.1) was applied for performing all of the data treatments: digitalization of the images and data scaling. Three-way PCA was performed by a self-made software developed with MATLAB, using the Tucker 3 algorithm. Most of the graphic representations were built with “UNSCRAMBLER” (Camo Inc.,

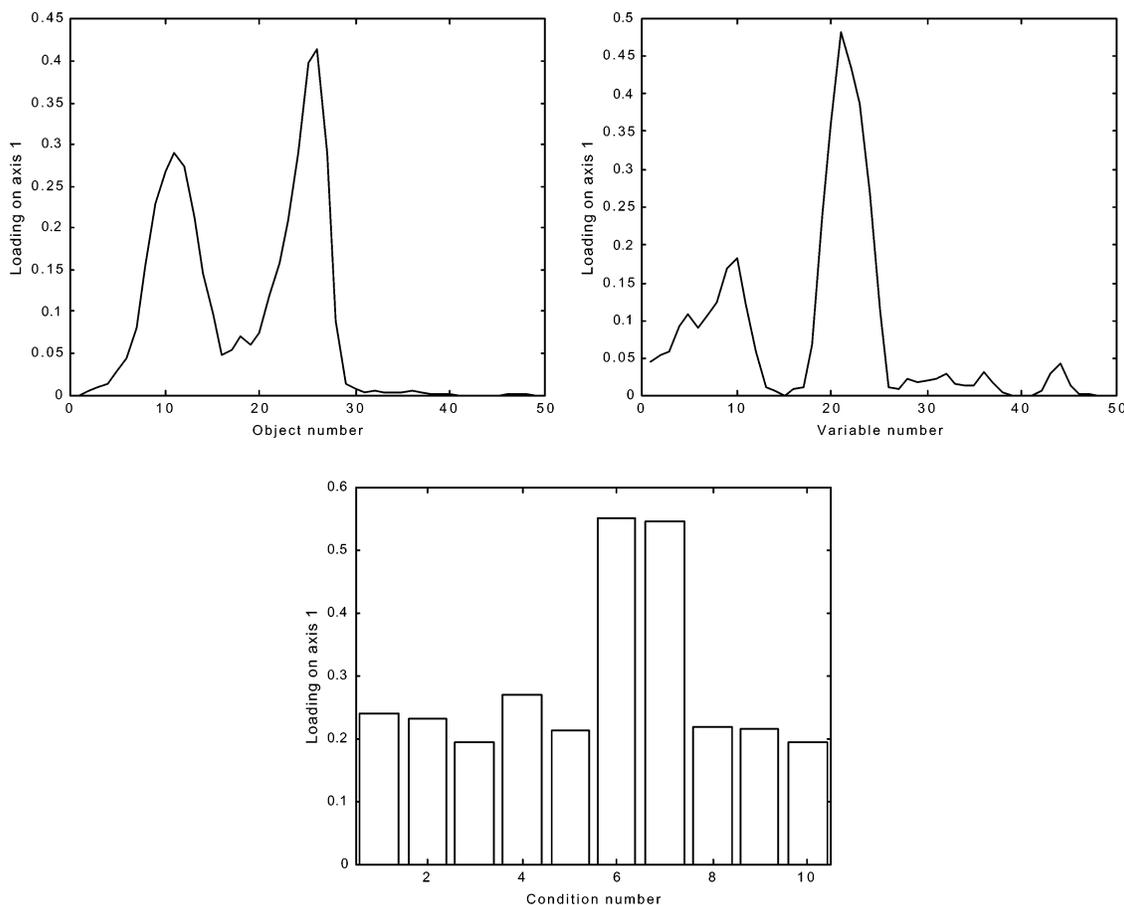


Figure 2. Plots of the three modes for the first entry of the core matrix g_{111} .

ver. 7.6), "MATLAB", "STATISTICA" (Statsoft Inc., ver. 5.1) and "EXCEL 2000" (Microsoft Corporation).

Results and Discussion

Rat Serum Dataset. When looking at the original 2D-PAGE images in Figure 1(a), sample HEA2 appears shifted with respect to the other samples of the same class. This peculiarity of sample HEA2 would lead to an incorrect analysis with three-way PCA because the major amount of explained variance would be probably led by this characteristic. Sample HEA2 was then shifted and matched to the other samples belonging to the same class.

Three-Way PCA. Before performing 3-way PCA, a maximum scaling procedure was applied to the dataset; this scaling technique produces some changes in the dataset: in the maximum-scaled patterns, some minor spots can be detected with respect to the original ones; in addition, maximum scaling is fundamental for obtaining a complete independence of the analysis from the staining step, often scarcely reproducible.

Three-way PCA performed on the normalized dataset gave the results represented in Figures 2 and 3 (two factors were retained for each mode); samples from 1 to 5 correspond to the control rat serum samples, while samples from 6 to 10 correspond to the nicotine-treated rat sera. The first two factors explain 59.8% of the total variance.

After body diagonalization³² the following core matrix is obtained; the cubic core matrix is reported according to the following unfolding

$$\begin{bmatrix} g_{111} & g_{121} & g_{112} & g_{122} \\ g_{211} & g_{221} & g_{212} & g_{222} \end{bmatrix} = \begin{bmatrix} 12.76 & -0.22 & 0.13 & 5.10 \\ 0.61 & 0.99 & -2.44 & 1.51 \end{bmatrix}$$

Since it is not super-diagonal, the plots of objects, variables and conditions can be interpreted jointly only for the first factor (the term g_{111} being by far the largest one).

The analysis is performed on one series of combinations at a time: this corresponds to examine one entry of the core matrix at a time. The first one, g_{111} , explains 51.6% of the total variance. The three plots, one for each mode, are represented in Figure 2. The first axis of the mode of conditions clearly discriminates samples 6 and 7 (ILL1 and ILL2) from the others; looking at the plots of the other two modes, it is possible to state that these two samples are characterized by larger values of the objects around 11 and 26 and of the variables around 21 and, at a lesser extent, around 10. By looking at the maps, it is easy to see that these samples are characterized by very intense spots in the rows around 11 and 26 (corresponding to pH 4.5–5 and 6–7) and in the columns around 21 (molecular mass 100–50 KDa).

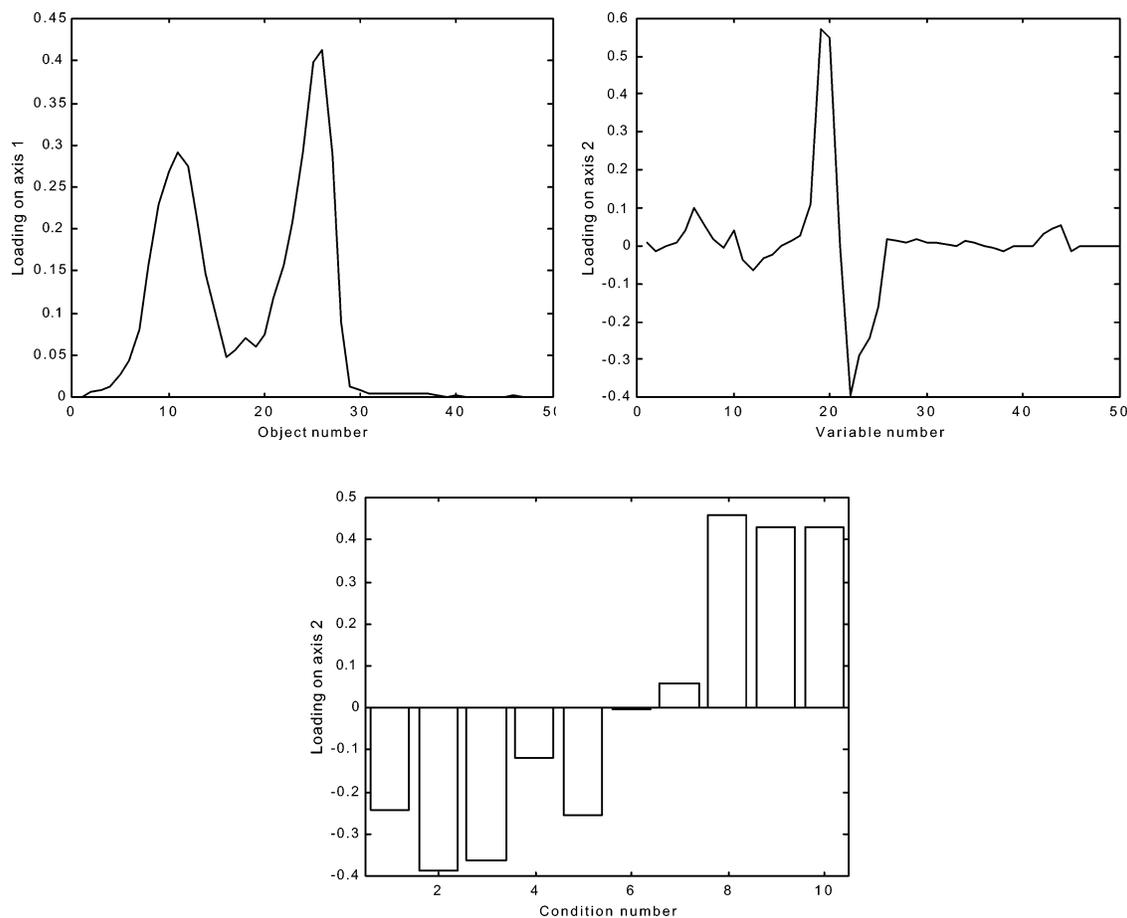


Figure 3. Plots of the three modes for the second entry of the core matrix g_{222} .

The second entry of the core matrix, g_{122} , explains 8.2% of the total variance. The three plots of this entry are represented in Figure 3. The second axis of the mode of conditions clearly separates the two classes of samples: the control ones at negative loadings and the treated ones at positive loadings. Looking at the plots of the objects and variables modes, the control samples appear characterized by small values of the objects around 11 and 26 and by large values of the variables around 22; the treated samples appear instead characterized by large values of the objects around 11 and 26 and of variables around 19.

It is therefore possible to notice that the information is mainly concentrated in objects around 11 and 26 (corresponding to pH 4.5–5 and 6–7) and in variables around 10 and 21 (molecular mass 200–150 KDa and 90–47 KDa).

Differences Analysis. The fundamental aim of the present study is to identify the regions of the maps responsible for the discrimination of the two classes of samples: the control and the treated ones. The identification of the differences occurring between the two classes of samples is performed by using the centroids (the averaged samples) of each class, in the space defined by variables and objects obtained from three-way principal component analysis. These centroids represent the average information concerning control and treated individuals contained in the first two factors which permit the class discrimination. The centroids can be re-projected in the original space, thus obtaining the corresponding 2D-map

images containing only the information accounted for by the first two 3-way factors. The images rebuilt in this way can be compared to identify the discriminant regions of the 2D-maps. This procedure allows a sort of filtering of the useful (discriminant) information contained in the 2D-maps.

The centroid of the control class is obtained on the basis of all of the five original maps belonging to this class, whereas the centroid of the treated class is calculated on the basis of only the three samples (ILL3, ILL4, ILL5) which appear the most different from the control class, as pointed out by three-way PCA. The two re-projected maps and their difference are represented in Figure 4(a). In Figure 4(b), the positive values (toward red) refer to regions which characterize the control sample, while the negative ones (toward blue) refer to regions which characterize the treated specimens. The control samples appear thus richer in spots than the diseased ones and the two classes show differences due both to the presence/absence of spots and to different relative intensities of the spots.

Human Lymph-Nodes Dataset. Three-way PCA was performed on the dataset of human healthy lymph-nodes (samples 1–4) and human lymph-nodes affected by a non-Hodgkin lymphoma (samples 5–8) after the application of a maximum scaling procedure.

Three-Way PCA. As in the previous dataset, a maximum scaling technique was applied: this allows for the appearance of minor spots in the maximum-scaled patterns, and it is

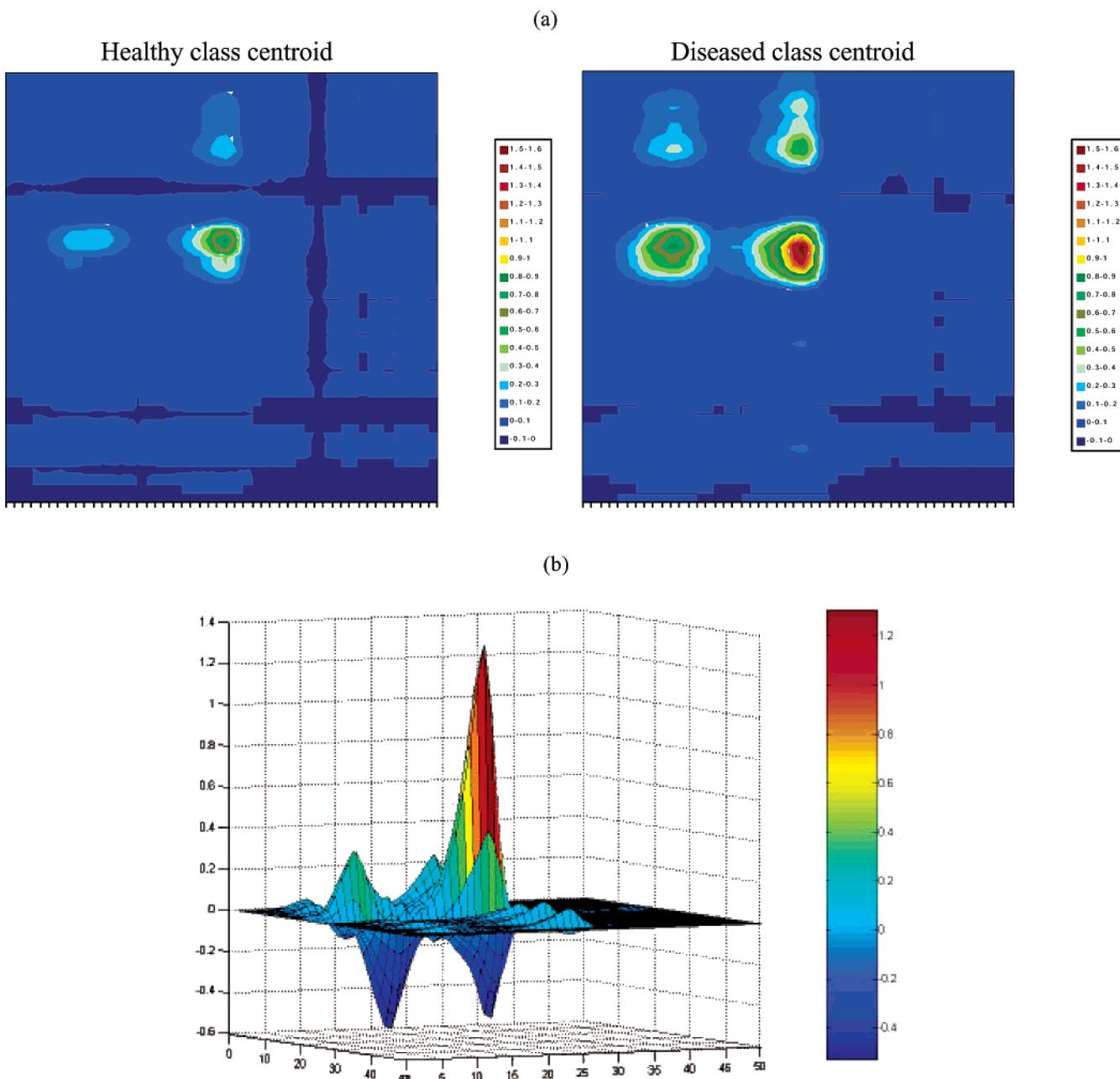


Figure 4. Contour plots of the centroids for the control and treated classes (a) and the correspondent map of the differences (b).

necessary because it makes the statistical analysis independent from the staining step.

The results of Three-way PCA performed on the normalized dataset are represented in Figure 5. Also, in this case, two components for each mode have been retained, and the model explains 50.8% of the total variance. After body-diagonalization, the following core matrix is obtained

$$\begin{bmatrix} 10.06 & 2.29 & -0.14 & 1.86 \\ 0.38 & -0.37 & 1.82 & 1.66 \end{bmatrix}$$

As for the previous dataset, it does not show a super-diagonal structure. However, the element $g_{1,1,1}$ is by far the largest one, and the element $g_{2,2,2}$ is quite close to the second largest. This behavior permits the joint interpretation of the three plots for

the first component and, at least to a certain degree, also for the second component.

From the plot of conditions (Figure 5), it can be seen that the data are grouped into three clusters, the control samples (samples 1–3, HEA1–3), the pathological ones (samples 6–8, ILL2–4), and two samples that lay at intermediate values along the second axis (samples 4 and 5, HEA4 and ILL1). Both classes show a sample (samples 4 and 5) which lays far from the others belonging to the same class.

The analysis of the plots of variables and of objects allows the identification of the variables and of the objects discriminating the three clusters and giving the greatest contribution to the model.

Samples 1–3 are characterized by large values of objects 22–27 (pH 5–6) and of variables 18–23 (mol.m. 62–50), whereas

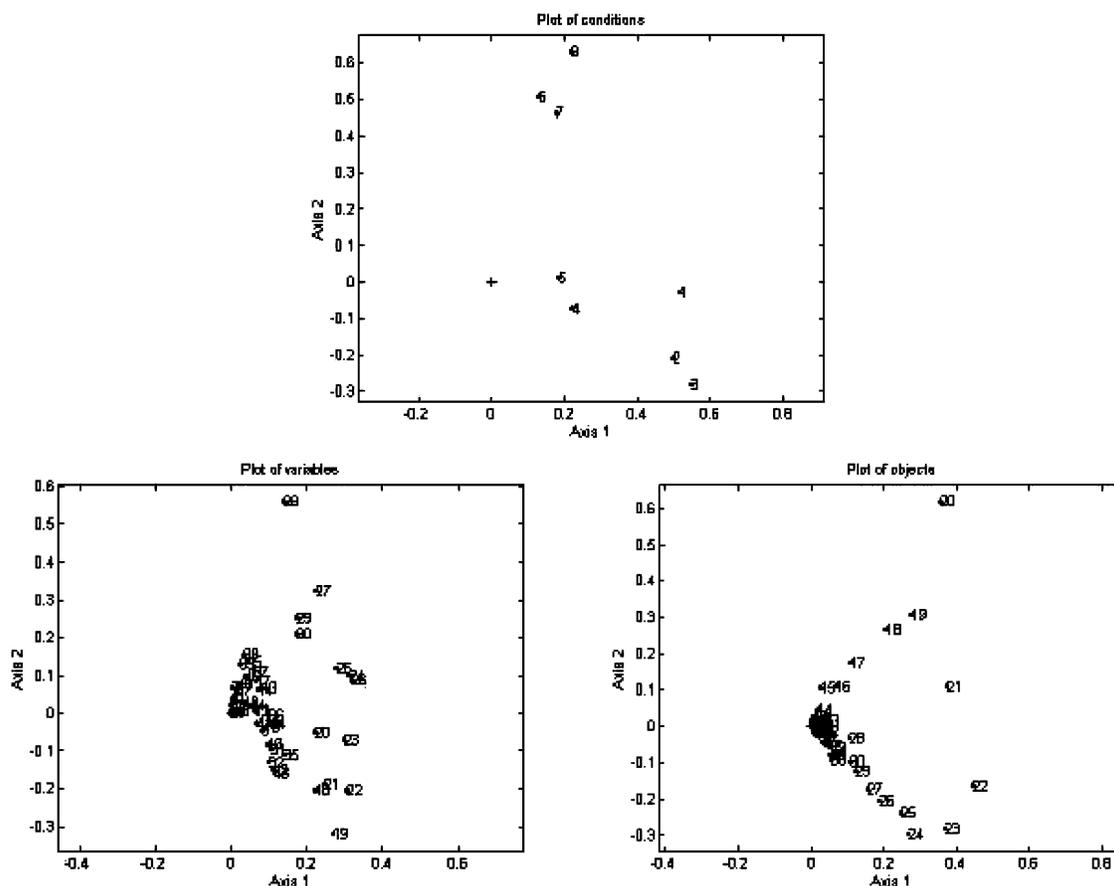


Figure 5. Results of three-way PCA applied to the human lymph-node dataset.

samples 6–8 are characterized mainly by large values of objects 18–20 and by large values of variables 27–30.

As in the previous case study, these findings are confirmed by the visual examination of the maps.

The plots of the objects and of the variables also suggest that the information is mainly resident in objects 15–30 (pH 4.5–7), whereas all the variables give some contribution to the model.

Differences Analysis. As for the previous dataset, the identification of the differences occurring between the two classes is performed on the basis of the centroids calculated for each class in the factors space. Samples HEA4 and ILL1 are excluded from the calculation of the centroids because these samples appeared intermediate between the groups of control and pathological samples, as pointed out by three-way PCA. So, the centroid of the control class was calculated on the basis of samples HEA1–3, whereas the centroid of the pathological one was calculated on the basis of samples ILL2–4. The two re-projected maps are reported in Figure 6(a). Figure 6(b) reports instead the difference between the two averaged samples: the regions toward red contain a signal intensity which is higher in the healthy samples, whereas colors toward blue refer to regions which contain a larger signal for the pathological samples. From this analysis, it appears clear that the differences between healthy and pathological samples are due entirely to the presence of larger signal intensities and a greater number of spots in the healthy samples.

Conclusions

In this paper, three-way PCA was performed on two different datasets: the first constituted by 10 samples of rat sera and the second constituted by 8 samples of human lymphonodes. Maximum scaling was applied to each dataset before performing three-way PCA to eliminate the influence of the staining procedure on the statistical analysis.

The applied method resulted a successful tool for the discrimination of the classes of samples present and for the identification of the zones responsible for the differences occurring between the samples belonging to the different classes. This last goal was realized by “differences analysis”, which allowed the identification of the regions which characterize each class of samples for both the considered dataset. It is necessary to stress that the method just presented represents a preliminary approach to the problem of comparing 2D-PAGE maps belonging to different classes: further studies are necessary before being able to use this method for diagnostic/prognostic purposes. Work is in progress to compare the present results with those obtained from the traditional approach.

Acknowledgment. The authors gratefully acknowledge financial support from MIUR (Ministero dell’Istruzione, dell’Università e della Ricerca, Rome, Italy; COFIN 2000). Professor P.G.R. is also supported by Grant No. I/R/294/02 from ASI (Agenzia Spaziale Italiana). The authors also thank N. Busti (GSK) for technical assistance in animal treatment and serum sample collection.

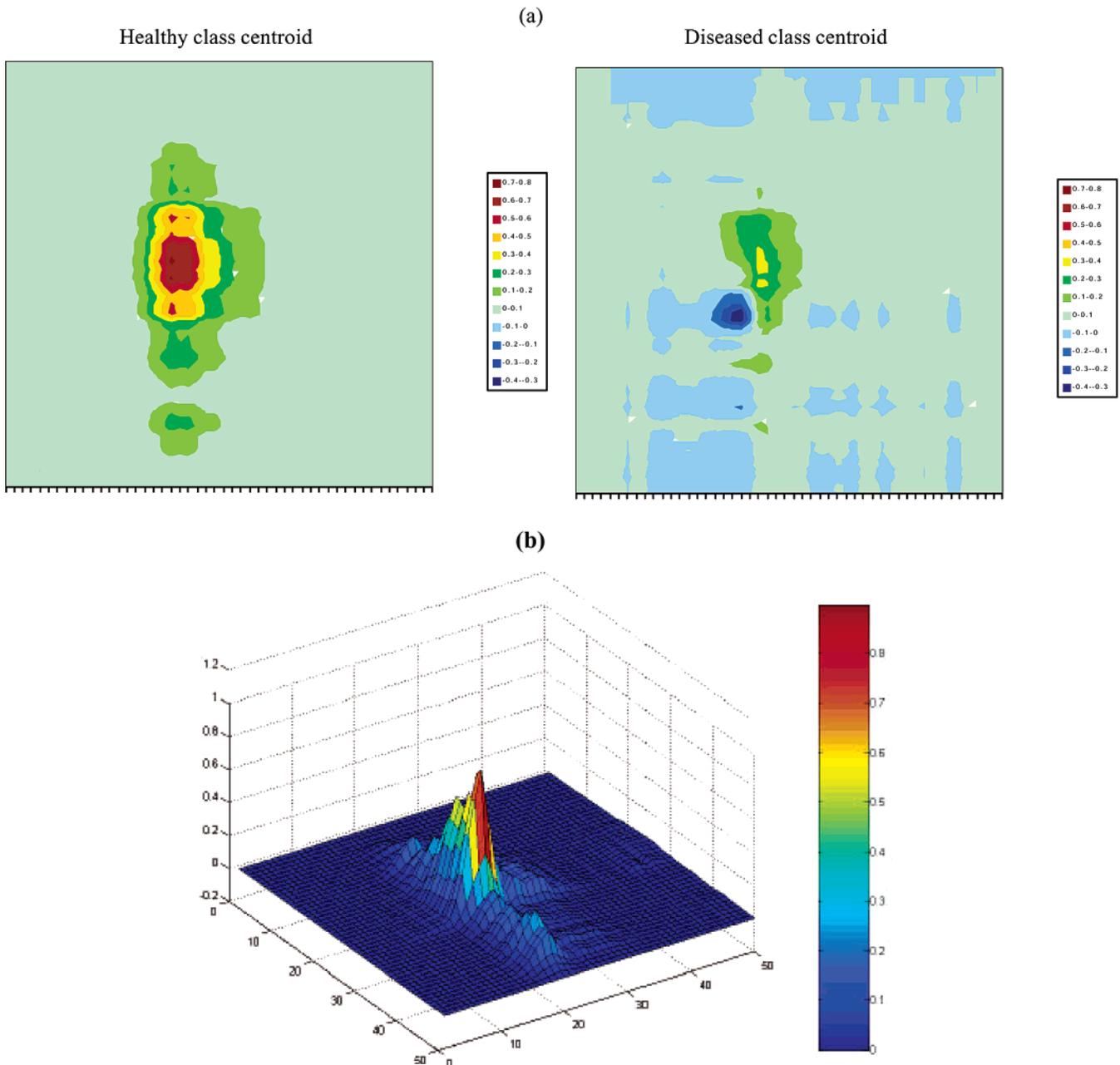


Figure 6. Contour plots of the centroids for the healthy and pathological classes (a) and the correspondent map of the differences (b).

References

- (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) Crenshaw T. R.; Cory J. G. *Adv. Enzyme Regul.* **2002**, *42*, 143–157.
- (4) Aksenov M. Y.; Aksenova M. V.; Butterfield D. A.; Geddes J. W.; Markesbery W. R. *Neuroscience* **2001**, *103*(2), 373–383.
- (5) Levine R. L.; Stadtman E. R. *Exp. Gerontol.* **2001**, *36*(9), 1495–1502.
- (6) Toda T. *Exp. Gerontol.* **2000**, *35*(6–7), 803–810.
- (7) Westergren-Thorsson G.; Malmstrom J.; Marko-Varga G. *J. Pharmaceut. Biomed. Anal.* **2001**, *24*(5–6), 815–824.
- (8) Bhatia K.; Lord R.; Stanton P. *Eur. J. Cancer* **2002**, *38*(S3), S156.
- (9) Hoffmann, F.; Kriegel, K.; Wenk, C. *Discrete Appl. Math.* **1999**, *93*, 75–88.
- (10) Lee, K. H. *Trends Biotechnol.* **2001**, *19*(6), 217–222.
- (11) Swets D. L.; Weng J. J. *IEEE T. Pattern Anal.* **1996**, *18*(8), 831–836.
- (12) Davis, J. C. *Statistics and Data Analysis in Geology*; John Wiley & Sons: New York, 1986.
- (13) Brereton, R. G. *Chemometrics: Application of Mathematics and Statistics to Laboratory Systems*; Ellis Norwood: New York, 1990.
- (14) Box, G. E. P.; Hunter, W. G.; Hunter, J. S. *Statistics for Experimenters*; Wiley: New York, 1978.
- (15) Massart, D. L.; Vandeginste, B. G. M.; Deming, S. N.; Michotte, Y.; Kaufman, L. *Chemometrics: a Textbook*; Elsevier: Amsterdam, 1988.
- (16) Leonard, R. B.; Mayer, J.; Sasser, M.; Woods, M. L.; Mooney, B. R.; Brinton, B. G.; Newcombgrayman, P. L.; Carroll, K. C. *J. Clin. Microbiol.* **1995**, *33*(10), 2723–2727.
- (17) Johansson, M. L.; Quednau, M.; Ahrne, S.; Molin, G. *Int. J. Syst. Bacteriol.* **1995**, *45*, 670–675.
- (18) Couto, M. M. B.; Vogels, J. T. W. E.; Hofstra, H.; Husintveld, J. H. J.; Vandervossen, J. M. B. M. *J. Appl. Bacteriol.* **1995**, *79*, 525–535.

- (19) Boon, N.; De Windt, W.; Verstraete, W.; Top, E. M. *FEMS Microbiol. Ecol.* **2002**, *39*, 101–112.
- (20) Alike, J. E.; AkenOva, M. E.; Fatokun, C. A. *Genet. Res. Crop Evol.* **1995**, *42*, 393–399.
- (21) Anderson, N. L.; EsquerBlasco, R.; Richardson, F.; Foxworthy, P.; Eacho, P. *Toxicol. Appl. Pharm.* **1996**, *137*, 75–89.
- (22) Vohradsky, J. *Electrophoresis* **1997**, *18*, 2749–2754.
- (23) Kovarova, H.; Radzioch, D.; Hajduch, M.; Sirova, M.; Blaha, V.; Macela, A.; Stulik, J.; Hernychova, L. *Electrophoresis* **1998**, *19*, 1325–1331.
- (24) 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.
- (25) Dewettinck, K.; Dierckx, S.; Eichwalder, P.; Huyghebaert, A. *Lait* **1997**, *77*, 77–89.
- (26) Kovarova, H.; Hajduch, M.; Korinkova, G.; Halada, P.; Krupickova, S.; Gouldsworthy, A.; Zhelev, N.; Strnad, M. *Electrophoresis* **2000**, *21*, 3757–3764.
- (27) Marengo, E.; Robotti, E.; Gianotti, V.; Righetti, P. G. *Ann. Chim. (Rome)* **2003**, *93*, in print.
- (28) Marengo, E.; Robotti, E.; Gianotti, V.; Righetti, P. G.; Cecconi, D.; Domenici, E. *Electrophoresis* **2003**, in print.
- (29) Tucker L. R. *Psychometrika* **1966**, *31*, 279–311.
- (30) Kroonenberg P. M. *Three-Mode Principal Component Analysis*; DSWO Press: Leiden, 1983.
- (31) Geladi P. *Chemometr. Intell. Lab.* **1989**, *7*, 11–30.
- (32) Kroonenberg P. M. *Chemometr. Intell. Lab.* **1989**, *15*, 143–157.
- (33) Henrion R. *Chemometr. Intell. Lab.* **1993**, *6*, 477–494.
- (34) Henrion R. *Chemometr. Intell. Lab.* **1994**, *25*, 1–23.
- (35) Henrion R.; Andersson C. A. *Chemometr. Intell. Lab.* **1999**, *47*, 189–204.
- (36) Sanchez, J. C.; Appel, R. D.; Golaz, O.; Pasquali, C.; Ravier, F.; Bairoch, A.; Hochstrasser, D. F. *Electrophoresis* **1995**, *16*, 1131–1151.
- (37) Rabilloud, T. *Electrophoresis* **1996**, *17*, 813–829.
- (38) Herbert, B.; Galvani, M.; Hamdam, M.; Olivieri, E.; McCarthy, J.; Pedersen, S.; Righetti, P. G. *Electrophoresis* **2001**, *22*, 2046–2057.
- (39) Galvani, M.; Hamdan, M.; Herbert, B.; Righetti, P. G. *Electrophoresis* **2001**, *22*, 2058–2065.

PR030002T

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

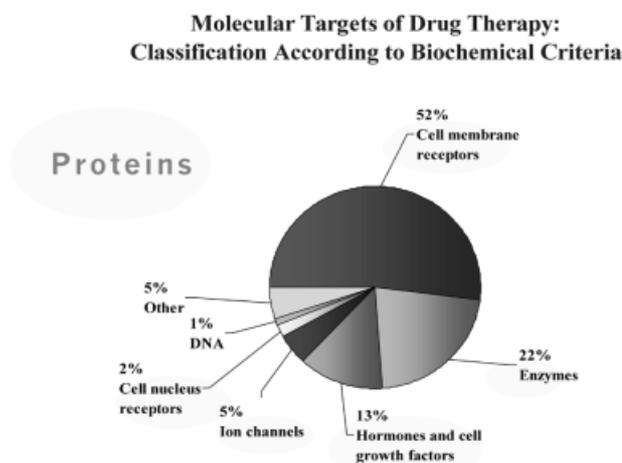


Figure 1 The distribution of important proteins acting as drug targets. Approximately 52% of them are cell membrane receptors, the second most abundant category being enzymes.

*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

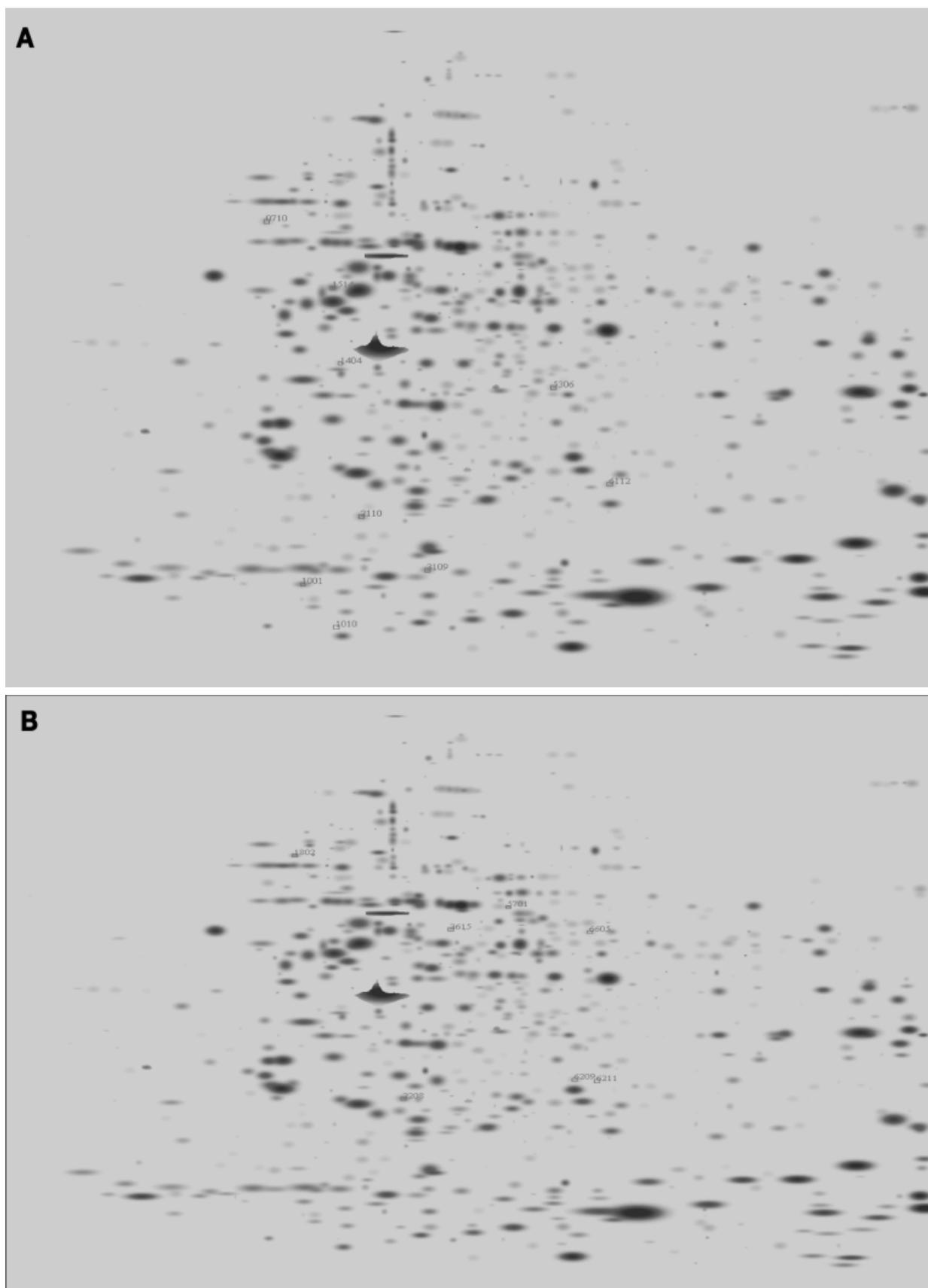
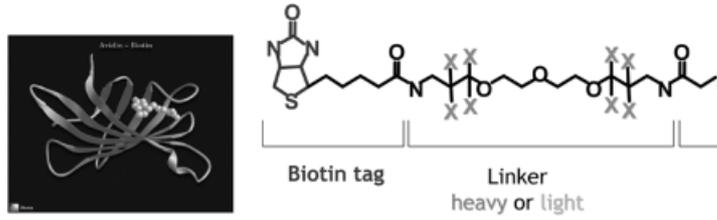


Figure 2 2-D pattern of Mantle Cell Lymphoma (MCL) tissue (A) and reactive lymph node (B). The 2-D electrophoresis was performed on a linear Immobiline gradient pH 3–10, followed by 8–18% SDS-PAGE. Proteins were detected by colloidal Coomassie staining. Image analysis of scanned gels was carried out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

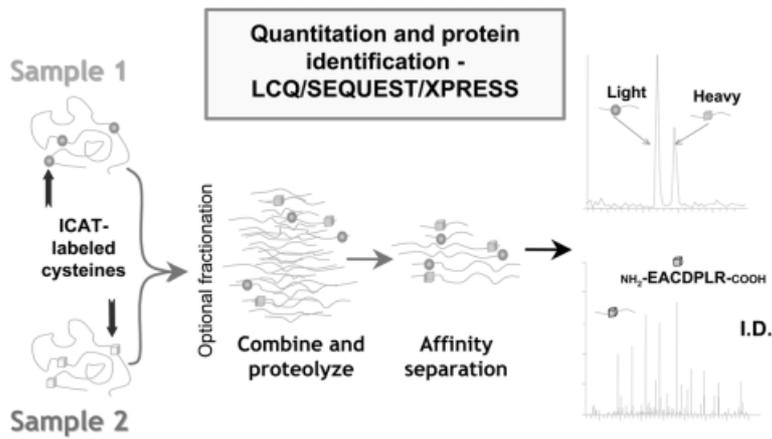


Figure 4 Protocol for the use of the ICAT label and affinity capture purification in quantitative proteomics (reprinted, with permission, from ref. 35).

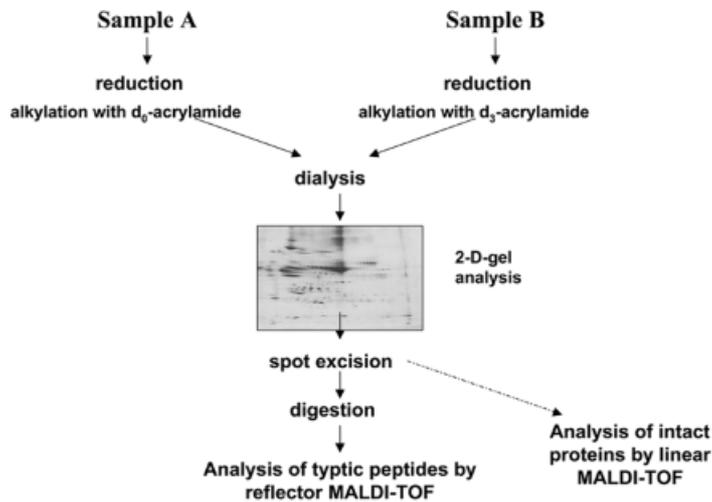


Figure 5 Protocol for the use of light/heavy acrylamide in quantitative proteomics of intact polypeptide chains (reprinted, with permission, from ref. 38).

Protein ID with Uninterpreted MS/MS

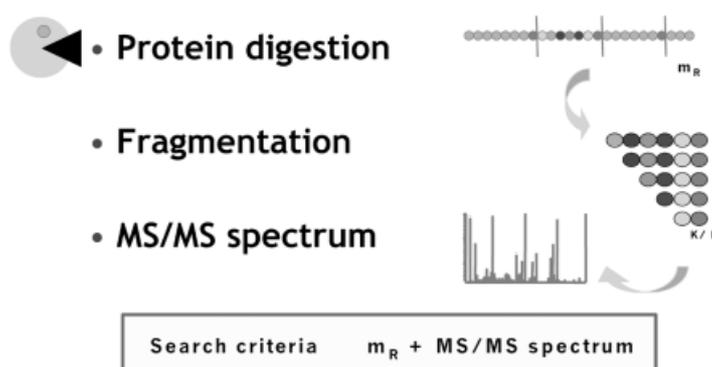


Figure 6 Protocol for identification of proteins extracted from a 2-D map *via* tryptic digestion and MALDI-TOF MS analysis, followed, if needed by an MS/MS sequencing procedure.

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

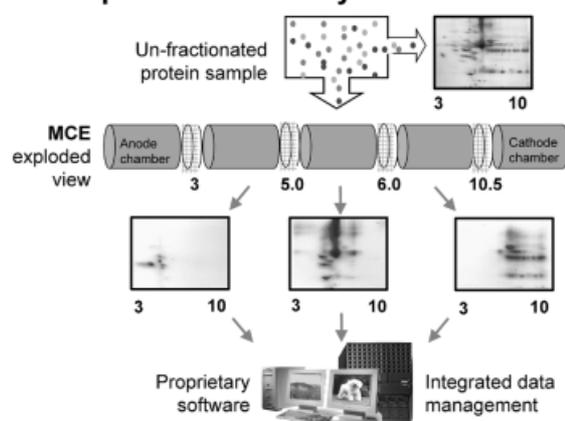


Figure 7 Scheme for sample pre-fractionation based on multicompartment electrolyzer. The upper right panel shows a 2-D map of unfractionated human serum vs. three different 2-D maps (lower panels) of three isoelectric fractions, captured into traps having membranes with pI 3.0–5.0; 5.0–6.0, and 6.0–10.5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

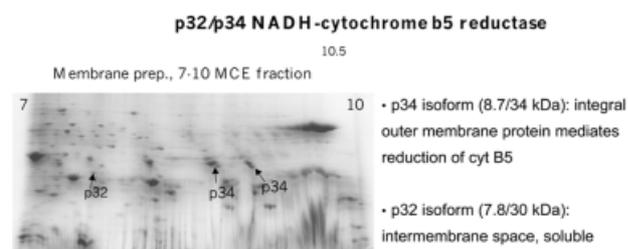


Figure 8 Example of pre-fractionation/enrichment of yeast membrane proteins and detection of truncated and isoforms of the p34 NADH-cytochrome b5 reductase (reprinted, with permission, from ref. 62).

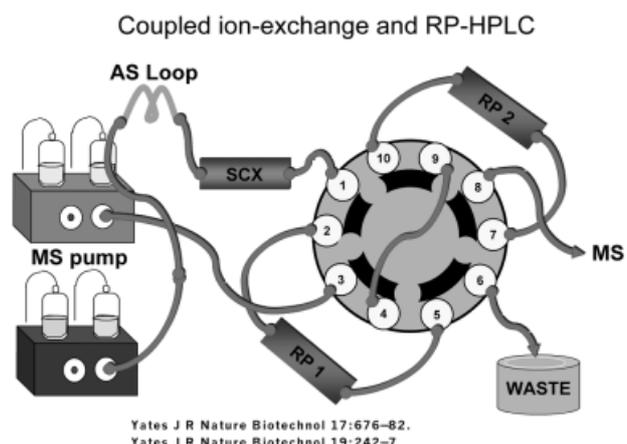


Figure 9 Scheme for coupled ion exchange and RP-HPLC in proteome analysis. SCX: strong cation exchanger column; AS loop: autosampler; MS: column eluate pumped to a mass spectrometer. It should be noted that, in most cases, such coupled columns are used for peptide fragment analysis.

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, Opitck *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 C_{18} 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 C_{18} -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 workstation to control this fully automated system. This system was applied

ProteinChip® Arrays

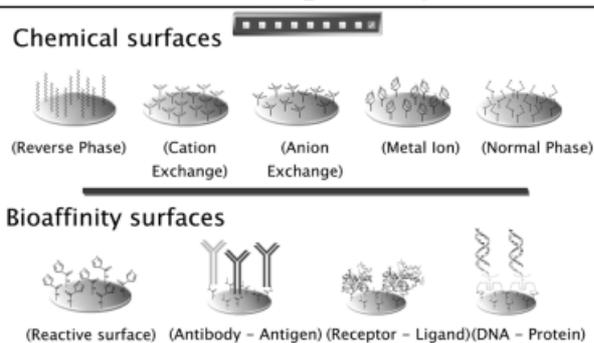


Figure 10 Scheme of different surfaces adopted in protein chip arrays (courtesy of CIPHERgen).

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

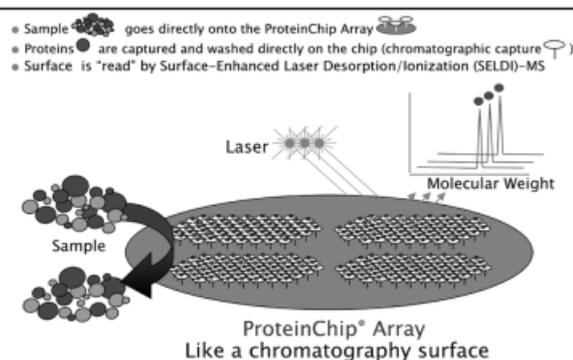


Figure 11 Scheme of processing/detection of protein species captured by different chemical surfaces (chromatographic adsorption; courtesy of CIPHERgen).

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

Daniela Cecconi¹
Aldo Scarpa²
Massimo Donadelli³
Marta Palmieri³
Mahmoud Hamdan⁴
Hubert Astner⁴
Pier Giorgio Righetti¹

¹Department of Agricultural
and Industrial Biotechnologies

²Department of Pathology,
Section of Anatomical Pathology

³Department of Neurological and
Visual Sciences,
Section of Biochemistry,
University of Verona

⁴Computational, Analytical &
Structural Sciences,
GlaxoSmithKline,
Verona, Italy

Proteomic profiling of pancreatic ductal carcinoma cell lines treated with trichostatin-A

A pancreatic adenocarcinoma cell line (Paca44) was treated with trichostatin-A (TSA), a potent inhibitor of histone deacetylases, in order to evaluate the effect of this drug on protein expression. Master maps of control and treated Paca44 cells were generated by analysis with the PDQuest software. The comparison between such maps showed up- and downregulation of 51 polypeptide chains, out of a total of 700 spots detected by a medium-sensitivity stain, micellar Coomassie Brilliant Blue. Fingerprinting by matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF)-mass spectrometry analysis enabled the identification of 22 of these spots. Among these proteins, of particular interest are the two downregulated proteins nucleophosmin and translationally controlled tumor protein, as well as the upregulated proteins programmed cell death protein 5 (also designated as TFAR19) and stathmin (oncoprotein 18). The modulation of these four proteins is consistent with our observation that TSA is able to inhibit cell growth of Paca44 by causing cell cycle arrest at the G2 phase and apoptotic cell death.

Keywords: Ductal carcinomas / Pancreatic tumors / Proteomics DOI 10.1002/elps.200305430

1 Introduction

Pancreatic cancer is an aggressive disease and is the fifth most frequent cause of cancer-related mortality in Western society. Around 90% of pancreatic neoplasms are thought to arise from ductal cells and are generally referred to as ductal adenocarcinoma (PDA). In recent years, PDA has become one of the best characterized human malignancies from a molecular standpoint and is in fact characterized by a unique molecular fingerprint constituted by frequent alterations of the *K-ras* [1, 2], *p53* [3], *p16INK4a* [4, 5], and *DPC4/Smad4* genes [5]. However, improved understanding of the underlying anomalies characterizing PDA is the only means to provide new markers for earlier diagnosis and to identify potential targets for therapeutic intervention. The importance of epigenetic events in human cancer has become increasingly apparent. These epigenetic events, comprising methylation of DNA and acetylation of histones, may be modulated by pharmacological intervention and several drugs have entered into clinical trials during the last decade [6, 7]. Those compounds that interfere with the acetylation/deacetylation pattern of histones seem to show particular promise.

Correspondence: Prof. Pier Giorgio Righetti, Department of Agricultural and Industrial Biotechnologies, Strada le Grazie 15, I-37134 Verona, Italy
E-mail: piergiorgio.righetti@univv.it
Fax: +39-045-8027929

Abbreviation: TSA, trichostatin-A

Chromatin modifying factors may be involved in cell proliferation and cancer [8, 9]. Generally, transcriptional efficiency is correlated with the relative activities of histone acetyltransferases (HATs) and histone deacetylases (HDACs). Acetylation is believed to separate the basic *N*-termini of histones from DNA that then becomes more accessible to transcription factors [10, 11]. Thus, histone acetylation leads to gene activation, while histone deacetylation leads to a tighter histone-DNA interaction and, accordingly, to gene repression. The role of histone acetylation and deacetylation in the genesis of cancer has been shown in recent studies [12]. During the last decade, a number of HDAC inhibitors have been identified that induce cultured tumor cells to undergo growth arrest, differentiation and/or apoptotic cell death [13]. These agents also inhibit the growth of cancer cells in animal models and several agents have been shown to slow tumor growth in animals at nontoxic doses. It has also been suggested that the action of HDAC inhibitors on gene expression is selective and restricted to a relatively small number of genes based on the results using differential display analysis of tumor cells cultured with trichostatin-A (TSA). In that study, only the expression of a small number of genes (8 of 340) was altered compared with untreated cells. Among these, two key genes involved in cell proliferation, namely *c-myc* and *p21*, were shown to have decreased or increased expression, respectively.

We report that the growth of the human pancreatic adenocarcinoma cell line Paca44 is strongly inhibited by TSA at submicromolar concentrations, and that the cellular

mechanisms underlying this effect consists in cell cycle arrest at the G2 phase and apoptotic cell death. In order to understand these effects, we have investigated the differential protein expression profiles of the pancreatic adenocarcinoma cell line Paca44, following treatment with TSA (a potent inhibitor of histone deacetylase) by using proteomics tools.

2 Materials and methods

2.1 Chemicals and materials

Urea, thiourea, CHAPS, iodoacetamide (IAA), tributylphosphine (TBP), and sodium dodecyl sulfate (SDS) were obtained from Fluka Chemie (Buchs, Switzerland). Bromophenol blue and agarose were from Pharmacia-LKB (Uppsala, Sweden). Acrylamide, *N,N'*-methylenebisacrylamide, ammonium persulfate, TEMED, the Protean IEF Cell, the GS-710 Densitometer as well as the linear Immobililine dry strips pH gradient 3–10 (17 cm long) were from Bio-Rad Laboratories (Hercules, CA, USA). Ethanol, methanol, and acetic acid were from Merck (Darmstadt, Germany). TSA was obtained from Sigma-Aldrich (St. Louis, MO, USA). A 3.3 mM solution of TSA in absolute ethanol was prepared and stored at -80°C until use.

2.2 Cell treatment with TSA

2.2.1 Cells and growth conditions

Paca44 human adenocarcinoma cells were grown in RPMI 1640 supplemented with 20 mM glutamine and 10% fetal bovine serum (FBS; BioWhittaker, Italy) and were incubated at 37°C with 5% CO_2 . Cells were seeded at a density of $25 \times 10^3/\text{cm}^2$.

2.2.2 Cell proliferation assay

Cells were plated in 96-well cell culture plates at sub-confluency (4×10^3 cells/well) and after 24 h treated with $0.2 \mu\text{M}$ TSA for 40 h. Cells were stained with Crystal Violet (Sigma), solubilized in PBS with 1% SDS, and measured photometrically ($A_{595\text{nm}}$) to determine cell viability. Three independent experiments were performed.

2.2.3 Cell cycle analysis

Cell cycle distribution was analyzed using propidium iodide (PI)-stained cells. Briefly, 10^6 cells were treated with $0.2 \mu\text{M}$ TSA for 48 h, washed with PBS, incubated with 0.1% sodium citrate dihydrate, 0.1% Triton X-100, 200 $\mu\text{g}/\text{mL}$ RNase A, 50 $\mu\text{g}/\text{mL}$ propidium iodide (Roche

Molecular Biochemicals, Indianapolis, IN, USA) and analyzed on a FACScalibur flow cytometer (Becton Dickinson, Franklin Lakes, NJ, USA). The percentage of cells in the various stages of the cell cycle was determined using the ModFitLT software program.

2.2.4 Apoptosis detection

Apoptosis was assessed by detection of histone-associated DNA fragments of cells treated for 40 h with $0.2 \mu\text{M}$ TSA. Cells were analyzed using the Cell-Death Detection ELISA^{Plus} assay (Roche Molecular Biochemicals). The assay is based on a quantitative sandwich-enzyme immunoassay principle using monoclonal antibodies anti-histone-biotin and anti-DNA-peroxidase (POD). POD was determined photometrically ($A_{405\text{nm}} - A_{492\text{nm}}$) with 2,2'-azino-bis(3-ethylbenzthiazoline-6-sulfonic acid (ABTS) as substrate. This allows the apoptotic determination of the enrichment of nucleosomes in the cytoplasmic fraction of cell lysates.

2.3 Two-dimensional gel electrophoresis

Protein extraction from cells untreated and treated with $0.2 \mu\text{M}$ TSA for 18 h was performed with lysis buffer (40 mM Tris, 1% NP 40, 1 mM Na_3VO_4 , 1 mM NaF, 1 mM PMSF, protease inhibitor cocktail). Cells were left in lysis buffer for 30 min in ice. After centrifugation at $14\,000 \times g$ at 4°C , for removal of particulate material, the protein solution was collected and stored at -80°C until used. Seventeen cm long, pH 3–10 immobilized pH gradient strips (IPG; Bio-Rad, Hercules, CA, USA) were rehydrated for 8 h with 450 μL of 2-D solubilizing solution (7 M urea, 2 M thiourea, 5 mM tributylphosphine, 40 mM Tris, and 20 mM IAA) containing 2 mg/mL of total protein from Paca44 cells. 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 70 000 Vh for each strip and the temperature was set at 20°C . For the second dimension, the IPG strips were equilibrated for 26 min by rocking in a solution of 6 M urea, 2% SDS, 20% glycerol, 375 mM Tris-HCl, pH 8.8. The IPG strips were then laid on an 8–18%T gradient SDS-PAGE with 0.5% agarose in the cathode buffer (192 mM glycine, 0.1% SDS and Tris to pH 8.3). The anodic buffer was a solution of 375 mM Tris HCl, pH 8.8. The electrophoretic run was performed by setting a current of 2 mA for each gel for 2 h, then 5 mA/gel for 1 h, 10 mA/gel for 20 h, and 20 mA/gel until the end of the run. During the whole run the temperature was set at 11°C . Gels were stained overnight with colloidal Coomassie blue (0.1% Coomassie Brilliant Blue G-250,

34% v/v methanol, 3% v/v phosphoric acid, and 17% w/v ammonium sulfate), while destaining was performed with a solution of 5% acetic acid until a clear background was achieved.

2.4 Protein pattern analysis

The 2-DE gels were scanned with a GS-710 densitometer (Bio-Rad), and analyzed with the software PDQuest Version 6.2 (Bio-Rad). A match set was created from the protein patterns of the two independent cellular extracts (control cell line, TSA-treated cell line). A standard gel was generated out of the image with the highest spot number. Spot quantities of all gels were normalized to remove nonexpression-related variations in spot intensity, so 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 optical density (OD). Statistical analysis of PDQuest allowed the study of proteins that were significantly increased or decreased in TSA-treated cell line. 51 spots were found to be differently expressed in TSA-treated Paca44, of which 29 spots were upregulated (with a significance level α of 0.05) and 22 spots downregulated (with $\alpha = 0.05$).

2.5 Protein identification by mass spectrometry

2.5.1 *In situ* digestion and extraction of peptides

The spots of interest were carefully excised from the gel with a razor blade, placed in Eppendorf tubes, and destained by washing three times for 20 min in 50% v/v acetonitrile, 2.5 mM Tris, pH 8.5. The gel pieces were dehydrated at room temperature and covered with 10 μ L

of trypsin (0.04 mg/mL) in Tris buffer (2.5 mM, pH 8.5) and left at 37°C overnight. The spots were crushed and peptides were extracted in 15 μ L of 50% acetonitrile, 1% v/v formic acid. The extraction was conducted in an ultrasonic bath for 15 min. The sample was centrifuged at 8000 \times g for 2 min, and the supernatant was collected.

2.5.2 MALDI-TOF analysis

The extracted peptides were loaded onto the 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 α -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 reflectron mode, with an accelerating voltage of 20 kV. Peptide masses were searched against SWISS-PROT, TrEMBL and NCBIInr databases by utilizing the ProteiLynx program (Micromass) or NCBIInr database by using the ProFound (<http://129.85.19.192>) program.

3 Results

Figure 1 shows that TSA is able to inhibit proliferation of Paca44 cells (Figure 1a) *via* mechanisms involving apoptotic cell death (Fig. 1b) and cell cycle arrest at the G2 phase (Fig. 1c). Figure 2 shows a standard 2-D map of the Paca44 cell line, stained with colloidal Coomassie blue. About 700 polypeptide spots could be revealed in the pH 3–10 interval with this medium-sensitivity stain. After matching the master map of the control Paca44 cells with the master map of the TSA-treated cells, 51 polypeptide chains, marked by black circles, were

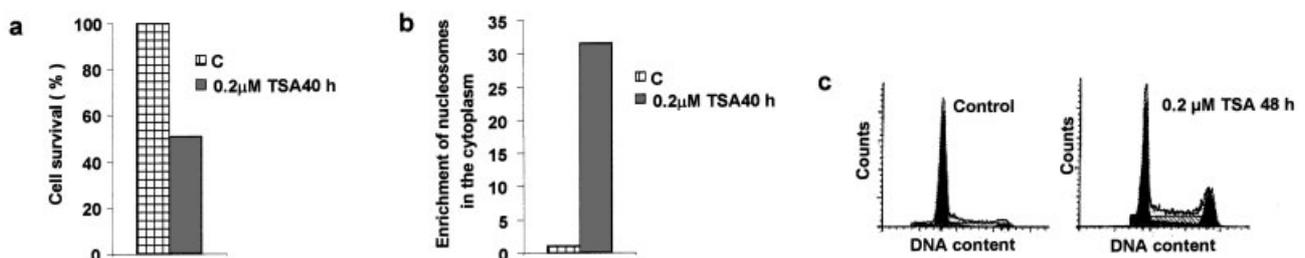


Figure 1. Effect of 0.2 μ M TSA on growth of Paca44 adenocarcinoma cell line. The cells were seeded in 96-well plates (a, b) or in 60 mm diameter plates (c) and incubated overnight. Cells were further incubated in the absence or presence of TSA for the time indicated in each panel. (a) Proliferation. Cell proliferation was determined using the Crystal Violet colorimetric assay as described in Section 2.2.2. Values are the means of triplicate wells. (b) Apoptosis. The cells were analyzed by detection of histone-associated DNA fragments in the cytoplasmic fraction of cell lysates as described in Section 2.2.4. The values are the means of triplicate wells. (1c) Cell cycle. The cell cycle distribution was analyzed by a flow cytometer after DNA staining with propidium iodide. Similar results were obtained from three independent experiments.

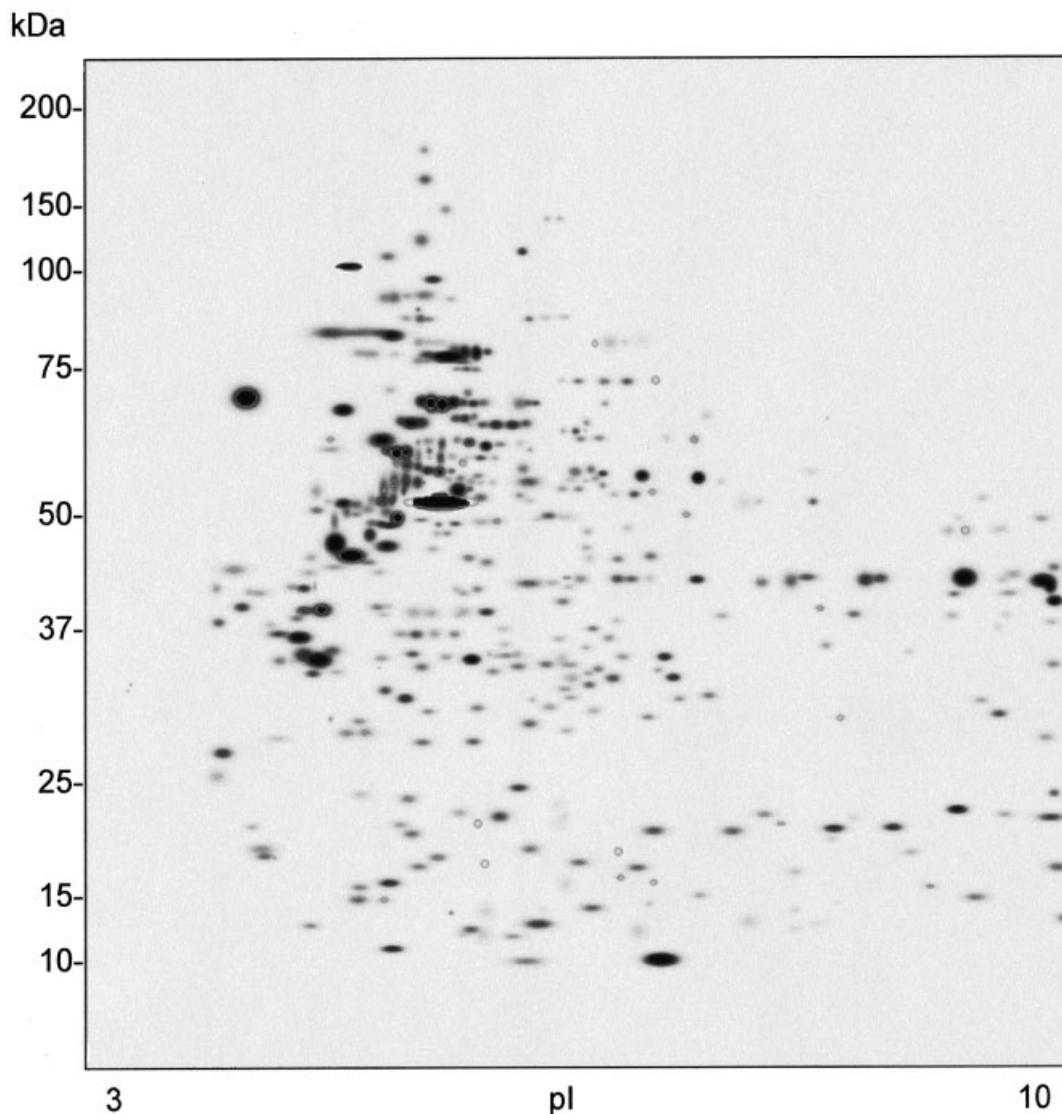


Figure 2. Master map of the pancreatic adenocarcinoma cell line Paca44. The 51 spots differently expressed after trichostatin A treatment are marked by black circles.

found to be differentially expressed after treatment with the anti-tumoral drug. All these spots were eluted and treated as described in Section 2.6. In Fig. 3, we have numbered the 22 spots which could be identified by MS analysis, 1–12 those downregulated, 13–22 those upregulated. Their identification, together with experimental and theoretically predicted pI and M_r values, are given in Table 1. Figure 4 gives an example of a gel area (central, low- M_r area of Fig. 3) in which some upregulated proteins could be found. The spot marked with a question mark in Fig. 4A represents a polypeptide which was 10-fold upregulated, which gave a good mass spectrum of the tryptic digest, but which could not be identified in any of the databases presently available.

4 Discussion

In the modern proteome world, differential analysis of protein expression is becoming a rapidly growing field for profiling a number of pathological states, such as cancer growth. The most widely accepted techniques, by and large, are still 2-D maps, with charge and mass coordinates [14], followed by spot excision and mass spectrometric identification of tryptic digests [15]. Although differential profiling can be today performed by a vast number of techniques (for a review, see [16]), in conventional 2-D maps the preferred method is still statistical analysis performed on master maps created from two independent cellular extracts (control vs. pathological or treated cells).

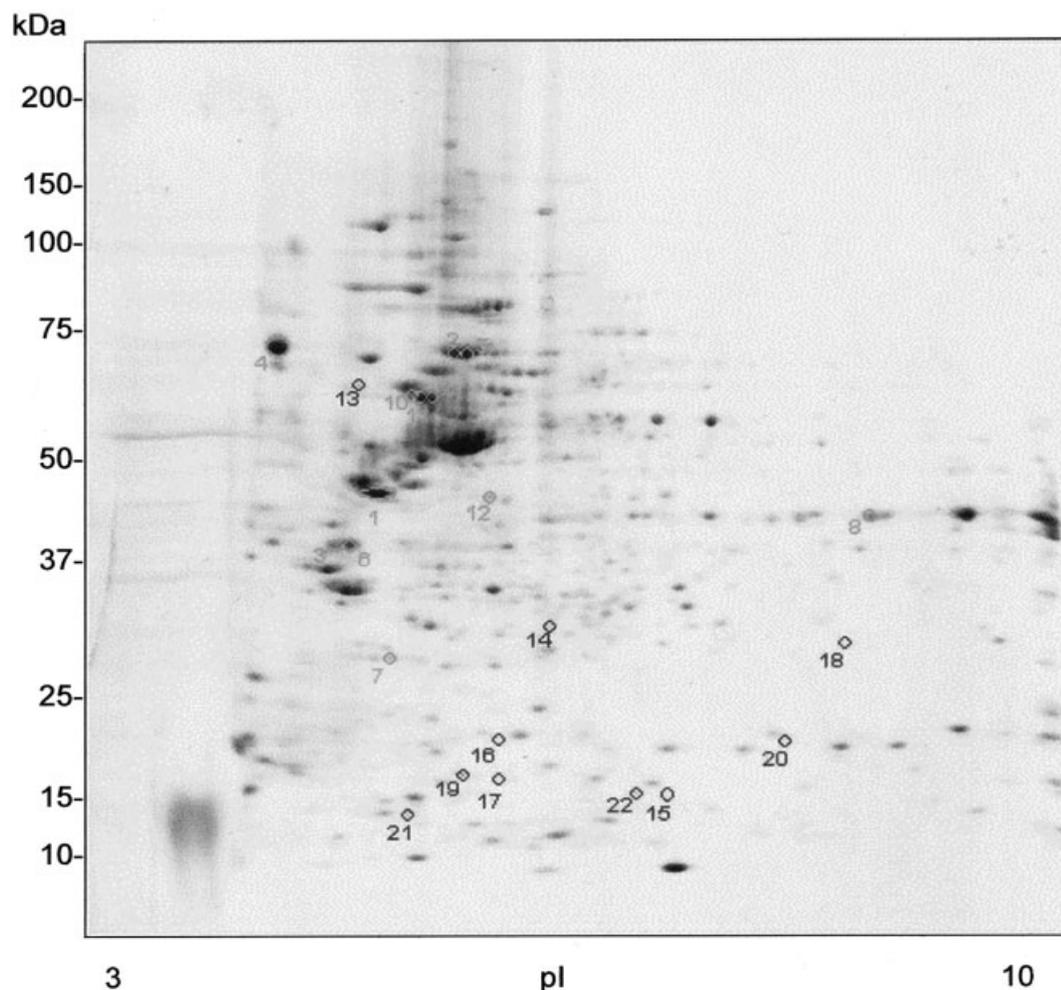


Figure 3. Two-dimensional master map of Paca44 cells in the pH 3–10 IPG interval. The identified spots are numbered as follows: 1–12, proteins down-regulated; 13–22, proteins up-regulated. All the numbered spots have been identified by peptide fingerprinting and MS analysis (see Table 1 for protein classification).

Table 1. Identified proteins from the Paca44 cell line 2-D gel

Spot No.	2-D Gel		Databank				Protein name	Accession No.	% Coverage	No. of peptides	Variation
	Exp. M_r (Da)	Exp. pI	Theor. M_r	Theor. pI	Z-Score	MOWSE-Score					
1	~ 40 000	~ 5.2	32 554	4.7	2.03	4.79 E12	Nucleophosmin (NPM) (Nucleolar phospho-protein B23) (Numatrin)	P06748	55.1	18	decreased 3
2	~ 72 000	~ 5.7	57 926	5.3	2.4	6.77 E31	60 kDa heat shock protein, mitochondrial precursor (Hsp60) (60 kDa chaperonin)	P10809	69.4	42	decreased 2
3	~ 35 000	~ 4.5	24 504	4.7	2.37	8.28 E12	Tropomyosin α 4 chain (Tropomyosin 4)	P07226	73.4	25	decreased 2
4	~ 75 000	~ 4.2	46 466	4.3	2.38	1.81 E19	Calreticulin precursor (CRP55) (Calregulin)	P27797	62	25	decreased 2

Table 1. Continued

Spot No.	2-D Gel		Databank								
	Exp. M_r (Da)	Exp. <i>pI</i>	Theor. M_r	Theor. <i>pI</i>	Z-Score	MOWSE-Score	Protein name	Accession No.	% Coverage	No. of peptides	Variation
5	~ 75 000	~ 6	57 926	5.3	2.38	1.46 E27	60 kDa heat shock protein, mitochondrial precursor (Hsp60) (60 kDa chaperonin)	P10809	60	36	decreased 2
6	~ 35 000	~ 4.9	32 798	4.7	2.36	4.83 E14	Tropomyosin α 3 chain (Tropomyosin 3)	P12324	65	31	decreased 3
7	~ 26 000	~ 5.4	19 582	4.9	2.41	1,09 E9	Translationally controlled tumor protein (TCTP)	P13693	47	14	decreased 3
8	~ 37 000	~ 9	37 406; 35 899	9.0; 8.6	2.4	5.69 E10; 1.17 E10	Heterogeneous nuclear ribonucleoproteins A2/B1; glyceraldehyde 3-phosphate dehydrogenase, liver (EC 1.2.1.12)	P22626; P04406	54; 48	16; 16	decreased 2
9	~ 55 000	~ 5.6	51 736; 46 142	5.0; 5.0	2.35	3.38 E12; 4.19 E9	ATP synthase β -chain, mitochondrial precursor (EC 3.6.3.14); protein disulfide isomerase A6 precursor (EC 5.3.4.1)	P06576; Q15084	41; 41	16; 13	decreased 3
10	~ 55 000	~ 5.4	51 736; 49 671	5.0; 4.8	2.32	4.24 E11; 1.20 E10	ATP synthase β -chain, mitochondrial precursor (EC 3.6.3.14); tubulin β -1 chain	P06576; P07437	38; 42	15; 14	decreased 2
11	~ 55 000	~ 5.5	51 736	5	2.41	1.63 E15	ATP synthase β -chain, mitochondrial precursor (EC 3.6.3.14)	P06576	53	24	decreased 2
12	~ 37 000	~ 5.5	34 274	5.9	2.23	2.57 E9	60S acidic ribosomal protein P0	P05388	47	16	decreased 2
13	~ 60 000	~ 4.6	47 626	4.7	2.35	1.24 E5	Chromatin assembly factor 1 subunit C	Q09028	20	9	increased 2
14	~ 30 000	~ 6.5	25 190	6.2	2.43	6.67 E7	Growth factor receptor-bound protein 2	P29354	43.8	16	increased 4
15	~ 18 000	~ 7	10 606	6.3	1.72	1.26 E5	Cytochrome <i>c</i> oxidase polypeptide Vb	P10606	40	9	increased 5
16	~ 20 000	~ 6.2	16 310; 17 160	5.5; 5.7	1.91; 2.01	2.22 E9; 5.89 E7	ARP2/3 complex 16 kDa subunit (P16-ARC); stathmin (phosphoprotein p19) (pp19) (oncoprotein 18)	Q15511; P16949	86; 77	14; 24	increased 8
17	~ 18 000	~ 6	14 276	5.7	1.77	1.54 E5	Programmed cell death protein 5 (TFAR19 protein)	O14737	60.8	8	increased 4
18	~ 26 000	~ 8.5	22 096	8.2	2.38	6.57 E8	Peroxisomal protein 1 (EC 1.11.1.-)	Q06830	68	15	increased 2
19	~ 18 000	~ 5.8	16 040	5.5	2.34		Deduced protein product shows significant homology to coactosin	NCBIInr. AAABB022.1	57	17	increased 2
20	~ 20 000	~ 8	16 363	8.5	1.98	1,92 E9	UEV protein (ubiquitin-conjugating E2 enzyme variant)	NCBIInr. AAH28673	75	13	increased 2
21	~ 13 000	~ 5	11 606	4.8	2.35	2.33 E7	Thioredoxin	P10599	69	10	increased 2
22	~ 13 500	~ 6.5	13 654	6.4	2.38	1.02 E6	Hint protein	P49773	59	6	increased 3

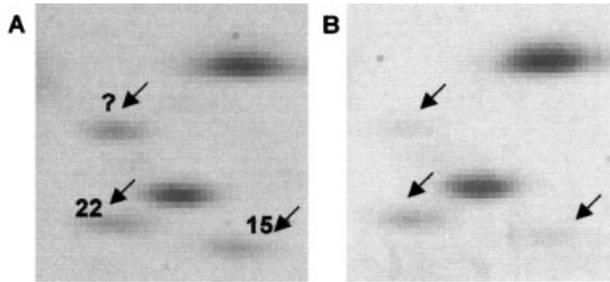


Figure 4. Comparison of 2-D gel patterns of some proteins in Paca44 cells (A) treated with TSA and (B) in Paca44 control. The corresponding spot numbers are shown in Table 1; the spot with a question mark corresponds to a 10-fold upregulated protein which could not be identified in any of the databases available, although it gave good MS spectra.

The available software packages, although labor-intensive, produce reliable results, widely accepted by the scientific community. To date, only two studies have used a proteomic approach in the field of pancreatic diseases [31, 32]. The first one [31] addressed a technical issue and was devoted to preparation of pancreatic juice for 2-D map analysis. The second one [32] reported the effect of the cytotoxic agent daunorubicin on protein expression of pancreatic cancer cells.

In the present study we show that TSA is able to inhibit cell proliferation of the pancreatic adenocarcinoma cell line Paca44 by cell growth arrest and apoptosis. We attempted to address the molecular basis of this effect by the study of protein expression profiles of cells before and after TSA treatment. This is the first paper addressing the issue of gene expression variation after HDAC inhibition, which is associated with cell cycle arrest and apoptotic cell death. Among the 22 proteins which were identified by MS analysis, of particular interest are the two downregulated proteins nucleophosmin and TCTP, as well as the upregulated proteins PDCD5 and stathmin. Their role will be briefly discussed below.

Nucleophosmin (NPM), a 3-fold downregulated 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 tumor-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]. Interestingly, 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].

Translationally controlled tumor protein (TCTP), a 3-fold downregulated polypeptide, seems to be involved in tumor reversion, *i.e.*, in the process by which some cancer cells lose their malignant phenotype. In a recent study, Tuynder *et al.* [22] have shown that TCTP is strongly downregulated in the reversion process of human leukemia and breast cancer cell lines. They hypothesized that tumor reversion is a biological process in its own right, involving a cellular reprogramming mechanism, overriding genetic changes in cancer, which triggers an alternative pathway leading to suppression of tumorigenicity. Recently, TCTP was also found, for the first time, in rat and human testes by Guillame *et al.* [23]. Interestingly, the mRNA of TCTP was recently reported to be overexpressed in human colon cancer by Chung *et al.* [24].

The programmed cell death protein 5 (PDCD5, also designated as TFAR19), here found to be upregulated by a factor of 4, is a recently discovered protein involved in the regulation of cell apoptosis [25–27]. The level of this protein in cells undergoing apoptosis is significantly increased compared with normal cells. Thus, its upregulation in TSA-treated cell lines, as here reported, is consistent with our finding of apoptotic cell death of Paca44 by TSA treatment.

Stathmin (oncoprotein 18, OC18) was upregulated by the TSA treatment by a factor of 8. Stathmin is a p53-regulated member of a novel class of microtubule-destabilizing proteins known to promote microtubule depolymerization during interphase and late mitosis [28]. Thus, high levels of stathmin could induce growth arrest at the G2 to mitotic boundary [29, 30]. This again is highly consistent with the observation that Paca44 showed a cell cycle arrest at the G2 phase. It is of interest to note that overexpression of stathmin *via* its effect of inhibiting polymerization of microtubules permits increased binding to these structures of vinblastin, a well-known chemotherapeutic agent, during treatment of human breast cancer [28]. Due to its effect of inhibiting cell proliferation *via* a mitotic block, the upregulation of stathmin here reported appears to be consistent with the anti-tumoral activity of TSA.

Work is in progress to extend the present data along the following lines: (i) explore the effect on the protein expression profile of combined drug treatments; (ii) follow the kinetics of drug exposure, in order to analyze the timing of the up- and downregulation profiles, and (iii) correlate

the protein expression with the mRNA levels for such altered proteins. The latter issue is of particular interest in order to clarify whether the process of TSA protein regulation occurs at transcriptional or post-translational level [32].

Supported by grants from AIRC (Associazione Italiana per la Ricerca sul Cancro) and from MIUR (No. 2001068593). PGR is additionally supported by grants from ASI (IIR/294/02) and by the European Community (grant No. QLG2-CT-2001-01903).

Received February 20, 2003

5 References

- [1] Almoguera, C., Shibata, D., Forrester, K., Martin, J., Arnheim, N., Perucho, M., *Cell* 1988, 53, 549–554.
- [2] Lemoine, N. R., Jain, S., Hughes, C. M., Staddon, S. L., Maillet, B., Hall, P. A., Kloppel, G., *Gastroenterology* 1992, 102, 230–236.
- [3] Scarpa, A., Capelli, P., Mukai, K., Zamboni, G., Oda, T., Iacono, C., Hirohashi, S., *Am. J. Pathol.* 1993, 142, 1534–1543.
- [4] Caldas, C., Hahn, S. A., Costa, L. T. D., Redston, M. S., Schutte, M., Seymour, A. B., Weinstein, C. L., Hruban, R. H., Yeo, C. J., Kern, S. E., *Nature Genet.* 1994, 8, 27–32.
- [5] Moore, P., Orlandini, S., Zamboni, G., Capelli, P., Rigaud, G., Falconi, M., Bassi, C., Lemoine, N., Scarpa, A., *Brit. J. Cancer* 2001, 84, 253–262.
- [6] Aparicio, A., Weber, J. S., *Curr. Opin. Investig. Drugs* 2002, 3, 627–633.
- [7] Kelly, W. K., O'Connor, O. A., Marks, P. A., *Expert Opin. Investig. Drugs* 2002, 11, 1695–1713.
- [8] Kouzarides, T., *Curr. Opin. Genet. Dev.* 1999, 9, 40–48.
- [9] Archer, S. Y., Hodin, R. A., *Curr. Opin. Genet. Dev.* 1999, 9, 171–174.
- [10] Tsukiyama, T., Wu, C., *Curr. Opin. Genet. Dev.* 1997, 7, 182–191.
- [11] Gregory, P. D., Horz, W., *Eur. J. Biochem.* 1998, 251, 9–18.
- [12] Jacobson, S., Pillus, L., *Curr. Opin. Genet. Dev.* 1999, 9, 175–184.
- [13] Marks, P. A., Richon, V. M., Rifkind, R. A., *J. Natl. Cancer Inst.* 2000, 92, 1210–1216.
- [14] Righetti, P. G., Stoyanov, A. V., Zhukov, M. Y., *The Proteome Revisited*, Elsevier, Amsterdam 2001.
- [15] James, P. (Ed.), *Proteome Research: Mass Spectrometry*, Springer, Berlin 2001.
- [16] Hamdan, M., Righetti, P. G., *Mass Spectrom. Rev.* 2002, 21, 287–302.
- [17] Colombo, E., Marine, J. C., Danovi, D., Falini, B., Pelicci, P. G., *Nat. Cell Biol.* 2002, 4, 529–533.
- [18] Okuda, M., *Oncogene* 2002, 21, 6170–6174.
- [19] Yang, C., Maignel, D. A., Carrier, F., *Nucleic Acids Res.* 2002, 30, 2251–2260.
- [20] Shackelford, G. M., Ganguly, A., MacArthur, C. A., *BMC Genomics* 2001, 2, 8–9.
- [21] Bischof, D., Pulford, K., Mason, D. Y., Morris, S. W., *Mol. Cell. Biol.* 1997, 17, 2312–2325.
- [22] Tuynnder, M., Susini, L., Prieur, S., Besse, S., Fiucci, G., Amson, R., Teleman, A., *Proc. Natl. Acad. Sci. USA* 2002, 99, 14976–14981.
- [23] Guillame, E., Pineau, C., Evrad, B., Dupaix, A., Moertz, E., Sanchez, J. C., Hochstrasser, D. F., Jegou, B., *Proteomics* 2001, 1, 880–889.
- [24] Chung, S., Kim, M., Choi, W., Chung, J., Lee, K., *Cancer Lett.* 2000, 156, 185–190.
- [25] Liu, H., Wang, Y., Zhang, Y., Song, Q., Di, C., Chen, G., Tang, J., Ma, D., *Biochem. Biophys. Res. Commun.* 1999, 254, 203–210.
- [26] Chen, Y., Sun, R., Han, W., Zhang, Y., Song, Q., Di, C., Ma, D., *FEBS Lett.* 2001, 509, 191–196.
- [27] Rui, M., Chen, Y., Zhang, Y., Ma, D., *Life Sci.* 2002, 71, 1771–1778.
- [28] Alli, E., Vash-Babula, J., Yang, J. M., Hait, W. N., *Cancer Res.* 2002, 62, 6864–6869.
- [29] Cassimeris, L., *Curr. Opin. Cell Biol.* 2002, 14, 18–24.
- [30] Mistry, S. J., Atweh, G. F., *Mt. Sinai J. Med.* 2002, 69, 299–304.
- [31] Wandschneider, S., Fehring, V., Jacobs-Emeis, S., Thiesen, H. J., Löhr, M., *Electrophoresis* 2001, 22, 4383–4390.
- [32] Möller, A., Soldan, M., Volker, U., Maser, E., *Toxicology* 2001, 160, 129–138.

Daniela Cecconi^{1*}
 Hubert Astner^{1*}
 Massimo Donadelli^{2*}
 Marta Palmieri²
 Edoardo Missiaglia³
 Mahmoud Hamdan⁴
 Aldo Scarpa³
 Pier Giorgio Righetti¹

¹Department of Agricultural
 and Industrial Biotechnologies

²Department of Neurological
 and Visual Sciences,
 Section of Biochemistry

³Department of Pathology,
 Section of Anatomical Pathology,
 University of Verona,

⁴Computational, Analytical &
 Structural Sciences,
 GlaxoSmithKline,
 Verona, Italy

Proteomic analysis of pancreatic ductal carcinoma cells treated with 5-aza-2'-deoxycytidine

A pancreatic adenocarcinoma cell line (PaCa44), which contains, among other alterations, a methylated p16 promoter, was treated with a chemotherapeutic agent, 5-aza-2'-deoxycytidine (DAC), in order to evaluate the effect of this drug on cell growth and protein expression. Cell proliferation was strongly inhibited by a 24 h DAC treatment and this inhibition lasted for at least 10 days. Master maps of control and treated PaCa44 cells were generated by analysis with the PDQuest software. The comparison between such maps showed up- and downregulation of 45 polypeptide chains, of which 32 were downregulated and 13 upregulated, out of a total of 700 spots detected by a medium-sensitivity stain, micellar Coomassie Brilliant Blue. Fingerprinting by mass spectrometry analysis enabled the identification of 36 of these spots. Among the major changes in DAC-treated cells: cofilin and profilin 1 are silenced; coactosin, peptidyl-propyl *cis-trans* isomerase A and cystatin B are decreased by 22-, 16- and 15-fold, respectively; stress-70 protein, superoxide dismutase and protein disulfide isomerase A3 are increased by 13-, 11-, and 5-fold, respectively. The significance of some of these major changes is discussed.

Keywords: Ductal carcinomas / Pancreatic tumors / Proteomics DOI 10.1002/elps.200305724

1 Introduction

Pancreatic cancer is an aggressive disease and is the fifth cause of cancer-related mortality in Western countries. Around 90% of pancreatic neoplasms are thought to arise from ductal cells and are generally referred to as pancreatic ductal adenocarcinoma (PDA). In recent years, PDA has become one of the best characterized human malignancies from a molecular standpoint and is in fact characterized by a relatively unique molecular fingerprint constituted by frequent alterations of the *K-ras* [1, 2], *p53* [3], *p16INK4a* [4, 5], and *DPC4/Smad4* genes [5]. However, improved understanding of the underlying anomalies characterizing PDA is the only means to provide new markers for earlier diagnosis and to identify potential targets for therapeutic intervention.

The importance of epigenetic events in human cancer, including histone acetylation [6, 7] and DNA methylation [8], has become increasingly apparent. Histone acetylation levels depend on the activity of the two counteracting enzymes histone acetyltransferases (HATs) and histone

deacetylases (HDACs) [9, 10]. Acetylation is thought to separate the basic *N*-termini of histones from DNA, which then become more accessible to transcription factors. Thus, histone acetylation leads to gene activation, while histone deacetylation leads to gene repression. It has been demonstrated that HDAC inhibitors, such as trichostatin A (TSA), lead to an increase in histone acetylation and induce an enhancement of the expression of specific genes that elicit growth arrest, differentiation, and apoptosis [11, 12]. We have recently explored the effects of TSA on gene expression in pancreatic cancer cells by proteomic profiling with 2-D maps, showing dysregulation of 51 proteins. The modulation of most of these proteins was consistent with the observation that TSA inhibited cell growth by cell cycle arrest at the G2 phase and apoptotic cell death [13].

Methylation of DNA occurs predominantly at the cytosine of CpG dinucleotides by DNA methyl transferase (DNMT) activities and is generally responsible for transcriptional repression. Recently, a link has been established between DNA methylation and histone deacetylation by studies showing that DNMT1 and MeCPs (methyl-CpG-binding proteins) interact physically with histone deacetylases [14–16] and that repression by methylated CpGs is partially relieved by TSA. Many tumor suppressor genes contain abnormal methylation of CpG islands in their promoters and appear to be silenced at the transcriptional level.

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: DAC, 5-aza-2'-deoxycytidine; DNMT, DNA methyl transferase; PDA, pancreatic ductal adenocarcinoma

* These authors contributed equally to this work.

To date, most of the studies investigating the induction of gene expression by the DNMT1 inhibitor 5-aza-2'-deoxycytidine (DAC) in pancreatic cancer have focused on the reactivation of few genes, such as *p16*, *DAP kinase*, and *RARbeta* [17]. The tumor suppressor gene *p16* is silenced by promoter methylation in up to 20% of both xenografted [18, 19] and primary pancreatic cancers [6]. However, the identification of other genes selectively hypermethylated in this cancer may be of great importance for diagnostic purpose and for further understanding the biology of this tumor.

In the present study, we have investigated the growth susceptibility of the human pancreatic cancer cell line PaCa44 to DAC and determined the effect of this drug on protein expression, *via* proteomic profiling, as implemented by 2-D map analysis [20].

2 Materials and methods

2.1 Chemicals and materials

Urea, thiourea, CHAPS, iodoacetamide (IAA), tributylphosphine (TBP), and sodium dodecyl sulfate (SDS) were obtained from Fluka Chemie (Buchs, Switzerland). Bromophenol blue and agarose were from Pharmacia-LKB (Uppsala, Sweden). Acrylamide, *N,N'*-methylenebisacrylamide, ammonium persulfate, TEMED, the Protean IEF Cell, the GS-710 Densitometer as well as the linear Immobiline dry strips pH gradient 3–10 (17 cm long) were from Bio-Rad Laboratories (Hercules, CA, USA). Ethanol, methanol, and acetic acid were from Merck (Darmstadt, Germany). DAC was obtained from Sigma-Aldrich (St. Louis, MO, USA). A 10 mM solution of DAC in DMSO was prepared and stored at -80°C until use.

2.2 Cell treatment with DAC

2.2.1 Cells and growth conditions

PaCa44 human pancreatic adenocarcinoma cells were grown in RPMI 1640 supplemented with 20 mM glutamine and 10% fetal bovine serum (FBS; BioWhittaker, Italy) and were incubated at 37°C with 5% CO_2 . Cells were seeded at a density of $12 \times 10^3/\text{cm}^2$.

2.2.2 Cell proliferation assay

Cells were plated in 96-well cell culture plates at subconfluency (2×10^3 cells/well). Cells were treated with 2.5 μM DAC for 24 h and cell proliferation was evaluated from 3 to 17 days after the beginning of the treatment. Cells

were stained with Crystal Violet (Sigma), solubilized in PBS with 1% SDS, and measured photometrically ($A_{595\text{nm}}$) to determine cell viability. The crystal violet staining procedure is a simple and reproducible assay of cytotoxicity based on the growth rate reduction reflected by the colorimetric determination of the stained cells [47]. Three independent experiments were performed.

2.2.3 Treatment of cells for proteomic analysis and cell lysis

Preliminary results using different concentrations of DAC showed that a 2.5 μM DAC treatment for 24 h was able to induce a 50% arrest of cell growth 3 days after beginning of the treatment, and that maximal demethylation of DNA is obtained 6 days after treatment. Thus, we have chosen 2.5 μM concentration and 6 days after beginning of treatment as a good compromise to obtain the best cytotoxic effect and a reasonable amount of material to perform proteomic analysis. Protein extraction from cells untreated and treated with 2.5 μM DAC for 24 h and harvested after 6 days from the beginning of the treatment was performed with lysis buffer (40 mM Tris, 1% NP-40, 1 mM Na_3VO_4 , 1 mM NaF, 1 mM PMSF, protease inhibitor cocktail). Cells were left in lysis buffer for 30 min on ice. After centrifugation at $14\,000 \times g$ at 4°C , for removal of particulate material, the protein solution was collected and stored at -80°C until use.

2.3 Two-dimensional gel electrophoresis

Seventeen cm long, pH 3–10 immobilized pH gradient strips (IPG; Bio-Rad Labs., Hercules, CA, USA) were rehydrated for 8 h with 450 μL of 2-D solubilizing solution (7 M urea, 2 M thiourea, 5 mM TBP, 40 mM Tris, and 20 mM IAA) containing 2 mg/mL of total protein from PaCa 44 cells [21]. Isoelectric focusing (IEF) was carried out with a Protean IEF Cell (Bio-Rad, Hercules, CA, USA), 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 hours \times voltage applied was 70 000 Vh for each strip and the temperature was set at 20°C . For the second dimension, the IPGs strips were equilibrated for 26 min by rocking in a solution of 6 M urea, 2% SDS, 20% glycerol, 375 mM Tris-HCl, pH 8.8. The IPG strips were then laid on an 8–18%T gradient SDS-PAGE with 0.5% agarose in the cathode buffer (192 mM glycine, 0.1% SDS, and Tris to pH 8.3). The anodic buffer was a solution of 375 mM Tris HCl, pH 8.8. The electrophoretic run was performed by setting a current of 2 mA for each gel for 2 h, then 5 mA/gel for 1 h, 10 mA/gel for 20 h, and 20 mA/gel until the end of the run. During the whole run the temperature was set at 11°C .

Gels were stained overnight with colloidal Coomassie blue (0.1% Coomassie Brilliant Blue G, 34% v/v methanol, 3% v/v phosphoric acid, and 17% w/v ammonium sulfate), while destaining was performed with a solution of 5% acetic acid until a clear background was achieved. Five replicas for each condition (control and DAC-treated cells) were made. In addition, the same experiments were repeated twice.

2.4 Protein pattern analysis

The 2-DE gels were scanned with a GS-710 densitometer (Bio-Rad), and analyzed with the software PDQuest Version 7.1 (Bio-Rad). A match set was created from the protein patterns of the two independent cellular extracts (control cell line, DAC-treated cell line). A standard gel was generated out of the image with the highest spot number. Spot quantities of all gels were normalized to remove nonexpression-related variations in spot intensity, so 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 optical density (OD). Statistical analysis of PDQuest allowed the study of proteins that were significantly increased or decreased in DAC-treated cell line. Forty-five spots were found to be differently expressed in DAC-treated PaCa44, of which 13 spots were upregulated (with a significance level $\alpha = 0.05$) and 32 spots downregulated (with $\alpha = 0.05$).

2.5 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, covered with 15 μ L 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. The extract was brought to dryness in a Speedvac and then resuspended with 10 μ L of a H_2O 0.1% TFA solution. The extracted peptides were further purified by ZipTip pipetting tips (Millipore, Bedford, MA, USA). The obtained solutions 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 α -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, TrEMBL and NCBI nr databases by utilizing the ProteinLynx program from Micromass, Profound from Prowl, and Mascot from Matrix Science.

3 Results

The effect of the DNMT1 inhibitor DAC on the proliferation of the PaCa44 pancreatic adenocarcinoma cell line was examined by measuring cell viability with a colorimetric assay as described in Section 2.2. Treatment with 2.5 μ M DAC for 24 h inhibited cell proliferation for at least 10 days (Fig. 1A). Microscopic examination of the cells at the sixth day after treatment showed important morphological alterations such as distinctive polynucleated cells, larger dimension, and heterogeneous shape which may suggest an extensive change of the gene expression profile (Figs. 1B, C).

A standard 2-D map of the total protein extract from PaCa44 cell line, stained with colloidal Coomassie blue, is shown in Fig. 2. About 700 polypeptide spots were revealed in the pH 3–10 interval with this medium-sensitivity stain. After matching the master map of the control PaCa44 cells with the master map of cells at 6 days after DAC treatment, a total of 45 polypeptide chains were found to be differentially expressed: 32 were downregulated and 13 upregulated.

All these spots were eluted and submitted to mass spectrometry analysis as described in Section 2.5. Thirty-six proteins could be identified by mass spectrometric analysis (Fig. 3). Of these, 2 were silenced, 22 were downregulated, and the remaining 12 were upregulated. Their identification, experimental and theoretically predicted pI , and M_r values are given in Table 1. The histogram in Fig. 4 displays the total changes detected of both downregulated and upregulated polypeptides. Figure 5 gives an example of a few spots exhibiting marked changes, such as spot No. 8201 (cofilin) that is silenced by DAC treatment. DAC treatment caused up- or downregulation in a number of proteins. The classification of these proteins according to the different biological processes in which they are involved is detailed in Table 2 and summarized in Fig. 6. It is of interest that most of the proteins which appear to be strongly influenced by DAC treatment are mostly involved in metabolism, cell growth and cell communication.

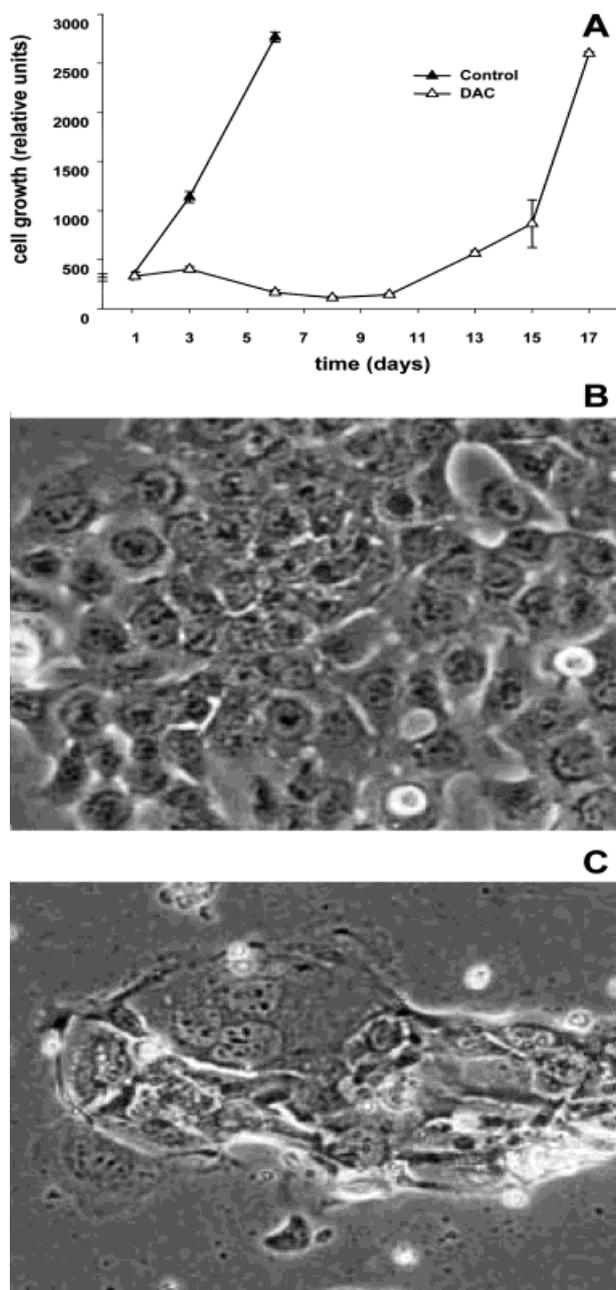


Figure 1. Effect of DAC on growth of the pancreatic cancer cell line PaCa44. (A) Cells were seeded in 96-well plates at a density of 2×10^3 cells/well in the absence (control) or presence of $2.5 \mu\text{M}$ DAC for 24 h. Cells were then cultured for 17 days and the culture medium was changed three times a week. At various times, cell viability was measured photometrically as described in the Section 2. Values are means of quadruplicate wells for each time point. The relative units of cell growth, in ordinate, correspond to the optical density at $595 \text{ nm} \times 1000$. (B), (C) Changes induced in cell morphology by treatment with DAC after 6 days. Photomicrographs of unstained PaCa44 cells were performed with Olympus-IX50, 10X-20X. (B) Untreated cells; (C) DAC-treated cells.

4 Discussion

Differential analysis of protein expression is increasingly being used for profiling a number of pathological states, including cancer. The most widely accepted techniques are still 2-D maps, with charge and mass coordinates [20], followed by spot excision and mass spectrometry identification of tryptic digests [22]. Although differential profiling can be performed by a vast number of techniques (for a review, see [23]), in conventional 2-D maps the preferred method is still statistical analysis performed on master maps created from two independent cellular extracts (control vs. pathological or treated cells). The available software packages, although labor-intensive, produce reliable results, widely accepted by the scientific community. In the present study we showed that DAC inhibits cell proliferation of the pancreatic adenocarcinoma cell line PaCa44, and attempted to address the molecular basis of this effect by the study of protein expression profiles of cells before and after DAC treatment. The significance of some of our findings is discussed below.

4.1 Silenced proteins

The proteomic analysis showed that 32 proteins were downregulated at 6 days after DAC treatment. We speculate that at this time indirect effects, such as induction of repressors, may cause reduction in protein expression. Two proteins were found to be silenced upon DAC treatment: cofilin and profilin 1 (Table 1). Cofilin is a widely distributed actin-modulating protein that has the ability to bind along the side of F (filamentous)-actin and to depolymerize it in a pH-dependent manner. Our data appear to be in agreement with those of Kanai *et al.* [24], who have recently reported strongly reduced mRNA expression for cofilin in DAC-treated human stomach and colorectal cancer cells. Of interest, Sinha *et al.* [25] reported that cofilin was overexpressed in pancreatic adenocarcinoma cell lines, which may suggest that suppression of cofilin might lead to cancer regression. Profilin 1 is a small actin-binding protein that is involved in diverse functions, such as maintaining cell structure integrity, cell motility, growth factor signal transduction, and metastasis [26, 27]. The suppression of this protein in DAC-treated PaCa44 cells is consistent with the role of this drug in inhibiting tumor cell growth.

4.2 Downregulated proteins

This is the largest set of modulated species, since it accounts for 32 polypeptides out of a total of 45 differentially expressed. Among these, six appear to be of particular relevance: coactosin-like protein (CLP), peptidyl-pro-

Table 1. Identified proteins from the Paca44 cell line 2-D gel

Spot ID	Protein	Gene name	SWISS-PROT accession number	MW (expt/pred)	pI (expt/pred)	No. Peaks matched (% coverage)	MAS-COT score	Z-Score	Trend in DAC-treated cell	Observations and reported functions
8201	Cofilin	CFL1	P23528	18710/25000	8.5/9	10 (49)	89	2.37	OFF	Tumor-related protein, whose expression is affected by DNA methylation; actin-modulating protein
6118	Profilin I	PFN1	P07737	14923/18000	8.7/8	8 (46)	68	2.39	OFF	Actin-binding protein involved in maintaining cell integrity, cell mobility, and tumor cell metastasis
4209	Coactosin-like protein	COTL1	Q14019	15935/20000	5.54/5	7 (45)	102	2.31	Decreased 22-fold	Human pancreatic cancer antigen; actin-binding protein
6211	Peptidyl-propyl <i>cis-trans</i> isomerase A	PPIA	P05092	17881/22000	8.2/6.5	16 (50)	94	2.38	Decreased 16-fold	Novel human hepatocellular carcinoma marker; overexpressed in metastatic tumors
6117	Cystatin B	CSTB	P04080	11140/15000	7.7/7	7 (61)	110	2.3	Decreased 15-fold	Inhibitors of lysosomal cysteine proteinases (cathepsins); involved in tumor cell invasion and metastasis
6210	No date in database								Decreased 10-fold	
8101	No date in databases								Decreased 9-fold	
3525	Rho GDP-dissociation inhibitor 2	ARHGDIB	P52566	22988/50000	5.1/4.5	14 (61)	115	2.17	Decreased 9-fold	Inhibitor for the Ras-related Rho family GTPases; decreased in drug-induced apoptosis by proteolysis mediated by caspase-3
7246	Peptidyl-propyl <i>cis-trans</i> isomerase A	PPIA	P05092	17881/22000	8.2/8.6	19 (65)	139	2.38	Decreased 9-fold	Novel human hepatocellular carcinoma marker; overexpressed in metastatic tumors
9109	FK508-binding protein	FKBP1A	P20071	11820/15000	8.5/8.9	7 (40)	77	1.7	Decreased 6-fold	Physiologic regulator of the cell cycle, cell from mice knock-out manifest cell cycle arrest in G1 phase
6205	Peptidyl-propyl <i>cis-trans</i> isomerase A	PPIA	P05092	17881/22000	8.2/8.4	18(67)	96	2.32	Decreased 5-fold	Novel human hepatocellular carcinoma marker; overexpressed in metastatic tumors
4308	Rho GDP-dissociation inhibitor 2	ARHGDIB	P52566	22988/32000	5.1/4.9	8 (40)	82	1.68	Decreased 5-fold	Inhibitor for the Ras-related Rho family GRPases; decreased in drug-induced apoptosis by proteolysis mediated by caspase-3

Table 1. Continued

Spot ID	Protein	Gene name	SWISS-PROT accession number	MW (expt/pred)	pI (expt/pred)	No. Peaks matched (% coverage)	MAS-COT score	Z-Score	Trend in DAC-treated cell	Observations and reported functions
3318	Rho GDP-dissociation inhibitor 1	ARHGDI1	P52565	23207/32000	5/4.6	10 (32)	82	1.67	Decreased 4-fold	Inhibitor for the Ras-related Rho family GTPases; decreased in drug-induced apoptosis by proteolysis mediated by caspase-3
4104	Tubulin-specific chaperone A	TBCA	075347	12716/16000	5.25/5	11 (61)	76	2.42	Decreased 4-fold	Involved in dimerization of the tubulin subunits, in degradation of defective tubulin subunits and in regulating tubulin polymerization
4404	Proteasome activator complex subunit 2. Chloride intracellular channel protein	PSME2 CLIC1	Q9UL46 000299	27362/37000 26923/37000	5.5/5.4 5.1/5.4	18 (59) 9 (43)	181		Decreased 4-fold	Modulates the proteasome-catalyzed production of antigenic peptides presented to the immune system on MHC class I molecules, chloride channel in nuclear membrane
5203	Stathmin (phosphoprotein p19) Oncoprotein 18	STMN1	P16949	17171/24000	6/5.5	10 (49)	86	2.06	Decreased 4-fold	Microtubule-destabilizing proteins upregulated in neoplastic cells; downregulation in malignant cells interferes with their progression through cell cycle and abrogates their transformed phenotype
7247	Peptidyl-propyl <i>cis-trans</i> isomerase A	PPIA	P05092	17881/22000	8.2/8.6	18 (71)	105	2.32	Decreased 4-fold	Novel human hepatocellular carcinoma marker; overexpressed in metastatic tumors
6323	GTP-binding nuclear protein RAN	RAN	P17080	24423/32000	7.6/7.1	10 (48)	75	2.24	Decreased 3-fold	Involved in Cdc2/cyclin B activation and entry into mitosis when this process is coupled to the progression of S-phase
6311	Proteasome subunit beta type 2	PSMB2	P49721	22836/30000	7/6.5	10 (54)	63	1.82	Decreased 3-fold	Subunit of proteasome involved in an ATP/ubiquitin-dependent nonlysosomal proteolytic pathway
7248	Single-stranded DNA-binding Protein	SSBP1	Q04837	17260/20000	9.8/8.9	11 (70)	182	2.31	Decreased 2-fold	Participates in DNA replication in eukaryotic cells and possibly is an intracellular regulator of proliferation

Table 1. Continued

Spot ID	Protein	Gene name	SWISS-PROT accession number	MW (expt/pred)	pI (expt/pred)	No. Peaks matched (% coverage)	MAS-COT score	Z-Score	Trend in DAC-treated cell	Observations and reported functions
4201	60S ribosomal protein L12	MRPL12	P52815	16394/25000	9.3/4.6				Decreased 2-fold	Component of the ribosomal elongation factor binding domain
7438	Adenylate kinase isoenzyme 2	AK2	P54819	26347/36000	8.2/8.6	17 (59)	87	2.28	Decreased 2-fold	Proapoptotic protein released from the mitochondrial intermembrane space into the cytoplasm during apoptosis
6505	Annexin I	ANXA1	P04083	38528/40000	7/6.3	19 (54)	127	2.4	Decreased 2-fold	Calcium- and phospholipid-binding protein upregulated in pancreatic carcinoma cell lines, involved in cell proliferation, mitogenic signal transduction and metastasis
5301	Proteasome subunit beta type 4	PSMB4	P28070	29192/32000	5.9/5.5	11 (46)	79	2.33	Decreased 2-fold	Subunit of proteasome involved in an ATP/ubiquitin-dependent nonlysosomal proteolytic pathway
6203	Peptidyl-propyl <i>cis-trans</i> isomerase A	PPIA	P05092	17881/22000	8.2/8.4	18 (58)	67	2.26	Decreased 2-fold	Novel human hepatocellular carcinoma marker; overexpressed in metastatic tumors
4403	Annexin III	ANXA3	P12429	36375/38000	5.8/5.5	17 (52)	132	2.36	Decreased 2-fold	Acts like inositol 1,2-cyclic phosphate 2-phosphohydrolase
5720	Stress-70 Protein	HSPA9B	P38646	73681/80000	6.1/5.6	20 (28)	140	1.37	Increased 13-fold	Chaperone involved in cell proliferation, differentiation and tumorigenesis
6329	Superoxide dismutase	SOD2	P04179	24722/30000	8.6/7.6	9 (37)	100	2.38	Increased 11-fold	Tumor suppressor gene in human pancreatic cancer, overexpression may be effective in growth suppression of pancreatic cancer
6321	Superoxide dismutase	SOD2	P04179	24722/30000	8.6/7	10 (43)	107	2.19	Increased 6-fold	Tumor suppressor gene in human pancreatic cancer, overexpression may be effective in growth suppression of pancreatic cancer
5641	Protein disulfide isomerase A3	GRP58	P30101	56782/60000	6.3/5.8	30 (51)	246	2.41	Increased 5-fold	Chaperone in the endoplasmic reticulum lumen, may regulate signaling by sequestering inactive and activated Stat3

Table 1. Continued

Spot ID	Protein	Gene name	SWISS-PROT accession number	MW (expt/pred)	pI (expt/pred)	No. Peaks matched (% coverage)	MAS-COT score	Z-Score	Trend in DAC-treated cell	Observations and reported functions
4652	60 kDa heat shock protein	HSPD1	P10809	61055/65000	5.8/5.4	21 (37)	151	2.38	Increased 4-fold	Chaperone that accelerates the maturation of pro-caspase-3 by upstream activator proteases during apoptosis
6105	β -2-Microglobulin	B2M	P01884	13706/15000	6.1/6.5	10 (60)	76	2.35	Increased 4-fold	Induces caspase-dependent apoptosis and cell cycle arrest
5511	Actin	ACTB	P02570	41710/52000	5.3/5.6	14 (41)	105	2.38	Increased 3-fold	Involved in various types of cell motility
4803	150 kDa oxygen-regulated protein	HYOU1	Q9Y4L1	111336/160000	5.2/5	26 (25)	158	2.28	Increased 3-fold	Endoplasmic reticulum-associated protein: overexpression in pancreatic beta cells is correlated with insulin secretion
5712	Stress-70 protein	HSPA9B	P38646	73681/80000	6.1/5.5	18 (30)	153	2.4	Increased 2-fold	Chaperone involved in cell proliferation, differentiation and tumorigenesis
3316	SET Protein	SET	Q01105	32103/31000	4.1/4.6	9 (31)	97	2.37	Increased 2-fold	Regulates G(2)/M transition by modulating cyclin B-CDK1 activity; interacts in vivo and in vitro with the cell cycle inhibitor p21(Cip1)
5607	Protein disulfide isomerase A3	GRP58	P30101	56782/60000	6.3/5.5	20 (36)	170	2.36	Increased 2-fold	Chaperone in the endoplasmic reticulum lumen, may regulate signaling by sequestering inactive and activated Stat3
5504	Actin	ACTB	P02570	41710/52000	5.3/5.5	21 (51)	176	2.31	Increased 2-fold	Involved in various types of cell motility

lyl *cis-trans* isomerase A (PPIA), cystatin B, Rho GDP-dissociation inhibitor (Rho GDI-2), stathmin, and annexin 1. CLP (22-fold) is a human filamentous actin (F-actin) binding protein [28]. CLP binds to actin filaments with a stoichiometry of 1:2 (CLP:actin subunits). Additionally, it binds to 5-lipoxygenase in a 1:1 molar ratio [29]. A recent report by Nakatsura *et al.* [30] indicated CLP as a tumor-associated antigen, suggesting this protein as a candidate for a vaccine for immunotherapy of cancer patients. PPIA is thought to control mitosis by binding to cell cycle regulatory proteins and altering their activity. It appears that the gene coding for this protein is almost exclusively overexpressed in aggressive metastatic or chemoresistant tumours [31]. In addition, PPIA has been recently pro-

posed as a novel candidate marker for hepatocellular carcinoma [32]. Interestingly, all five isoforms of the PPIA are downregulated upon DAC treatment (16-fold for the main isoform, 9-, 5-, 4-, and 2-fold for four additional, minor isoforms, see Table 1).

Cystatin B (also called stefin B) belongs to the cystatin superfamily of cysteine protease inhibitors and target cysteine proteases, such as cathepsin B. Cystatin B has been implicated in malignant progression [33] and alterations in its expression, processing, and localization have been observed at various levels in malignant human tumors [34]. Progressively higher levels of cystatin B have been associated, *e.g.*, with short survival in patients

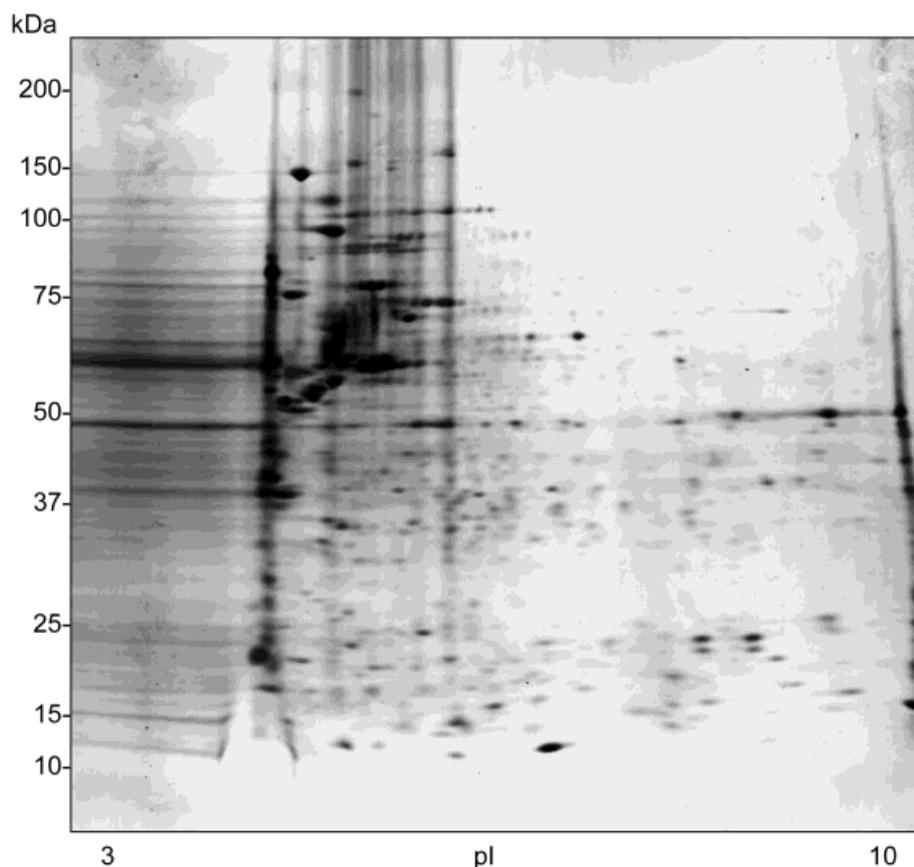


Figure 2. Master map of the pancreatic adenocarcinoma cell line Paca44, developed in the IPG 3–10 interval.

Table 2. Biological process ontology

Metabolism (42.86%)	Downregulated: ANXA1, FKBP1A, PPIA, RAN, SSBP1, PSMB2, MRPL12, PSMB4.	Upregulated: SET, HSPD1, GRP58, SOD2
Cell growth and/or maintenance (32.14%)	Downregulated: PFN1, CLIC1, SSBP1, STMN1, RAN, TBCA	Upregulated: SET, HSPD1, GRP58
Cell communication (25%)	Downregulated: ANXA1, CFL1, STMN1, RAN, ARHGDI, ARHGDI	Upregulated: GRP58
Response to external stimulus (21.43%)	Downregulated: ANXA1, PPIA, PSME2, ARHGDI	Upregulated: B2M, SOD2
Response to stress (14.29%)	Downregulated: ANXA1, PPIA	Upregulated: HYOU1, SOD2
Actin cytoskeleton reorganization (7.14%)	Downregulated: CFL1, ARHGDI	Upregulated: /
Cell motility (7.14%)	Downregulated: ANXA1	Upregulated: ACTB
Small molecule transport (3.57%)	Downregulated: CLIC1	Upregulated: /
Viral replication (3.57%)	Downregulated: PPIA	Upregulated: /
Oncogenesis (3.57%)	Downregulated: /	Upregulated: SET

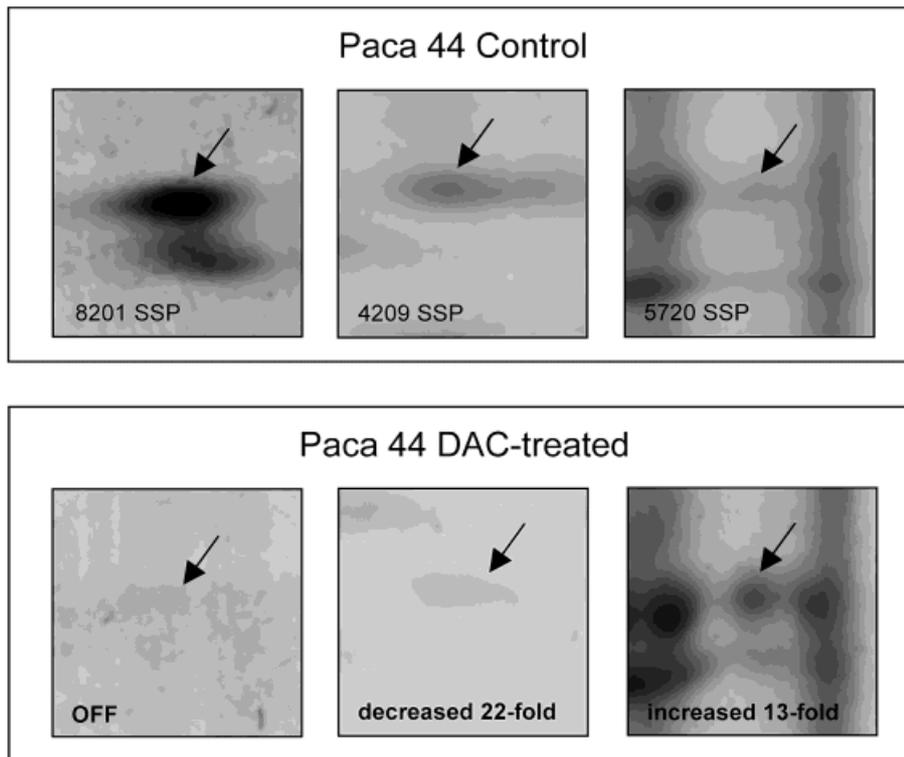


Figure 5. Comparison of 2-D gel patterns of some proteins in Paca44 cells treated with DAC (upper row) and in Paca44 control (lower row). The corresponding spot numbers are reported in Table 1.

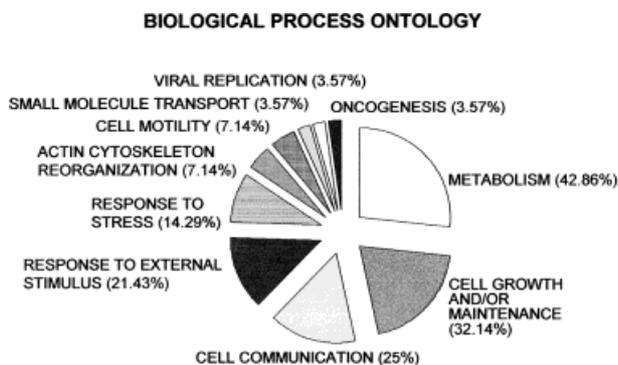


Figure 6. Classification of the proteins modulated by DAC treatment according to the biological function in which they are involved. Obtained by the program FatiGO (<http://fatigo.bioinfo.cnio.es>). Note that single proteins may belong to more categories (see also Table 2), which explains the total sum being substantially larger than 100%.

with colorectal cancer [35]. In human squamous cell carcinoma of the lung and in a number of other cancer types cystatin B has been found to be significantly increased as compared to normal tissues, and proved to be a prognostic factor [36]. This latter observation together with our finding that the level of cystatin B is markedly decreased (15-fold) after DAC treatment suggests a role for this protein in maintaining tumor cell growth.

All of the three Rho GDI-2 isoforms (9-fold for the main isoform, 5- and 4-fold for two additional isoforms) were downregulated by DAC. Rho GDI-2 is a cytosolic protein participating in the regulation of both the GDP/GTP cycle and the membrane association/dissociation cycle of Rho/Rac proteins. Rho GDI proteins are in general overexpressed in breast tumors and human bladder cancer and their increased synthesis correlates with malignancy [37, 38]. In agreement with our findings, Kovarova *et al.* [39] have reported, by 2-D map analysis followed by MALDI-MS, that Rho GDI proteins are significantly downregulated in CEM T-lymphoblastic leukemia cell lines after treatment with bohemin.

Stathmin (4-fold) is a member of a novel class of microtubule-destabilizing proteins that regulate the dynamics of microtubule polymerization and depolymerization. Stathmin is expressed at high levels in a wide variety of human cancers. Inhibition of stathmin expression in malignant cells interferes with their orderly progression through the cell cycle and abrogates their transformed phenotype. Thus, stathmin provides an attractive molecular target for disrupting the mitotic apparatus and arresting the growth of malignant cells [40]. The downregulation of stathmin upon DAC treatment gives support to the findings of Mistry and Atweh [40], who suggest stathmin as a therapeutic target in cancer therapy. Annexin 1 is a member of a family of calcium- and phospholipids binding proteins

related by amino acid sequence homology. Annexins 1 and 2 are substrates for protein tyrosine kinases. Both these proteins appear to be involved in mitogenic signal transduction and cell proliferation. All evidence so far gathered suggests that over-expression of annexins occurs in a variety of malignancies and well correlates with their progression [41–44].

4.3 Upregulated proteins

The most interesting among the 13 upregulated proteins is superoxide dismutase (SOD, 11- and 6-fold for a minor isoform), which is a component of the antioxidant system located in the mitochondrion. It has been recently shown that SOD could be a tumor suppressor in human pancreatic cancer, and suggested that delivery of the SOD gene might be used for pancreatic cancer gene therapy [45, 46].

4.4 Conclusions

In conclusion, the identification in a representative pancreatic cancer cell line of proteins, whose expression is altered by DNMT1 inhibition, may serve to understand the biology of the pancreatic adenocarcinoma and the molecular mechanisms involved in the response to DAC treatment. This information may represent a powerful tool for pancreatic cancer diagnosis and therapy with DNA methylation inhibitors.

Supported by grants from AIRC (Associazione Italiana per la Ricerca sul Cancro) to AS & PGR and from MIUR (No. 2001068593); Fondazione Cassa di Risparmio di Verona (Bandi 2001 & 2002, AS & PGR), and by the European Community (grant No. QL2-CT-2001-01903). PGR is additionally supported by grants from ASI I/R/294/02).

Received August 8, 2003

5 References

- [1] Almoguera, C., Shibata, D., Forrester, K., Martin, J., Arnheim, N., Perucho, M., *Cell* 1988, 53, 549–554.
- [2] Lemoine, N. R., Jain, S., Hughes, C. M., Staddon, S. L., Maillet, B., Hall, P. A., Kloppel, G., *Gastroenterology* 1992, 102, 230–236.
- [3] Scarpa, A., Capelli, P., Mukai, K., Zamboni, G., Oda, T., Iacono, C., Hirohashi, S., *Am. J. Pathol.* 1993, 142, 1534–1543.
- [4] Caldas, C., Hahn, S. A., Costa, L. T. D., Redston, M. S., Schutte, M., Seymour, A. B., Weinstein, C. L., Hruban, R. H., Yeo, C. J., Kern, S. E., *Nature Genet.* 1994, 8, 27–32.
- [5] Moore, P., Orlandini, S., Zamboni, G., Capelli, P., Rigaud, G., Falconi, M., Bassi, C., Lemoine, N., Scarpa, A., *Brit. J. Cancer.* 2001, 84, 253–262.
- [6] Kouzarides, T., *Curr. Opin. Genet. Dev.* 1999, 9, 40–48.
- [7] Archer, S. Y., Hodin, R. A., *Curr. Opin. Genet. Dev.* 1999, 9, 171–1744.
- [8] Herman, J. G., *Semin. Cancer Biol.* 1999, 9, 359–367.
- [9] Tsukiyama, T., Wu, C., *Curr. Opin. Genet. Dev.* 1997, 7, 182–191.
- [10] Gregory, P. D., Horz, W., *Eur. J. Biochem.* 1998, 251, 9–18.
- [11] Jacobson, S., Pillus, L., *Curr. Opin. Genet. Dev.* 1999, 9, 175–184.
- [12] Marks, P. A., Richon, V. M., Rifkind, R. A., *J. Natl. Cancer Inst.* 2000, 92, 1210–1216.
- [13] Cecconi, D., Scarpa, A., Donadelli, M., Palmieri, M., Hamdan, M., Aster, H., Righetti, P. G., *Electrophoresis* 2003, 24, 1871–1978.
- [14] Rountree, M. R., Bachman, K. E., Baylin, S. B., *Nat. Genet.* 2000, 25, 269–277.
- [15] Robertson, K. D., Ait-Si-Ali, S., Yokochi, T., Wade, P. A., Jones, P. L., Wolffe, A. P., *Nat. Genet.* 2000, 25, 338–342.
- [16] Jones, P. L., Veenstra, G. J., Wade, P. A., Vermaak, D., Kass, S. U., Landsberger, N., Strouboulis, J., Wolffe, A. P., *Nat. Genet.* 1998, 19, 187–191.
- [17] Ueki, T., Toyota, M., Sohn, T., Yeo, C. J., Issa, J. P., Hruban, R. H., Goggins, M., *Cancer Res.* 2000, 60, 1835–1839.
- [18] Sorio, C., Bonora, A., Orlandini, S., Moore, P., Cristofori, P., Dal Negro, G., Gaviraghi, G., Falconi, F., Pederzoli, P., Scarpa, A., *Virchows Archiv* 2001, 438, 154–159.
- [19] Schutte, M., Hruban, R. H., Geradts, J., Maynard, R., Hillgers, W., Rabindran, S. K., Moskaluk, C. A., Hahn, S. A., Schwarte-Waldhoff, I., Schmiegel, W., Baylin, S. B., Kern, S. E., Herman, J. G., *Cancer Res.* 1997, 57, 3126–3130.
- [20] Righetti, P. G., Stoyanov, A. V., Zhukov, M. Y., *The Proteome Revisited*, Elsevier, Amsterdam 2001.
- [21] Herbert, B., Galvani, M., Hamdan, M., Olivieri, E., McCarthy, J., Pedersen, S., Righetti, P. G., *Electrophoresis* 2001, 22, 2046–2057.
- [22] James, P. (Ed.), *Proteome Research: Mass Spectrometry*, Springer, Berlin 2001.
- [23] Hamdan, M., Righetti, P. G., *Mass Spectr. Rev.* 2002, 21, 287–302.
- [24] Kanai, Y., Ushijima, S., Saito, Y., Nakanishi, Y., Sakamoto, M., Hirohashi, S., *J. Cancer Res. Clin. Oncol.* 2001, 127, 697–706.
- [25] Sinha, P., Hutter, G., Kottgen, E., Dietel, M., Schadendorf, D., Lage, H., *Electrophoresis* 1999, 20, 2952–2960.
- [26] Hu, E., Chen, Z., Fredrickson, T., Zhu, Y., *Exp. Nephrol.* 2001, 9, 265–274.
- [27] Feldner, J. C., Brandt, B. H., *Exptl. Cell Res.* 2002, 272, 93–108.
- [28] Provost, P., Doucet, J., Stock, A., Gerisch, G., Samuelsson, B., Radmark, O., *Biochem. J.* 2001, 359, 255–263.
- [29] Provost, P., Doucet, J., Hammarberg, T., Gerisch, G., Samuelsson, B., Radmark, O., *J. Biol. Chem.* 2001, 276, 16520–16527.
- [30] Nakatsura, T., Senju, S., Ito, M., Nishimura, Y., Itoh, K., *Eur. J. Immunol.* 2002, 32, 826–836.
- [31] Meza-Zepeda, L. A., Forus, A., Lygren, B., Dahlberg, A. B., Godager, L. H., South, A. P., Marenholz, I., Lioumi, M., Flores, V. A., Maelansmo, G. M., Serra, M., Mischke, D., Nizetic, D., Ragoussis, J., Tarkkanen, M., Nesland, J. M., Knuutila, S., Myklebost, O., *Oncogene* 2002, 21, 2261–2269.
- [32] Lim, S. O., Park, S. J., Kim, W., Park, S. G., Kim, H. J., Kim, Y. I., Sohoni, T. S., Noh, J. H., Jung, G., *Biochem. Biophys. Res. Commun.* 2002, 291, 1031–1037.

- [33] Calkins, C. C., Sameni, M., Koblinski, J., Sloane, B. F., Moin, K., *J. Histochem. Cytochem.* 1998, *46*, 745–751.
- [34] Kos, J., Lah, T. T., *Oncol. Reports* 1998, *5*, 1349–1361.
- [35] Kos, J., Krasovec, M., Cimerman, N., Nielsen, H. J., Christensen, I. J., Brunner, N., *Clin. Cancer Res.* 2000, *6*, 505–511.
- [36] Ebert, E., Werle, B., Julke, B., Kopitaer-Jerala, N., Kos, J., Lah, T., Abrahamson, M., Spiess, E., Ebert, W., *Adv. Exptl. Med. Biol.* 1997, *421*, 259–265.
- [37] Seraj, M. J., Harding, M. A., Zgildea, J. J., Welch, D. R., Theodorescu, D., *Clin. Exp. Metastasis* 2000, *18*, 519–525.
- [38] Fritz, G., Brachetti, C., Bahlmann, F., Schmidt, M., Kaina, B., *Br. J. Cancer* 2002, *87*, 635–644.
- [39] Kovarova, H., Hajduch, M., Korinkova, G., Halada, P., Krupickova, S., Gouldsworth, A., Zhelev, N., Strnad, M., *Electrophoresis* 2000, *21*, 3757–3764.
- [40] Mistry, S. J., Atweh, G. F., *Mt. Sinai J. Med.* 2002, *69*, 299–304.
- [41] De Coupade, C., Gillet, R., Bennoun, M., Briand, P., Russo-Marie, F., Solito, E., *Hepatology* 2000, *31*, 371–380.
- [42] Wu, W., Tang, X., Hu, W., Lotan, R., Hong, W. K., Mao, L., *Clin. Exp. Metastasis* 2002, *19*, 319–326.
- [43] Rhee, H. J., Kim, G. Y., Huh, J. W., Kim, S. W., Na, D. S., *Eur. J. Biochem.* 2000, *267*, 3220–3225.
- [44] Pencil, S. D., Toth, M., *Clin. Exp. Metastasis* 1998, *16*, 113–121.
- [45] Weydert, C., Roling, B., Liu, J., Hinkhouse, M. M., Ritchie, J. M., Oberley, L. W., Cullen, J. J., *Mol. Cancer Therapy* 2003, *2*, 361–369.
- [46] Cullen, J. J., Weydert, C., Hinkhouse, M. M., Ritchie, J. M., Domann, F. E., Spitz, D., Oberley, L. W., *Cancer Res.* 2003, *63*, 1297–303.
- [47] Saotome, K., Morita, H., Umeda M., *Toxicology in vitro* 1989, *3*, 317–321.

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.

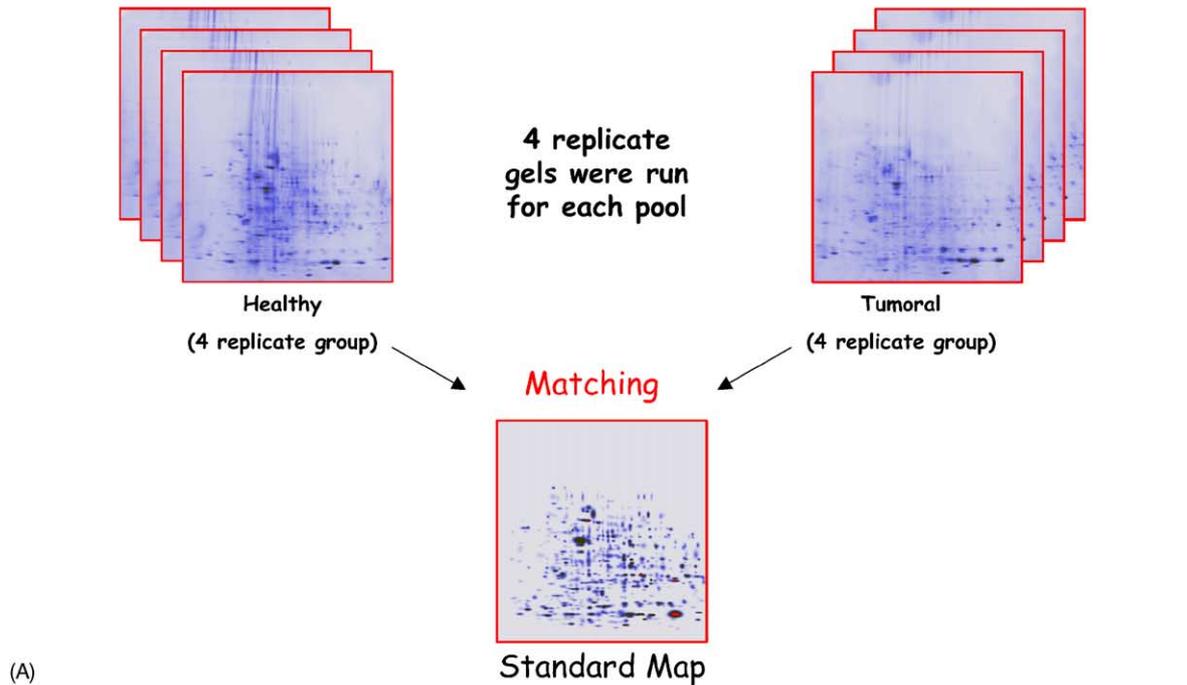
There has been a sudden burst, in the last few years, of methods describing novel approaches to quantitative proteomics. Quite a few of them have been reviewed in a number of papers dedicated to these topics [10–13]. Such methods comprise not only differential proteomics in the well-ingrained two-dimensional (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

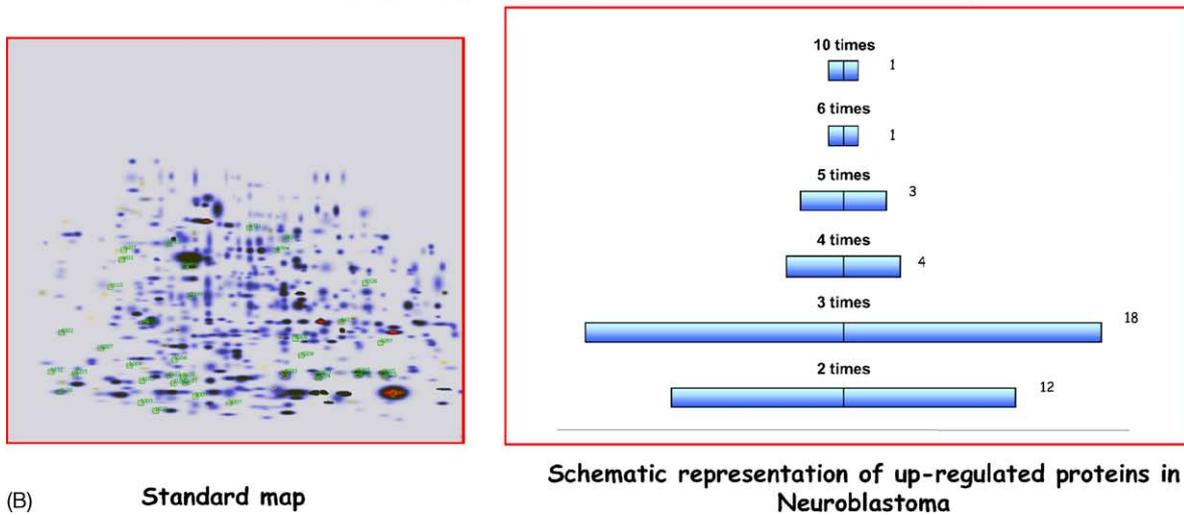
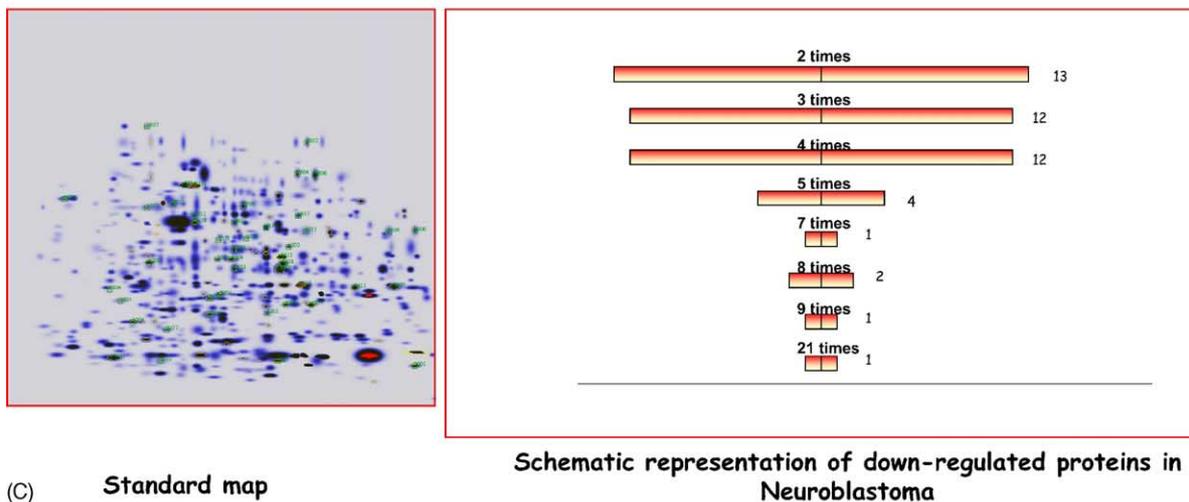


Fig. 1. Experimental design for generating sets of 2D maps from control and treated samples and for comparing master maps, stained with colloidal Coomassie Blue, for detection of up- and down-regulated proteins in the paired samples. All comparative steps performed with the PDQuest software. (A) Creating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

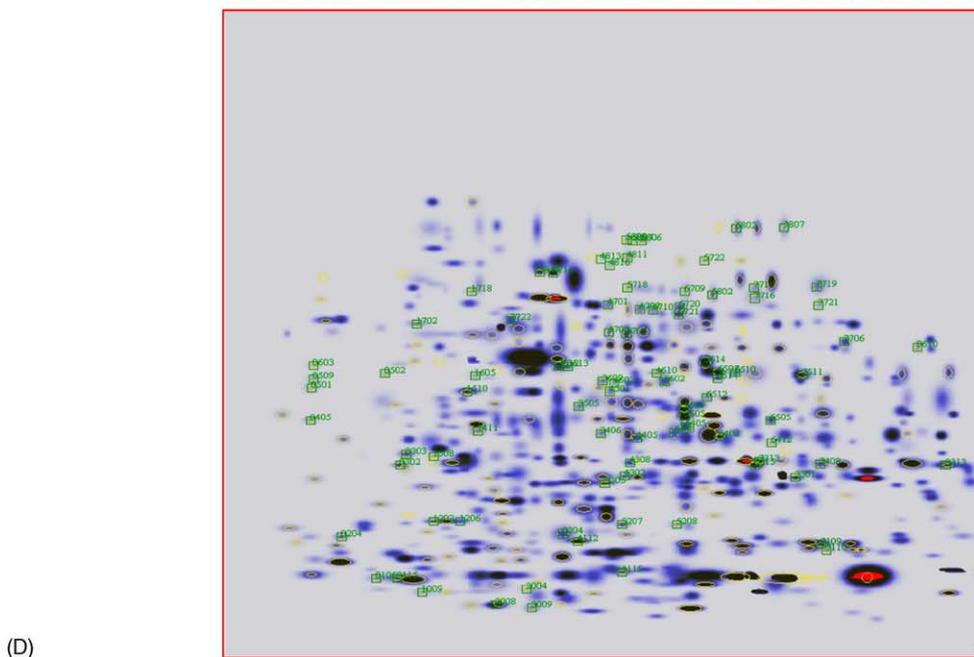


Fig. 1. (Continued).

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.

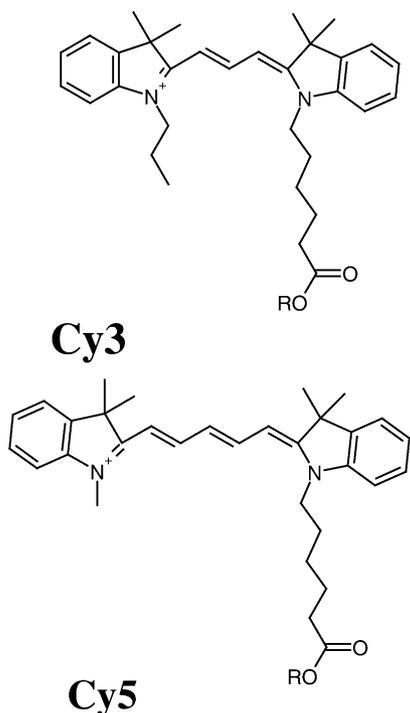


Fig. 2. Chemical formulae of the Cy3 and Cy5 dyes.

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

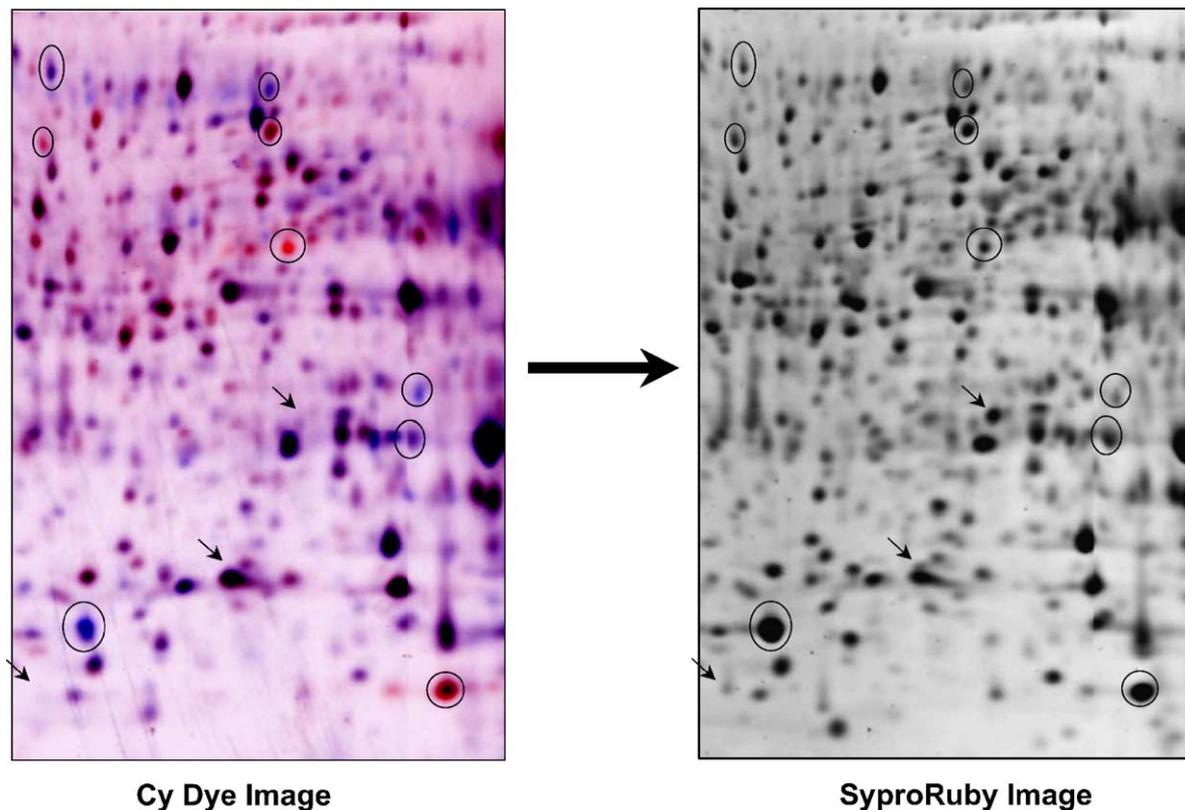


Fig. 3. Comparison of 2D-DIGE imaging and Sypro Ruby post-staining. Left panel: merged Cy dye image of HB4a lysate labelled with Cy3 (red) and HBc3.6 lysate labelled with Cy5 (blue). The same gel was post-stained with Sypro Ruby (right panel). Circles represent differentially expressed proteins detectable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

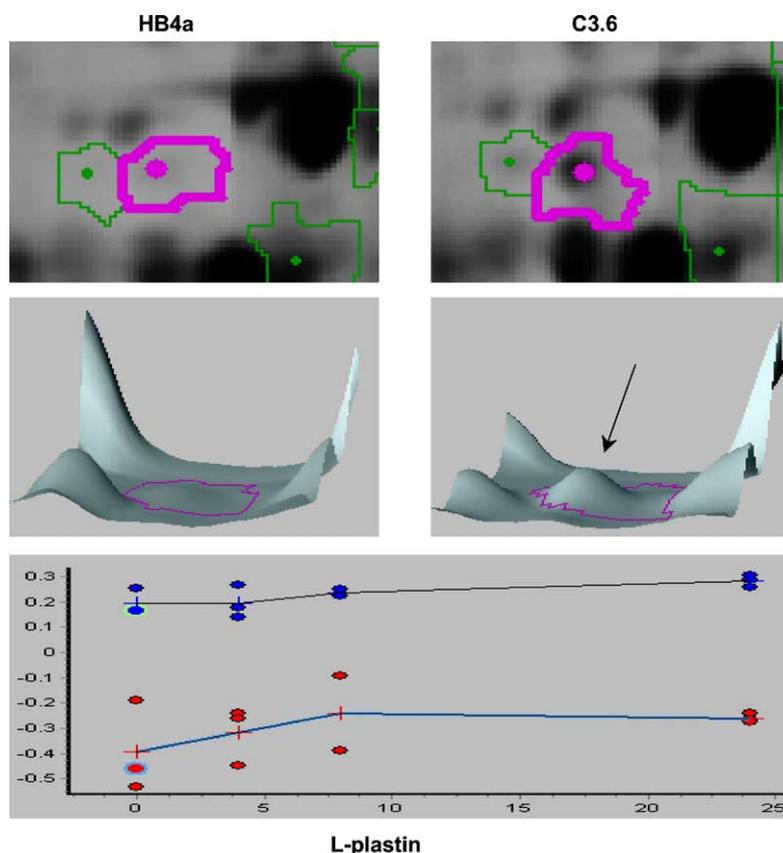


Fig. 4. Detailed gel area on a Cy3/Cy5 labelled sample pool, showing the induction of L-plastin in tumour cells. Bottom panel: kinetics of L-plastin induction. From [22] with permission.

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

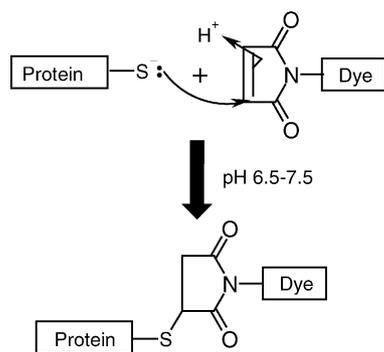


Fig. 5. Reaction scheme of maleimido cyanine dyes with the $-SH$ group of proteins. From [26] with permission.

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-

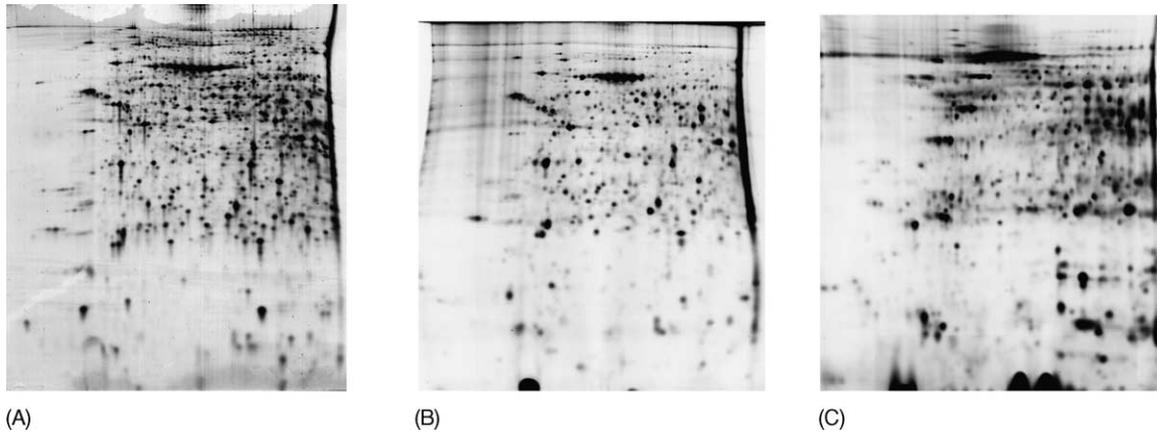


Fig. 6. 2D gel images of (A) silver stained, (B) Cy3 minimal dye-labelled and (C) Cy3 saturation dye-labelled liver homogenate (50 μ g, pH 4–7 18 cm IPG strip). Whilst the silver stained image and the Cy3 minimal dye image are very similar, the Cy3 saturation dye image shows an altered spot pattern.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

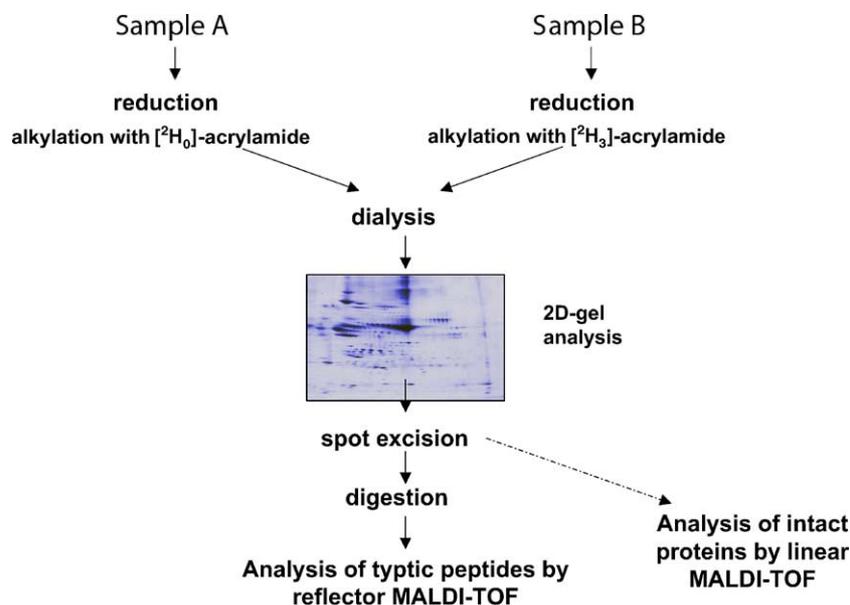


Fig. 7. Scheme for differential labelling of two samples with $[^2\text{H}_0]/[^2\text{H}_3]$ -acrylamide (alkylation of Cys residues). The central map refers to rat sera, labelled separately with either $[^2\text{H}_0]$ - or $[^2\text{H}_3]$ -acrylamide and mixed in a 30:70% ratio. From [34] with permission.

“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-

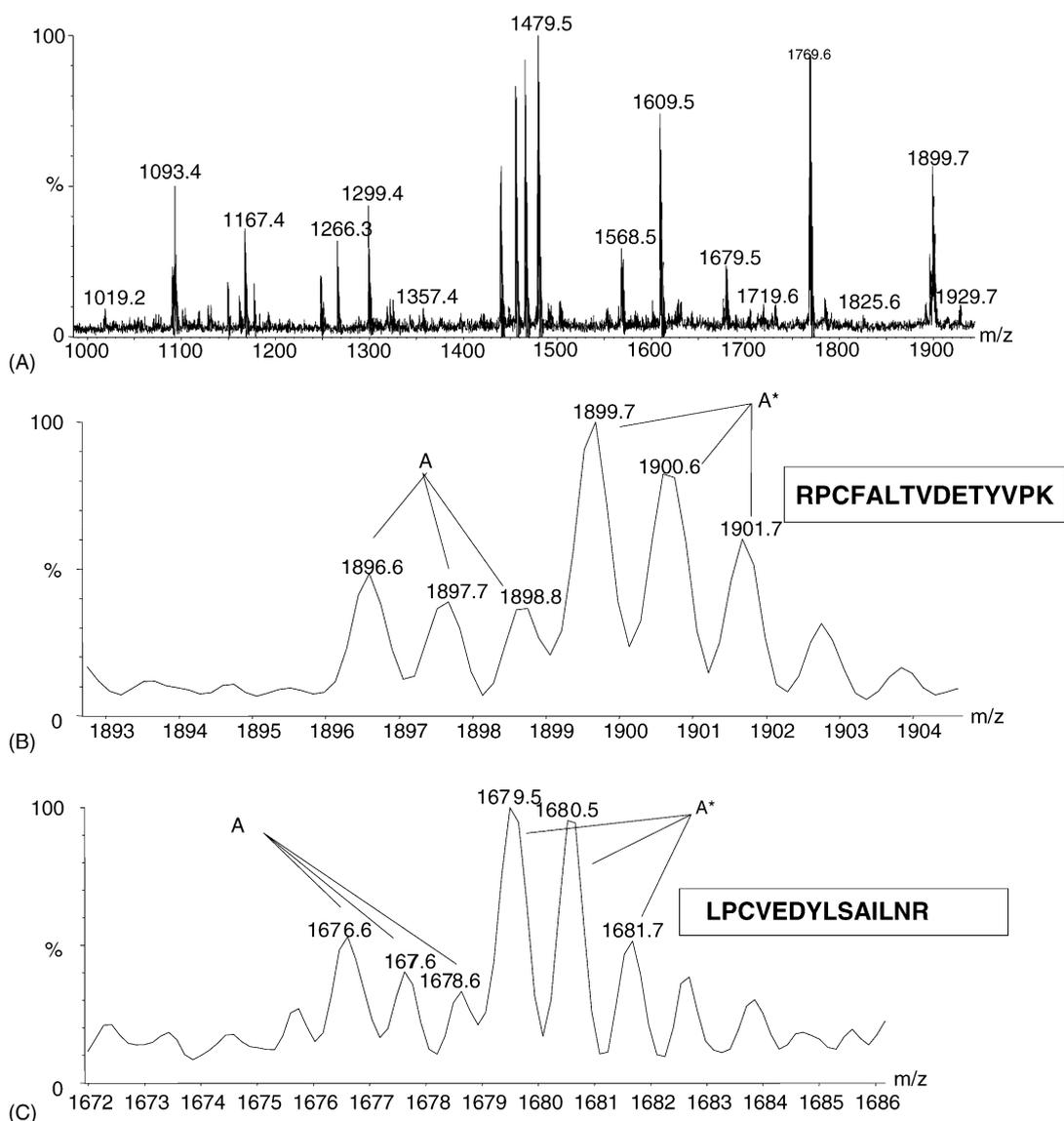


Fig. 8. (A) Reflector MALDI mass spectrum of an in-situ digest of apo-transferrin taken from the 2D map of rat sera displayed in Fig. 7, that were alkylated with $[^2\text{H}_0]$ - and $[^2\text{H}_3]$ -acrylamide and mixed in a 30%/70% ratio. (B) and (C) are two short intervals taken from (A), and are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

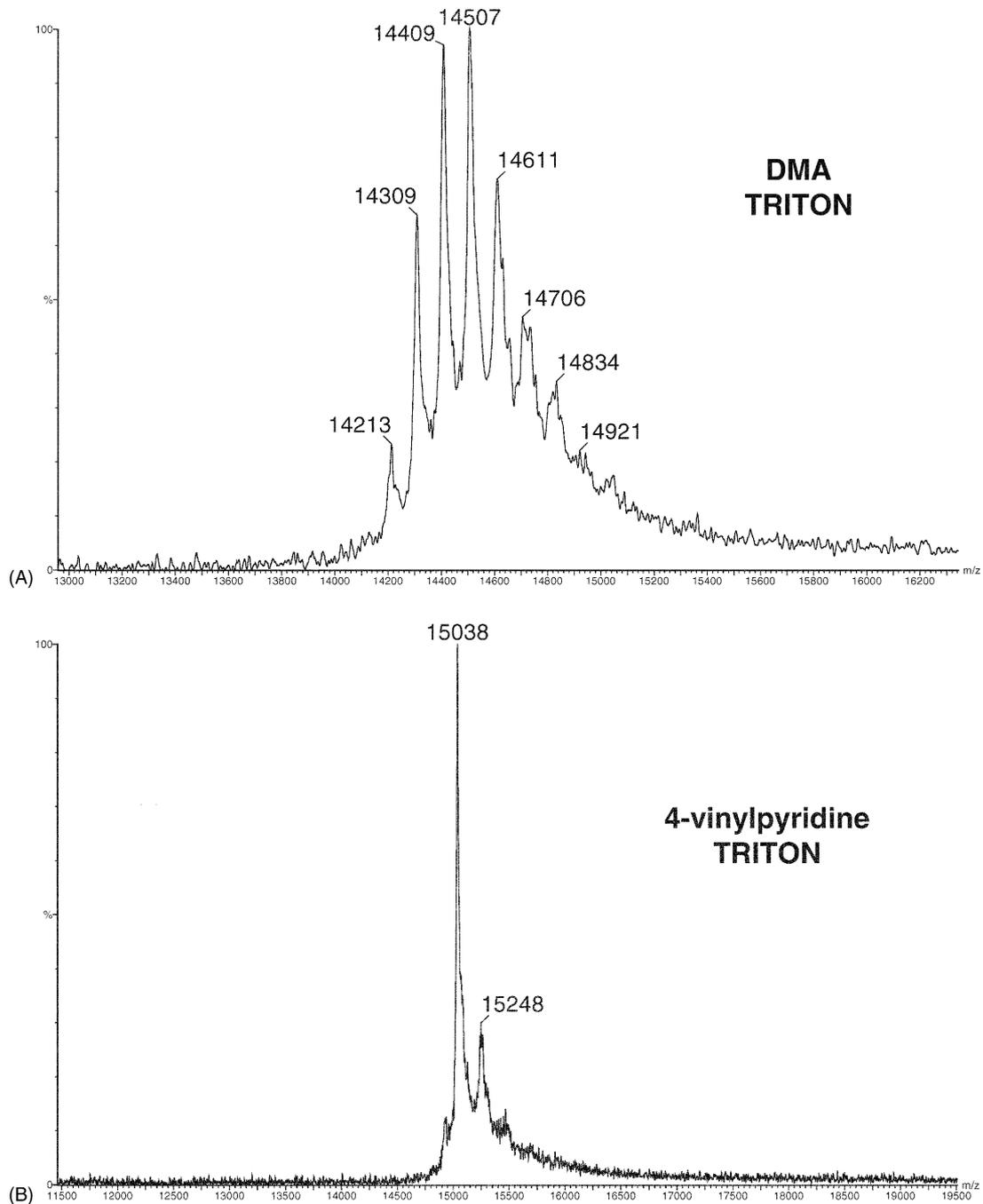


Fig. 9. MALDI-TOF mass spectra of bovine α -lactalbumin after 1 h incubation with DMA (A) or 4-vinylpyridine (B), both in presence of the surfactant 2% Triton X-100. Note that, in panel B, the peak at m/z 15248 represents an adduct of LCA with the MALDI matrix, sinapinic acid. From [36] with permission.

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).

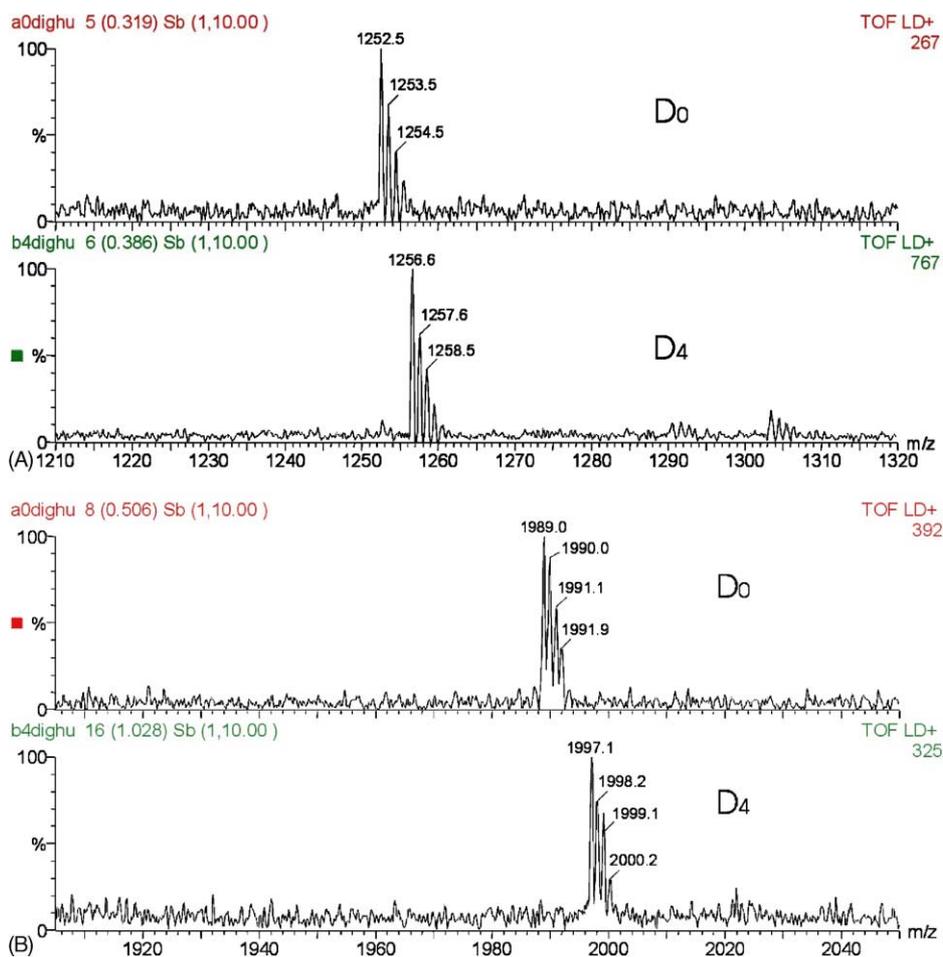


Fig. 10. (A) Zoom-in to monoisotopic distributions of the m/z 1252.5 peptide, obtained by tryptic digestion of α -lactalbumin, labelled with [$^2\text{H}_0$] 2-VP (D₀) and the corresponding m/z 1256.6 peptide labelled with [$^2\text{H}_4$] 2-VP (D₄). The peptide contains a single Cys residue. (B) Zoom-in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

Emilio Marengo · Elisa Robotti · Daniela Cecconi
Mahmoud Hamdan · Aldo Scarpa · Pier Giorgio Righetti

Identification of the regulatory proteins in human pancreatic cancers treated with Trichostatin A by 2D-PAGE maps and multivariate statistical analysis

Received: 8 January 2004 / Revised: 26 May 2004 / Accepted: 26 May 2004 / Published online: 15 July 2004
© Springer-Verlag 2004

Abstract In this paper, principal component analysis (PCA) is applied to a spot quantity dataset comprising 435 spots detected in 18 samples belonging to two different cell lines (Paca44 and T3M4) of control (untreated) and drug-treated pancreatic ductal carcinoma cells. The aim of the study was the identification of the differences occurring between the proteomic patterns of the two investigated cell lines and the evaluation of the effect of the drug Trichostatin A on the protein content of the cells. PCA turned out to be a successful tool for the identification of the classes of samples present in the dataset. Moreover, the loadings analysis allowed the identification of the differentially expressed spots, which characterise each group of samples. The treatment of both the cell lines with Trichostatin A therefore showed an appreciable effect on the proteomic pattern of the treated samples. Identification of some of the most relevant spots was also performed by mass spectrometry.

Keywords PCA · Chemometrics · Human pancreatic tumour · Trichostatin A · Protein identification

E. Marengo (✉) · E. Robotti
Department of Environmental and Life Sciences,
University of Eastern Piedmont, Spalto Marengo 33,
15100 Alessandria, Italy
E-mail: emilio.marengo@mfn.unipmn.it
Tel.: +39-131-287428
Fax: +39-131-287416

D. Cecconi · P. G. Righetti
Department of Industrial Biotechnologies,
University of Verona, Strada le Grazie 15,
37134 Verona, Italy

M. Hamdan
Computational, Analytical and Structural Sciences,
GlaxoSmithKline, Via Fleming 4,
37135 Verona, Italy

A. Scarpa
Department of Pathology,
Section of Anatomical Pathology, University of Verona,
37134 Verona, Italy

Introduction

Since each cell or biological fluid has a rich protein content (often comprising thousands of proteins of different structure and size), an effective method for achieving their separation is necessary. In the field of proteomics [1, 2], the separation of proteins is usually achieved by two-dimensional (2D) electrophoresis, a very powerful tool which performs two successive electrophoretic runs: the first run (through a pH gradient) separates the proteins with respect to their isoelectric point, while the second run (through a porosity gradient or a highly sieving, constant concentration gel) separates them according to their molecular mass. This technique produces a two-dimensional map, a so-called 2D-PAGE (polyacrylamide gel electrophoresis), with the proteins appearing as spots spread all over the gel matrix. A 2D-PAGE map may thus be considered as a “snapshot” of the protein content of the investigated cell at a given point of its life cycle.

In this new post-genomic and proteomic era, the investigation of the protein content of different cell types has become fundamental. In fact, 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 pathological tissue: these differences may consist of changes in the relative abundance or even in the appearance/disappearance of some proteins [3–11]. The comparison of 2D-PAGE maps belonging to healthy subjects with samples belonging to individuals affected by any pathology thus becomes a fundamental tool for both diagnostic and prognostic purposes [3–11]. The 2D-PAGE technique is also widely applied in the field of drug development [12–15], especially for cancer: two-dimensional gel-electrophoresis may be used to investigate if the treatment with a particular drug has played the expected role on the protein content of the pathological cell and to evaluate which effect was produced (e.g. up- or down-regulation, appearance/disappearance of pathological chains).

Unfortunately, the comparison of 2D-PAGE maps is not a trivial problem; its difficulty is principally due to:

- The high complexity of the sample, which can produce maps with thousands of spots
- The complex sample pre-treatment, characterised by several purification/extraction steps, which may contribute to the appearance of maps with spurious spots due to accidental chemical modifications
- The sometimes small differences which often occur in the 2D-PAGE maps of treated and reference samples, which are much more difficult to recognise in complex maps.

Usually, the comparison of 2D-PAGE maps is performed by means of some specific softwares (e.g. Melanie III or PDQuest) [16–19], which exploit the following three-step method:

- The 2D-PAGE images to be compared are aligned, so that all images are reduced to the same size. This step needs the choice of at least two positively identified spots in all the maps; the maps are then matched to each other on the basis of the position of these two spots
- The spots present on each map are independently revealed
- The maps are matched to each other in order to identify the common information (spots present in all the maps) and the different one (spots detected only on some of the samples). If the comparison is performed on a set of replicate maps this step produces a “synthetic” map which summarises the common information and contains only the spots present in all the compared maps.

The great amount of information produced by the comparison of 2D maps can be investigated by modern multivariate techniques like principal component analysis (PCA) [20–22], classification methods [23, 24] and multidimensional scaling (MDS) [25].

Our research group has developed a new method based on both fuzzy logic and classification methods for the comparison of the proteomic pattern of classes of 2D-PAGE maps [26, 27]. This method has also been applied together with MDS for the study of 2D maps from control and diseased individuals [28]. Different proteomics patterns have also been investigated by the use of three-way principal component analysis [29].

PCA has been applied to the comparison of 2D-PAGE maps on the basis of the spot volume since the mid-1980s by Anderson et al. [30] in the USA and Tarroux et al. [31] in France. Recently, it has been applied to the study of DNA and RNA fragments of several biological systems [32–35] and to the characterisation of proteomic patterns of different classes of tissues [36–41]. Another recent application of PCA is for the characterisation of the anticancer activity of bohemine, a new omoleucine-derived synthetic cyclin-dependent kinase inhibitor, by Kovarova et al. [42].

In this paper, PCA is applied to a dataset comprising 18 samples belonging to two different cell lines (Paca44 and T3M4) of pancreatic human cancer before and after the treatment with a new drug (Trichostatin A). This approach focuses on the evaluation of the efficacy of the drug (reflected in a difference in the protein content of control and treated samples) and to the identification of the differences occurring between the samples (control/treated samples and Paca44/T3M4 cell lines). Some of the proteins responsible for the identified differences in the control and treated Paca44 samples were also characterised by mean of mass spectrometry with the matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) technique.

Theory

Principal component analysis

Principal component analysis [20–22] 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 residual variance contained in the original dataset: the first PC explains the maximum amount of variance contained in the overall dataset, while the second one explains the maximum residual variance. The PCs are then calculated hierarchically, so that experimental noise and random variations are contained in the last PCs. 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*. The graphical representation of scores by means of PCs allows the identification of groups of samples showing a similar behaviour (samples close to one another in the graph) or different characteristics (samples far from each other). By looking at the corresponding loading plot, it is possible to identify the variables, which are responsible for the analogies or the differences detected for the samples in the score plot. From this point of view, PCA is a very powerful visualisation tool, which allows the representation of multivariate datasets by means of only few PCs identified as the most relevant.

Cluster analysis

Cluster analysis techniques allow one to investigate the relationships between the objects or the variables of a

dataset in order to recognise the existence of groups. The most commonly used approaches belong to the class of agglomerative hierarchical methods [43], in which 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, in which 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 the linking method.

Experimental

The dataset comprised 18 2D maps, which were divided into four classes:

- Four replicate 2D maps of a Paca44 cell line pool
- Five replicate 2D maps of a T3M4 cell line pool
- Four replicate 2D maps of a Paca44 cell pool treated for 48 h with Trichostatin A
- Five replicate 2D maps of a T3M4 cell pool treated for 48 h with Trichostatin A.

Because we used pools of cell lines (grown under the same conditions), four to five replicates of 2D maps for each sample were deemed amply sufficient for reducing variability due to experimental errors. This strategy is common practice in today's proteome analysis. Figure 1 represents an example, for each class, of the experimental 2D maps obtained.

Software

Principal component analysis was performed with UNSCRAMBLER (Camo Inc., version 7.6, Norway). Cluster analysis was performed with STATISTICA (Statsoft Inc., version 5.1, USA). Graphical representations were performed with both UNSCRAMBLER and STATISTICA. The 2D-PAGE maps were scanned with a GS-710 densitometer (Bio-Rad Laboratories, Hercules, CA, USA) and analysed with the software PDQuest Version 6.2 (Bio-Rad Laboratories, Hercules, CA, USA).

Chemicals and materials

Urea, thiourea, 3-[(cholamidopropyl)dimethylammonium]-1-propane-sulfonate (CHAPS), iodoacetamide (IAA),

tributylphosphine (TBP) and sodium dodecyl sulfate (SDS) were obtained from Fluka Chemie (Buchs, Switzerland). Bromophenol blue and agarose were from Pharmacia-LKB (Uppsala, Sweden). Acrylamide, N,N' -methylenebisacrylamide, ammonium persulfate, TEMED, the Protean IEF Cell, the GS-710 Densitometer and the 17-cm-long, immobilised pH 3–10 linear gradient strips were from Bio-Rad Laboratories (Hercules, CA, USA). Ethanol, methanol and acetic acid were from Merck (Darmstadt, Germany). Trichostatin A (TSA) was obtained from Sigma-Aldrich Ltd. (St. Louis, MO, USA). A 3.3 mM solution of TSA in absolute ethanol was prepared and stored at -80°C until use.

Cell treatment with TSA

Paca44 and T3M4 cells were grown in RPMI 1640 supplemented with 20 mM glutamine and 10% (v/v) FBS (BioWhittaker, Italy) and were incubated at 37°C with 5% (v/v) CO_2 . Subconfluent cells were treated with 0.2 mM TSA for 48 h.

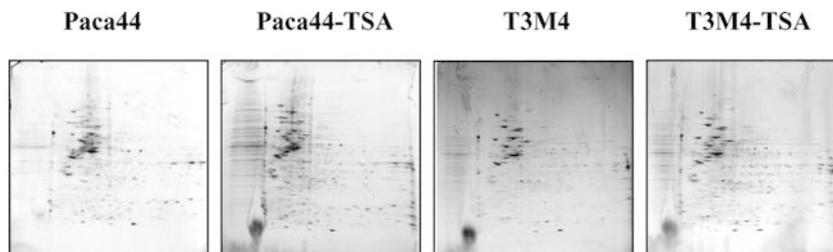
Cell lysis

Protein extraction from cells was performed with lysis buffer (40 mM Tris, 1% v/v NP40, 1 mM Na_3VO_4 , 1 mM NaF, 1 mM PMSF, protease inhibitor cocktail). Cells were left in lysis buffer for 30 min in ice. After centrifugation at $14,000\times g$ at 4°C for removal of particulate material, the protein solution was collected and stored at -80°C until used.

Two-dimensional gel electrophoresis

Seventeen-centimeter-long, pH 3–10 immobilized pH gradient strips (IPG; Bio-Rad Labs., Hercules, CA, USA) were rehydrated for 8 h with 450 μL of 2D solubilizing solution (7 M urea, 2 M thiourea, 5 mM tributylphosphine, 40 mM Tris and 20 mM iodoacetamide) containing 2 mg mL^{-1} of total reduced/alkylated protein from sample cells. Isoelectric focussing (IEF) was carried out with a Protean IEF Cell (Biorad, Hercules, CA, USA) with a low initial voltage and then by applying a voltage gradient up to 10,000 V with a limiting current of 50 μA . The total product time \times voltage applied was 70,000 Vh for each strip, and the tempera-

Fig. 1 2D-PAGE maps of the real samples of pancreatic human cancer: examples of control (untreated) Paca44 cells, treated Paca44 cells, control (untreated) T3M4 cells and treated T3M4 cells



ture was set at 20°C. For the second dimension, the IPGs strips were equilibrated for 26 min by rocking in a solution of 6 M urea, 2% w/v SDS, 20% v/v glycerol, 375 mM Tris-HCl, pH 8.8. The IPG strips were then laid on an 8–18% T gradient SDS-PAGE with 0.5% w/v agarose in the cathode buffer (192 mM glycine, 0.1% w/v SDS and Tris to pH 8.3). The anodic buffer was a solution of 375 mM Tris HCl, pH 8.8. The electrophoretic run was performed by setting a current of 2 mA for each gel for 2 h, then 5 mA/gel for 1 h, 10 mA/gel for 20 h and 20 mA/gel until the end of the run. During the whole run the temperature was set at 11°C. Gels were stained overnight with colloidal Coomassie blue (0.1% w/v Coomassie Brilliant Blue G, 34% v/v methanol, 3% v/v phosphoric acid and 17% w/v ammonium sulphate); destaining was performed with a solution of 5% v/v acetic acid until a clear background was achieved.

Protein identification by mass spectrometry

In situ digestion and extraction of peptides

The spots of interest were carefully excised from the gel with a razor blade, placed in Eppendorf tubes, and destained by washing three times for 20 min in 50% v/v acetonitrile, 2.5 mM Tris, pH 8.5. The gel pieces were dehydrated at room temperature and covered with 10 μ L of trypsin (0.04 mg mL⁻¹) in Tris buffer (2.5 mM, pH 8.5) and left at 37°C overnight. The spots were crushed and peptides were extracted in 15 μ L of 50% v/v acetonitrile, 1% v/v formic acid. The extraction was conducted in an ultrasonic bath for 15 min. The sample was centrifuged at 8,000 \times g for 2 min, and the supernatant was collected.

MALDI-TOF analysis

The extracted peptides were loaded onto the 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⁻¹ cyano-4-hydroxycinnamic acid in acetonitrile/ethanol (1:1 v:v), and allowed to dry. Measurements were performed by using a TofSpec 2E MALDI-TOF instrument (Micromass, Manchester, UK), operated in reflectron mode, with an accelerating voltage of 20 kV. Peptide masses were searched against SWISS-PROT, TrEMBL and NCBI nr databases by utilizing the ProteinLynx program from Micromass, Profound from Prowl, and Mascot from Matrix Science.

Results and discussion

Protein pattern analysis with the PDQuest software

The 2D gels of all the samples (Paca44 control and treated with TSA, T3M4 control and treated with TSA)

were scanned with a GS-710 densitometer (Bio-Rad), and analysed with the software PDQuest. A match-set was created from the protein patterns of the 18 replicates 2D maps. A standard gel was generated out of the image with the highest spot number. Spot quantities of all gels were normalized to remove non-expression-related variations in spot intensity, so 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 optical density (OD). The analysis with the PDQuest allowed two types of comparisons: between the two different cell lines (Paca44 versus T3M4), and between the control and TSA-treated cell lines (control versus TSA-treated) in order to detect protein variations that were at least two-fold. The Student's *t*-test analysis allowed the identification of 60 spots up-regulated (with a significance level α of 0.05) and 45 spots down-regulated (with $\alpha=0.05$) in the T3M4 cell line with respect to the PaCa44 cell line; and 11 spots up-regulated (with a significance level α of 0.05) and two spots down-regulated in TSA-treated cell lines in respect to the control samples.

Figures 2a and b show the results of comparison between the two different cell lines. In Fig. 2a the 45 spots with a higher optical density than in Paca44 (which were thus less intense in the T3M4 cell line) are marked in red, whereas in Fig. 2b the 60 spots with a higher optical density than in T3M4 are marked in red.

Figures 2c and d show the results of comparison between the control and TSA-treated cell lines. In Fig. 2c the two spots down-regulated in TSA-treated cell lines (which were thus more intense in the control) are marked in red, while in Fig. 2d the 11 spots up-regulated in TSA-treated cell lines are marked in red.

Principal component analysis

The differential analysis performed by PDQuest on the 18 2D maps allowed the identification of 435 spots. The matching procedure produced a dataset comprising 435 variables (the optical density of each matched spot) and 18 objects (the 18 samples), thus giving a data matrix of dimensions 18 \times 435. All variables were autoscaled before performing PCA. The autoscaling procedure transforms the variables so that they all have a null average value and a unit variance: this last feature is fundamental, since it allows all the variables to bring the same amount of information to the overall dataset. In the present case, the autoscaling procedure is particularly suitable: it gives the small and large spots the same relevance, thus enhancing the detection of the differences between the four classes of 2D maps, which are often to be searched for among the smallest spots (less abundant proteins) rather than among the largest ones (more abundant proteins). Since PCA was performed on a singular matrix, with more variables than samples, a NIPALS

Fig. 2a–d Results of a comparison between the two different cell lines: **a** the 45 spots with lower optical density than in T3M4 (thus more intense in Paca44 cell line) are marked in red; **b** the 60 spots of lower intensity than in T3M4 are marked in red. **c** and **d** show the results of comparison between the control and TSA-treated cell lines: in panel **c** the two spots down-regulated in TSA-treated cell lines (more intense in the control) are marked in red, while in panel **d** the 11 spots up-regulated in TSA-treated cell lines are marked in red

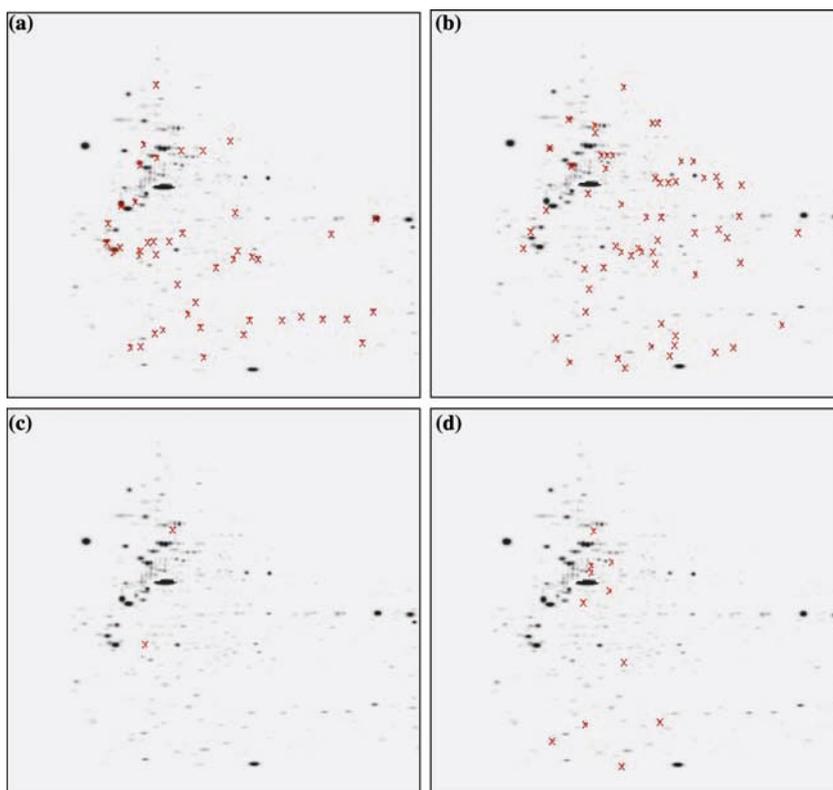


Table 1 Results of PCA performed on the overall dataset: percentage of explained variance and percentage of cumulative explained variance

	Explained variance (%)	Cumulative explained variance (%)
PC1	37.59	37.59
PC2	12.90	50.49
PC3	8.27	58.75
PC4	5.42	64.17
PC5	4.70	68.87
PC6	4.33	73.20
PC7	3.47	76.67

algorithm was used for PCs calculation. The results of PCA are given in Table 1. The first three PCs explain more than 58% of the total variance contained in the original dataset and were considered for the successive analysis.

Figure 3 represents the score plots of the first three principal components. The first score plot (Fig. 3a) shows the samples' co-ordinates on PC_2 and PC_1 : the four classes of samples appear completely separated along the two PCs. In fact, at large positive values along PC_1 , there are the samples belonging to the cell line Paca44, while the samples belonging to the second cell line, T3M4, are grouped at large negative scores on the same PC. The information about the cell type differences is then explained by the first PC. The information about the differences occurring between control and drug-treated samples is instead explained by the second PC:

the drug-treated samples of both cell lines in fact appear at large negative scores on PC_2 and the control samples at large positive scores on the same PC. However, PC_2 is dominated by the relative down-regulation of the spots in the Paca44 cell line, since the samples belonging to this cell line show the largest variations of the scores on the second PC. Thus PC_2 describes the TSA general effect, with a larger contribution due to the Paca44 cell line. The first two PCs are thus able to account for the differences occurring between the four classes investigated; however, PC_3 also explains a significant amount of variance, which is worthy of being interpreted. The score plot of PC_3 versus PC_1 (Fig. 3b) shows the control samples belonging to the T3M4 cell line at large positive scores on PC_3 together with the TSA-treated ones of the Paca44 cell line; the other two groups of samples (TSA-treated samples of the T3M4 cells, and control samples of Paca44 cells) being grouped at large negative scores on the same PC. However, as for PC_2 , the third PC is dominated by the relative down-regulation of the spots in the T3M4 cell line, since the samples belonging to this cell line show the largest change of the scores on PC_3 . The third PC therefore mainly describes information about the effect of TSA on the T3M4 cell line (i.e. complementary information with respect to that accounted for by PC_2). Figure 4 reports the cell survival of the two cell lines which present a similar sensitivity to a 48-h treatment with Trichostatin A. With respect to this further information, the third PC mainly accounts for the sensitivity of the T3M4 cell line to the treatment with TSA.

Fig. 3a, b Score plots of the first three PCs: **a** PC₂ versus PC₁ and **b** PC₃ versus PC₁

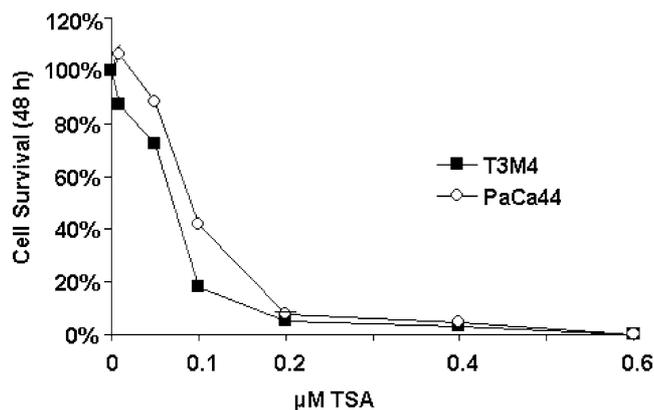
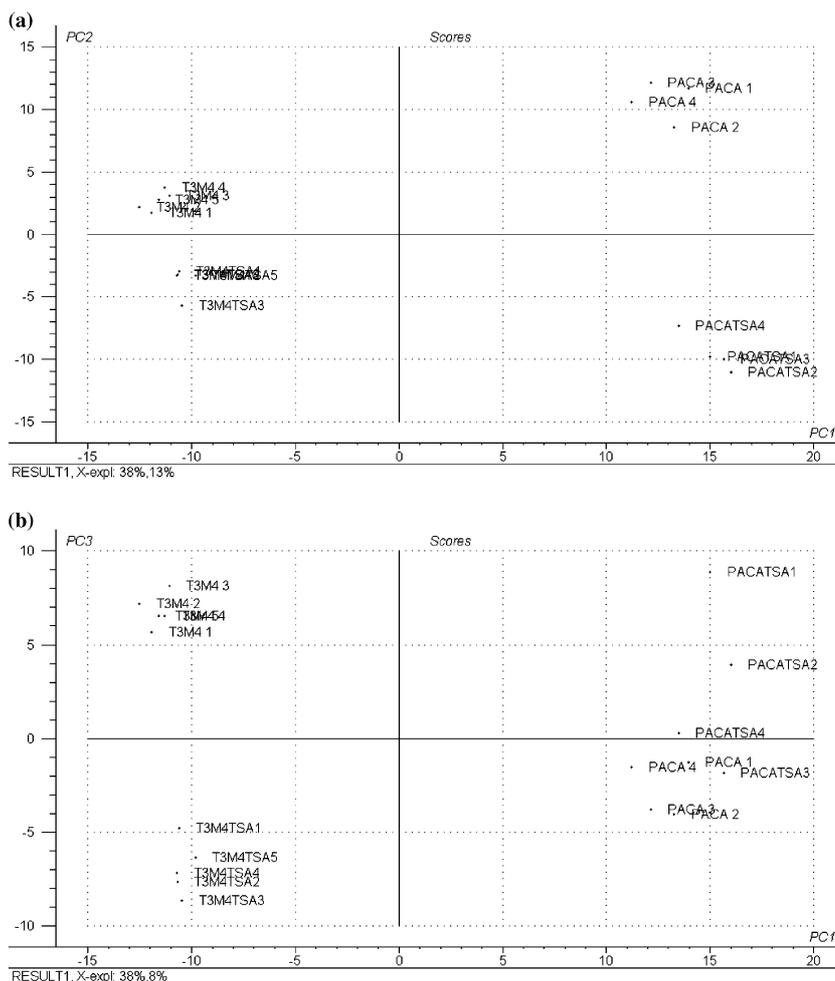


Fig. 4 Cell survival after a 48-h treatment with Trichostatin A for Paca44 and T3M4 cell lines

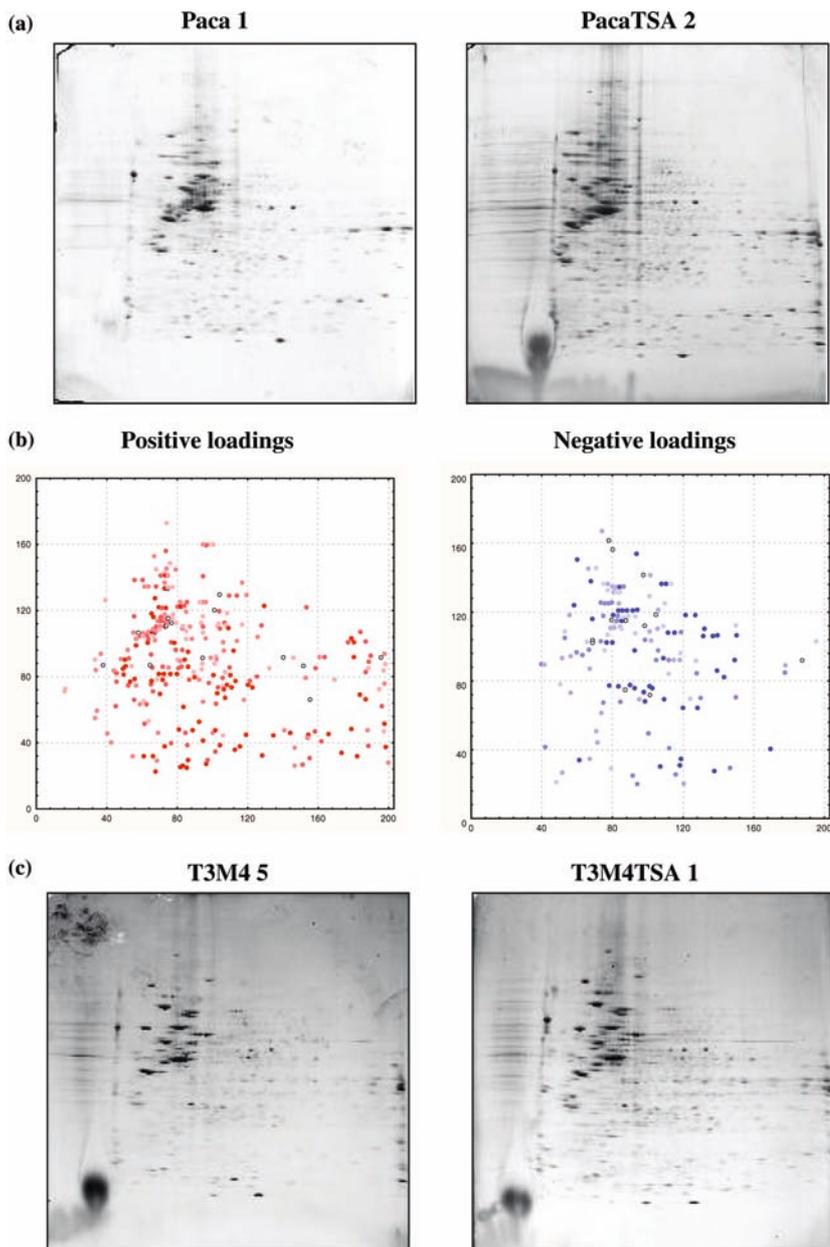
From the previous considerations, it is possible to state that the first three PCs allow a synthetic and exhaustive representation of the investigated dataset:

- 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 the sensitivity to TSA, mainly for the T3M4 cell line.

The loadings of the three significant PCs provide information on the spots responsible for the regulatory effect (i.e. they can allow the identification of the differences occurring between the 2D-PAGE maps of the four groups of samples). For this purpose, the loading plots of these three components are reported in Figs. 5, 6 and 7, in which the two central maps report the spots as circles centred in the (x,y) -position revealed by PDQuest analysis. The red-coloured spots have a large positive loading on the correspondent PC, whereas the blue ones identify the spots with large negative loadings. Thus the spots are represented in a colour scale in which the increasing red or blue tone is proportional to its loading. The colour of the spots changes from those which show a small influence (small positive or negative loading, light red or light blue) towards those which show a large influence (large positive or negative loadings, dark blue or red); the spots marked as a black circle do not have a relevant loading on the first three PCs. Figure 5 shows the loadings of the spots on PC₁; the red-coloured circles identify those spots showing a larger optical density in the Paca44 cells or spots, which are identified in this cell line but not in the

Fig. 5 Loading plots of PC₁ (coloured maps) with two examples of 2D-PAGE maps of real samples (Paca44 control and treated) characterised by large values of the red-coloured spots (top) and two examples of 2D-PAGE maps of real samples (T3M4 control and treated) characterised by large values of the blue-coloured spots (bottom)



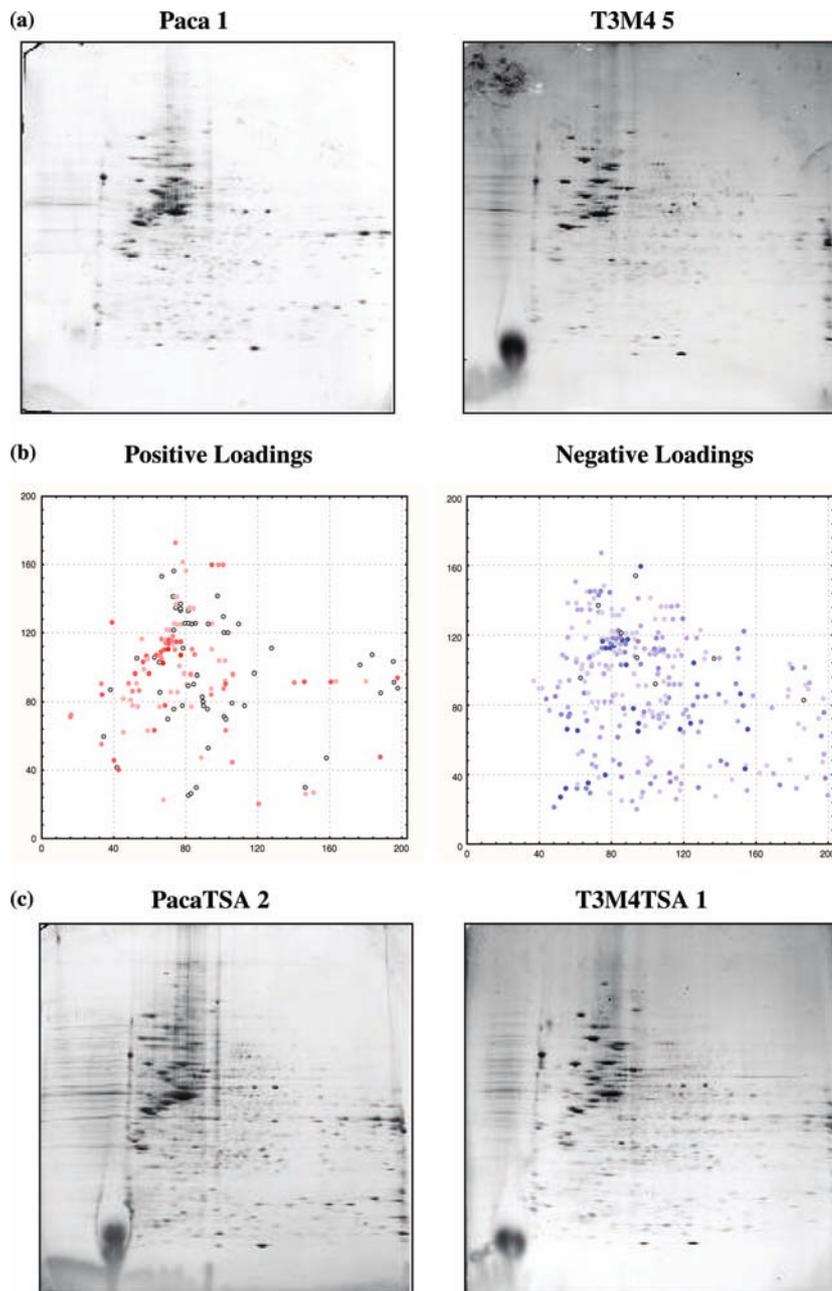
other one; the blue-coloured circles identify the spots that are more intense in the T3M4 cell line or those which are present in this cell line but missing in Paca44 cells. The two examples of real samples at top of Fig. 5 are characterised by large optical densities of the red-coloured spots and small values of the blue-coloured ones (Paca44). At the bottom of Fig. 5 there are two examples of real samples of the T3M4 cell line, characterised by large optical densities of the blue-coloured spots and small values of the red-coloured ones.

Figure 6 represents the loadings plots of the second component: the red-coloured spots identify spots characterised by a larger optical density in the control samples or missing in the TSA-treated ones, whereas the blue-coloured spots represent the opposite situation [i.e. spots with a larger density in the treated samples or

missing in the control ones (of both cell lines)]. Figure 6 (top and bottom) represents two examples of control real samples of the two cell lines (characterised by large values of the red-coloured spots) and two examples of TSA-treated real samples of the two cell lines (characterised by large values of the blue-coloured spots).

The loading plots of the third PC are represented in Fig. 7: the red-coloured circles identify those spots showing a larger optical density in the control T3M4 cells and the TSA-treated Paca44 cells or spots which are absent in the other two classes of samples; the blue-coloured circles instead show the opposite behaviour: they represent the spots which show a larger optical density in the control Paca44 cells and the TSA-treated T3M4 cells or absent in the other two classes. In this case too, an example of a real sample of each class is

Fig. 6 Loading plots of PC₂ (coloured maps) with two examples of 2D-PAGE maps of real samples (Paca44 and T3M4 control) characterised by large values of the red-coloured spots (top) and two examples of 2D-PAGE maps of real samples (Paca44 and T3M4 treated) characterised by large values of the blue-coloured spots (bottom)



presented: the top figure represents two examples of real samples characterised by large values of the red-coloured spots, whereas the bottom figure reports two examples of real samples characterised by large values of the blue-coloured spots.

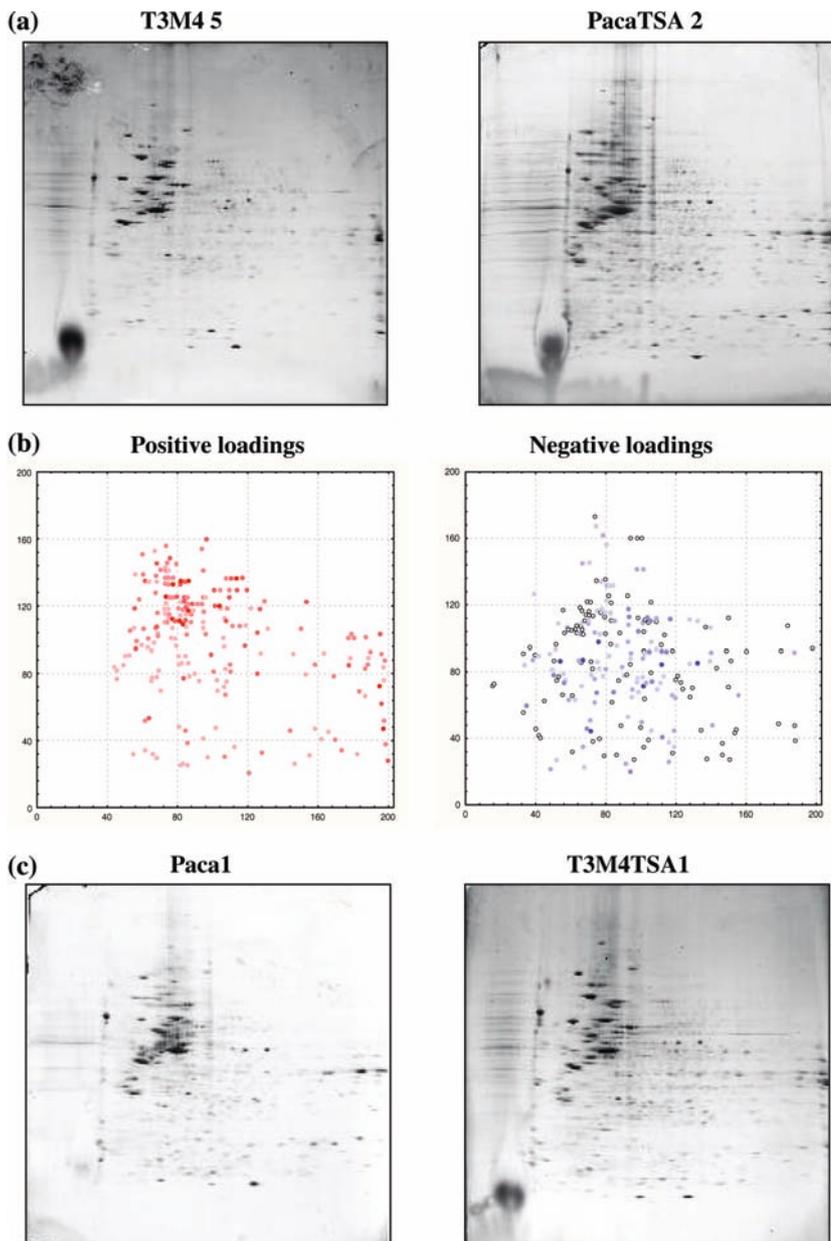
The conclusions derived by means of PCA show a very good agreement with those obtained by PDQuest analysis of the 2D-PAGE maps. The spots identified by PDQuest as the most characterising ones were also identified by means of PCA; in this last case, however, a larger number of spots were identified. Analysis of 2D-PAGE maps by dedicated software usually allows the identification of only those spots which exhibit at least a two-fold variation in the protein content. PCA is a robust tool which allows the detection of variations lower

than the classical two-fold, since the changes due to the natural variability of the experimental steps are explained by the last PCs, which are not taken into account. The total information obtained by PCA is then larger than that obtained by dedicated software; for example, in the present case, the existence of the three patterns identified by PC₁–PC₃ (Figs. 5, 6 and 7,) could not be achieved by conventional PDQuest analysis.

Cluster analysis

Since, as just pointed out, the first three principal components are able to separate the four classes of samples present in the dataset and to account for the reasons of

Fig. 7 Loading plots of PC_3 (coloured maps) with two examples of 2D-PAGE maps of real samples (Paca44 treated and T3M4 control) characterised by large values of the red-coloured spots (top) and two examples of 2D-PAGE maps of real samples (Paca44 control and T3M4 treated) characterised by large values of the blue-coloured spots (bottom)



the differences occurring between them, they are used to perform a cluster analysis to verify how the samples are grouped by means of the first three PCs. The cluster analysis was performed by calculating a dendrogram with the Ward method; the distances were computed using the Euclidean distance. Figure 8 reports the obtained dendrogram; the ordinate label (D_{leg}/D_{max}) $\times 100$ is a percentage dissimilarity scale expressing the linking distance (D_{leg}) of the groups of objects as a fraction of the maximum possible distance (D_{max}). The samples are separated into two main groups, the first comprising the samples belonging to Paca44 cells, and the other given by the samples belonging to T3M4 cells. The two groups are then separated into two sub-groups each, at a normalised distance of more than 40%: both the cell lines considered are correctly separated in control and treated

samples. The dendrogram obtained by considering the first three PCs is then able to correctly separate the four classes of samples, thus confirming the conclusions just derived by means of PCA.

Mass spectrometry

Mass spectrometry analysis was performed only on the Paca44 cell line as a result of the small size of the samples belonging to the T3M4 cell line. Some of the differentially expressed spots between control and treated Paca44 samples were identified by MALDI-TOF, as reported in Table 2. The identified spots are represented in Fig. 9 as black squares, and the SSP is indicated near each spot.

Fig. 8 Dendrogram calculated on the basis of the first three PCs (Ward method, Euclidean distances)

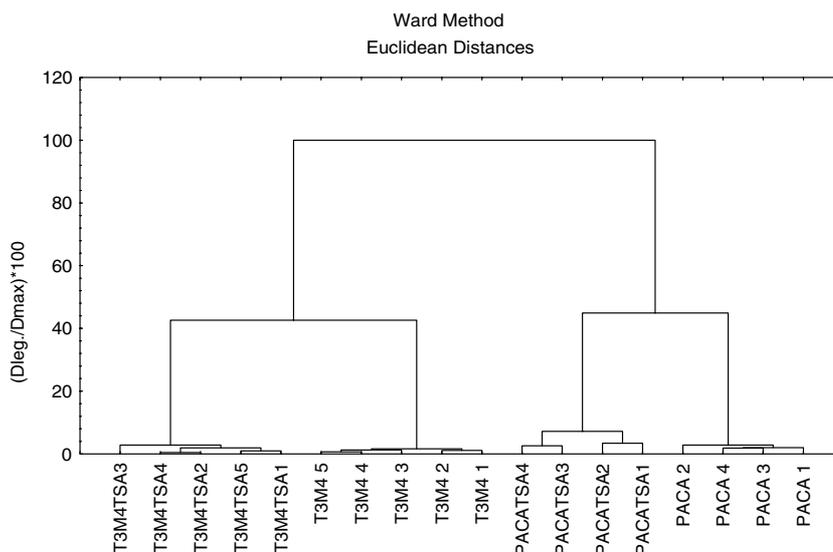
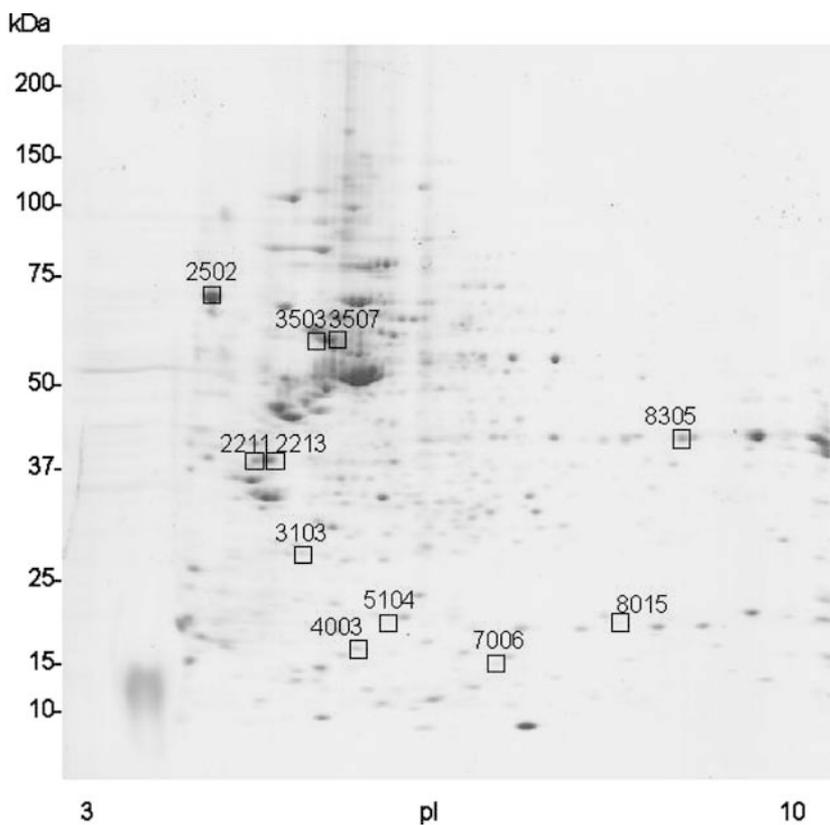


Fig. 9 Spots identified by MS analysis: the *number* near each spot identifies the SSP number



Biological significance of some interesting identified proteins

Among the proteins which were identified by MALDI-TOF analysis, of particular interest are the down-regulated translationally controlled tumour protein (TCTP) as well as the up-regulated protein stathmin (OP18). Their roles will be briefly discussed below.

Translationally controlled tumour protein (TCTP), a three-fold down-regulated polypeptide, seems to be involved in tumour reversion, that is, in the process by which some cancer cells lose their malignant phenotype. In a recent study, Tuynder et al. [44] showed that TCTP is strongly down-regulated in the reversion processes of human leukemia and breast cancer cell lines.

Table 2 Summary of the identified proteins from Paca44 cell line 2D gels. For spot numbers, refer to Fig. 2

Spot SSP	2D gel	Databank									
		Exp. Mr (Da)	Theor. Mr.	Theor. pI	Z-Score	MOWSE-score	Protein name	Accession number	Coverage (%)	No. of peptides	Variation
2211	≈35,000	≈4.5	24,504	4.7	2.37	8.28 E12	Tropomyosin alpha four chain (Tropomyosin 4)	P07226	73.4	25	Decreased 2
2502	≈75,000	≈4.2	46,466	4.3	2.38	1.81 E19	Calreticulin precursor (CRP55) (Calregulin)	P27797	62	25	Decreased 2
2213	≈35,000	≈4.9	32,798	4.7	2.36	4.83 E14	Tropomyosin alpha three chain (Tropomyosin 3)	P12324	65	31	Decreased 3
3103	≈26,000	≈5.4	19,582	4.9	2.41	1.09 E9	Translationally controlled tumor protein (TCTP)	P13693	47	14	Decreased 3
8305	≈37,000	≈9	37,406 and 35,899	9.0 and 8.6	2.4	5.69 E10 and 1.17 E10	Heterogeneous nuclear ribonucleoproteins A2/B1 and Glyceraldehyde 3-phosphate dehydrogenase, liver (EC 1.2.1.12)	P22626 and P04406	54 and 48	16 and 16	Decreased 2
3507	≈55,000	≈5.6	51,736 and 46,142	5.0 and 5.0	2.35	3.38 E12 and 4.19 E9	ATP synthase beta chain, mitochondrial precursor (EC 3.6.3.14) and protein disulfide isomerase A6 precursor (EC 5.3.4.1)	P06576 and Q15084	41 and 41	16 and 13	Decreased 3
3503	≈55,000	≈5.4	51,736 and 49,671	5.0 and 4.8	2.32	4.24 E11 and 1.20 E10	ATP synthase beta chain, mitochondrial precursor (EC 3.6.3.14) and Tubulin beta-1 chain	P06576 and P07437	38 and 42	15 and 14	Decreased 2
5104	≈20,000	≈6.2	16,310 and 17,160	5.5 and 5.7	1.91 and 2.01	2.22 E9 and 5.89 E7	ARP2/3 complex 16 kDa subunit (P16-ARC) and Stathmin (Phosphoprotein p19) (pp19) (Oncoprotein 18)	O15511 and P16949	86 and 77	14 and 24	Increased 8
4003	≈18,000	≈5.8	16,040	5.5	2.34		Deduced protein product shows significant homology to coactosin	NCBIInr: AAA88022.1	57	17	Increased 2
8015	≈20,000	≈8	16,363	8.5	1.98	1.92 E9	UEV protein (ubiquitin-conjugating E2 enzyme variant)	NCBIInr: AAH28673	75	13	Increased 2
7006	≈13,500	≈6.5	13,654	6.4	2.38	1.02 E6	Hint Protein	P49773	59	6	Increased 3

Matching peptides versus total number of peptides submitted to database search, sequence coverage and score resulted for peptide fingerprinting matching are listed. Protein accession number, theoretical pI and Mr were obtained from the SWISS-PROT and

NCBI databases. MOWSE and Z-Scores (output of the identification softwares ProteinProbe and ProFound, respectively) are measures of the statistical significance of the identification hits

Stathmin (Oncoprotein 18, OP18) was eight-fold up-regulated by the TSA treatment. Stathmin is a p53-regulated member of a novel class of microtubule-destabilizing proteins known to promote microtubule depolymerization during interphase and late mitosis [45]. Thus, high levels of stathmin could induce growth arrest at the G2 to mitotic boundary [46, 47]. This suggests a cell cycle arrest at the G2 phase of Paca44 cell treated with TSA. Due to its effect of inhibiting cell proliferation via a mitotic block, the up-regulation of stathmin reported here appears to be consistent with the anti-tumoural activity of TSA.

Conclusions

Principal component analysis is applied here to a dataset comprising by 18 samples belonging to control (untreated) and drug-treated pancreatic human cancer cells;

the samples belong to two different cell lines: Paca44 and T3M4. PCA turned out to be a successful tool for the identification of the classes of samples present in the dataset; moreover, the loadings analysis allowed the identification of the regulatory spots, which characterise each group of samples. Thus, the treatment of both cell lines with Trichostatin A showed an appreciable effect on the proteomic pattern of the control samples. The separation of the samples into four groups by mean of the first three PCs was also confirmed by cluster analysis. The conclusion driven by PCA resulted in good agreement with those obtained from the application of the differential analysis provided by PDQuest.

The MALDI-TOF analysis performed on the Paca44 cell line allowed the identification of some of the spots differentially expressed in control versus treated Paca44 samples. The biological significance of some of the proteins differentially expressed upon TCA treatment is discussed.

Acknowledgements Supported by grants from AIRC (Associazione Italiana Ricerca sul Cancro, Milano, Italy), FIRB 2001 (No. RBNF01KJHT), MIUR (Ministero dell'Istruzione, dell'Università e della Ricerca, Rome, Italy; COFIN 2003), Fondazione Cassa di Risparmio di Verona and the European Community, Grant No. QLG2-CT-2001-01903 and No. QLG-CT-2002-01196.

References

- Righetti PG, Stoyanov A, Zhukov M (2001) The proteome revisited: theory and practice of all relevant electrophoretic steps. Elsevier, Amsterdam
- Wilkins MR, Williams KL, Appel RD, Hochstrasser DF (1997) Proteome research: new frontiers in functional genomics. Springer, Berlin Heidelberg New York
- Fountoulakis M, Schlaeger EJ (2003) *Electrophoresis* 24(1–2):260–275
- Kim J, Kim SH, Lee SU et al (2002) *Electrophoresis* 23:4142–4156
- Gromov PS, Ostergaard M, Gromova I et al (2002) *Prog Biophys Mol Biol* 80:3–22
- Dwek MV, Rawlings SL (2002) *Mol Biotechnol* 22:139–152
- Kostanjevecki V, Lannoy K, Daniels A et al (2002) *Neurobiol Aging* 23 (suppl 1):1383
- Castegna A, Aksenov M, Thongboonkerd V et al (2002) *J Neurochem* 82:1524–1532
- Sickmann A, Marcus K, Schafer H et al (2001) *Electrophoresis* 22:1669–1676
- Sinha P, Kohl S, Fischer J et al (2000) *Electrophoresis* 21:3048–3057
- Tsuji T, Shiozaki A, Kohno R et al (2002) *Neurochem Res* 27:1245–1253
- Ryan TE, Patterson SD (2002) *Trends Biotechnol* 20 (suppl S):S45–S51
- Cacabelos R (2002) *Ann Med* 34:357–379
- Poirier F, Pontet M, Labas V et al (2001) *Electrophoresis* 22:1867–1877
- Steiner S, Witzmann FA (2000) *Electrophoresis* 21:2099–2104
- Westergren-Thorsson G, Malmstrom J, Marko-Varga G (2001) *J Pharm Biomed Anal* 24:815–824
- Bhatia K, Lord R, Stanton P (2002) *Eur J Cancer* 28:S156
- Hoffmann F, Kriegel K, Wenk C (1999) *Discrete Appl Math* 93:75–88
- Lee KH (2001) *Trends Biotechnol* 19:217–222
- Massart DL, Vandeginste BGM, Deming SN, Michotte Y, Kaufman L (1988) *Chemometrics: a textbook*. Elsevier, Amsterdam
- Brereton RG (1990) *Chemometrics: application of mathematics and statistics to laboratory systems*. Ellis Norwood, New York
- Box GEP, Hunter WG, Hunter JS (1978) *Statistics for experimenters*. Wiley, New York
- Mardia KV, Kent JT, Bibby JM (1979) *Multivariate analysis*. Academic, London
- James M (1985) *Classification algorithms*. Collins, London
- Cox TF, Cox MAA (1994) *Multidimensional scaling*. Chapman and Hall, London
- Marengo E, Robotti E, Gianotti V, Righetti PG (2003) *Ann Chim (Rome)* 93:105–116
- Marengo E, Robotti E, Righetti PG, Antonucci F (2003) *J Chromatogr A* 1004:13–28
- Marengo E, Robotti E, Gianotti V, Righetti PG, Domenici E, Ceconi D (2003) *Electrophoresis* 24:225–236
- Marengo E, Leardi R, Robotti E, Righetti PG, Antonucci F, Ceconi D (2003) *J Proteome Res* 2:351–360
- Anderson NL, Hofmann JP, Gemmill A, Taylor J (1984) *Clin Chem* 30:2031–2036
- Tarroux P, Vincens P, Rabilloud T (1987) *Electrophoresis* 8:187–199
- Leonard RB, Mayer J, Sasser M, Woods ML, Mooney BR, Brinton BG, Newcomb-gayman PL, Carroll KC (1995) *J Clin Microbiol* 33:2723–2727
- Johansson ML, Quednau M, Ahrne S, Molin G (1995) *Int J Syst Bacteriol* 45:670–675
- Couto MMB, Vogels JTWE, Hofstra H, Husiintveld JHJ, Vandervossen JMBM (1995) *J Appl Bacteriol* 79:525–535
- Boon N, De Windt W, Verstraete W, Top EM (2002) *FEMS Microbiol Ecol* 39:101–112
- Alika JE, Akenova ME, Fatokun CA (1995) *Genet Res Crop Evol* 42:393–399
- Anderson NL, EsquerBlasco R, Richardson F, Foxworthy P, Eacho P (1996) *Toxicol Appl Pharm* 137:75–89
- Vohradsky J (1997) *Electrophoresis* 18:2749–2754
- Kovarova H, Radzioch D, Hajduch M, Sirova M, Blaha V, Macela A, Stulik J, Hernychova L (1998) *Electrophoresis* 19:1325–1331
- Lacroix-Desmazes S, Bayry J, Misra N, Horn MP, Villard S, Pashov A, Stieltjes N, d'Oiron R, Saint-Remy J, Hoebeke J, Kazatchkine MD, Kaveri SV (2002) *New Engl J Med* 34:662–667
- Dewettinck K, Dierckx S, Eichwalder P, Huyghebaert A (1997) *Lait* 77:77–89
- Kovarova H, Hajduch M, Korinkova G, Halada P, Krupickova S, Gouldsworthy A, Zhelev N, Strnad M (2000) *Electrophoresis* 21:3757–3764
- Massart DL, Kaufman L (1983) Hierarchical clustering methods. In: Elving PJ, Winefordner JD (eds) *The interpretation of analytical chemical data by the use of cluster analysis*. Wiley, New York
- Tuynder M, Susini L, Prieur S, Besse S, Fiucci G, Amson R, Telerman A (2002) *Proc Natl Acad Sci USA* 99:14976–14981
- Alli E, Vash-Babula J, Yang JM, Hait WN (2002) *Cancer Res* 62:6864–6869
- Cassimeris L (2002) *Curr Opin Cell Biol* 14:18–24
- Mistry SJ, Atweh GF (2002) *Mt Sinai J Med* 69:299–304

Application of Fuzzy Logic Principles to the Classification of 2D-PAGE Maps Belonging to Human Pancreatic Cancers Treated with Trichostatin-A

Emilio Marengo*, Elisa Robotti
Department of Environmental and Life
Sciences - University of Eastern Piedmont -
Spalto Marengo 33 - 15100 Alessandria – Italy
– emilio.marengo@mfn.unipmn.it

Daniela Cecconi
Department of Industrial
Biotechnologies -
University of Verona -
Strada le Grazie 15 -
37134 Verona – Italy

Aldo Scarpa
University of Verona,
Department of
Pathology, Section of
Anatomical
Pathology, 37134
Verona, Italy

Pier Giorgio Righetti
Department of Industrial
Biotechnologies -
University of Verona -
Strada le Grazie 15 -
37134 Verona – Italy

Abstract - 2D gel-electrophoresis is the most widespread technique applied in the study of cellular protein patterns. The complexity of this technique is reflected in the production of complex samples (2D gel maps) which are very difficult to compare. In this paper a method is proposed based on fuzzy logic principles. The method can be summarised in four steps: 1) digitalisation of the image; 2) fuzzyfication of the digitalised image in order to consider the natural variability of the applied technique; 3) comparison of the samples and calculation of a similarity matrix; 4) application of Multidimensional Scaling technique to the similarity matrix, in order to verify the presence of clusters of samples. The investigated dataset consists of 18 samples belonging to two different cell lines (Paca 44 and T3M4) of control (untreated) and drug-treated pancreatic ductal carcinoma cells.

I. INTRODUCTION

2D gel electrophoresis is certainly the most widespread technique in the field of proteomics [1,2]. This technique achieves a two-dimensional separation of the cellular protein extract: the first dimension performs the separation according to the isoelectric point (through a pH gradient), while the second dimension performs the separation according to the molecular mass (through a porosity gradient or a highly sieving, constant concentration gel). The final result is a two-dimensional map, called “2D-PAGE” (Poly Acrylamide Gel Electrophoresis), with the separated proteins appearing as spots spread all over the gel matrix. The particular complexity of the specimen (biological) is reflected in the complexity of the resulting 2D-maps which may contain even thousands of spots, different for structure and size.

The great advantage of the 2D-PAGE technique is, together with the achievement of the proteins separation, the achievement of a final result which may be considered as a

“snapshot” of the protein content of the investigated cell at a given point of its life cycle.

2D gel-electrophoresis has recently gained increasing fortune, thanks to the development of the researches in both proteomic and genomic fields. In fact, it is commonly acknowledged that the physiological state of a particular cell or tissue is directly related to its protein content: the onset of a particular disease may cause differences in the protein pattern characterising the pathological tissue; the differences occurring may consist of changes in the relative abundance or even in the appearance/disappearance of some proteins [3-7]. This aspect opens the way to use 2D-PAGE maps for both diagnostic and prognostic purposes. In a similar way, drug-design studies can be performed [8-9]: 2D gel-electrophoresis may be applied to investigate the effect played by a particular drug on the cellular protein profile (e.g., up- or down- regulation, appearance/disappearance of pathological chains).

In the just highlighted panorama, the comparison of 2D-PAGE maps belonging to healthy and diseased subjects or belonging to diseased and drug-treated individuals becomes almost a necessity. Unfortunately, the comparison of 2D-PAGE maps is not a trivial problem to solve; its difficulty is due to several aspects:

- a high sample complexity, with maps containing also thousands of spots;
- a highly complex sample pre-treatment which may cause the appearance of spurious spots on the map, due to accidental chemical reactions;
- a large number of experimental factors influencing the electrophoretic run (polymerisation conditions, staining and de-staining times and temperatures etc...), which produce spots not due to systematic variations but to accidental ones;

- the differences, sometimes very small, which characterise the 2D-PAGE maps belonging to treated and reference samples, that are difficult to identify in complex maps.

In the most of cases, the comparison of 2D-PAGE maps is achieved by means of specific softwares (e.g. Melanie III or PDQuest) [10-12]. These methods are based on a multi-step procedure:

- 1) the spots present on the maps are revealed and quantified independently for each map;
- 2) the 2D-PAGE images are previously aligned, in order to reduce all images to the same size. To complete this step, the choice of at least two spots of sure identification in all the maps is required. The maps are then matched one to the other on the basis of the position of the landmark spots. This step performs a distortion of the images, in fact the spots are shrunk or widened in order to obtain the optimal matching;
- 3) the common information, spots present in all the matched maps, is identified. When replicate maps are compared, this step produces a "synthetic" map which contains only the spots present in all the compared maps (the common information).

This procedure produces large spot volume datasets, usually treated by means of multivariate statistical tools, like Principal Component Analysis (PCA) [13-14]. PCA has found one of its first applications in the middle eighties by Anderson *et al.* [15] in USA and by Tarroux *et al.* [16] in France. Recently, it has been applied to the study of the proteomic patterns of several biological tissues [17-18] and of DNA and RNA fragments [19-20]. One of its recent applications is also the study of the effects of drugs on diseases [21].

These approaches certainly are very powerful and effective but they are all based on the previous image processing by dedicated softwares. These softwares, as just described, involve steps in which the images are distorted and stretched. It is surely preferable to work directly on the images, without the interference of human contribution (e.g. the choice of the sure spots in step 2 of the software procedure).

Recently, a new method for the comparison of the proteomic pattern of classes of 2D-PAGE maps, based on both fuzzy logic principles and classification methods, has been developed by our research group [22-23]: this method has also been coupled to MDS for the study of 2D-maps from control and diseased individuals [24]. Different cellular protein patterns have also been investigated by means of Three-way Principal Component Analysis [25].

In the present paper the approach based on fuzzy logic is applied to a dataset constituted by 18 samples, belonging to two different cell lines (Paca44 and T3M4) of pancreatic human cancer, before and after the treatment with a new drug (Trichostatin-A). This approach is focused on the evaluation of the existence of differences between diseased and treated samples.

II. THEORY

The applied procedure is based on a multi-step analysis:

- digitalisation of the samples;
- fuzzyfication of the digitalised samples;
- calculation of the similarity indices between the samples;
- treatment of the similarity matrix by means of Multidimensional Scaling (MDS).

A. Digitalisation

In this case, as in the common practice, each 2D-PAGE was scanned by a densitometer which reveals the optical density in each point of the 2D-map. The grey-scale images thus obtained can be transformed into a grid of a given size. In this case, the digitalisation of the maps is obtained by a grid with a pass of 0.001×0.001 m, that produces a map of 200×200 cells. This step is achieved by transforming each map into an image of 200×200 pixels. By this procedure, each 2-D map is converted into a matrix, corresponding to the grid, containing the values of the grey-scale intensity in each pixel. The values smaller than 0.35 were cut off and substituted by null values, in order to eliminate the effect due to the background. When the signal was greater than the selected threshold, the cell value was put to 1. In figure 1, a sample before and after the digitalisation step is presented.

B. Fuzzyfication

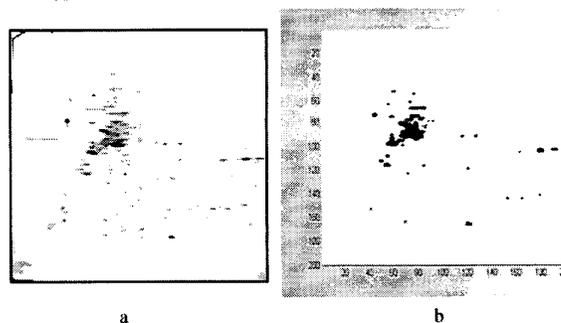


Fig. 1. Sample PACA1 before (a) and after (b) the digitalisation step

The fuzzyfication procedure represents the fundamental step of the method. Because of the high variability of the experimental result obtained by 2D gel-electrophoresis, the position, size and shape of the spots on the maps may change even sensitively in replicate maps of the electrophoretic run performed on the same biological sample. This variability can be treated by mean of fuzzy logic principles.

The presence of a spot in position x_i, y_j is substituted by a gaussian probability distribution, so that the probability to find a signal in the neighbour cell in position x_k, y_l , can be calculated by the two-dimensional gaussian function:

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

where σ_x and σ_y are two constants corresponding to the standard deviation of the gaussian function along each of the two dimensions and the correlation between the two dimensions x and y is considered negligible, since a complete independence of the two electrophoretic runs is expected. The two parameters σ_x and σ_y are considered at different levels ($\sigma_x \neq \sigma_y$) according to the assumption that one of the two dimensions presents a larger variability. The first dimension (isoelectric point) is run on commercial strips, while the second one (molecular mass) is run on gels directly polymerised in our lab. The second dimension presents typically a higher variability, estimated to be 50% larger than that on the first dimension. Since molecular mass is represented in our case by the rows of the digitalised matrix, $\sigma_x = 1.5 * \sigma_y$.

Obviously, increasing the value of both σ_x and σ_y shall extend the effect of each spot at a larger distance, making the virtual map more fuzzy. The choice of the gaussian function was dictated by the fact that each spot can be properly described by a function with the highest intensity in the centre of the spot itself and decreasing values for increasing distance from the spot.

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 neighbour cells x_i, y_j containing spots:

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

The sum runs on all the cells of the grid.

By this procedure, each digitalised map is transformed into a “fuzzy” map, where each cell contains the sum of the influence of all the spots present on the digitalised map, which represents a sort of probability of the presence of a spot in the correspondent position. The application of this procedure performs a fuzzy change of the spots, whose position and size become blurred, according to a gaussian dependence. The main advantages with respect to the traditional approach are: 1) the elimination of the operator’s intervention in the matching step; 2) the possibility of retaining each original replicate maps instead of replacing them by the “synthetic” map containing only the information common to all replicates.

C Similarity Index

The similarity indices are calculated for each pair of samples, on the basis of the “fuzzy” maps. The superposition of the

two “fuzzy” grids k and l allows the computation of the common signal (SC_{kl}) and of the total signal (ST_{kl}):

$$SC_{kl} = \sum_{i=1,n} \min(S_i^k, S_i^l) \quad (3)$$

$$ST_{kl} = \sum_{i=1,n} \max(S_i^k, S_i^l) \quad (4)$$

The similarity index can then be computed as the ratio between SC_{kl} and ST_{kl} :

$$S_{kl} = \frac{SC_{kl}}{ST_{kl}} \quad (5)$$

Obviously, a similarity index which equals 1 refers to two identical maps, while a null value refers to two maps showing no common structure.

D. Multidimensional Scaling

Multidimensional scaling is a branch of multivariate data analysis geared towards dimensional reduction and graphical representation of data. Given a set of n objects and a measure of their similarity S_{ij} , Multidimensional Scaling (MDS) consists in the search for a low dimensional space in which the objects are represented by points in the space and such that the distances between the points match as much as possible with the original similarities of the objects [26]. The space is usually Euclidean, but this is not a constraint. There are several different approaches to MDS, depending on the notion of *matching* of the similarities, on the metrics, on the method used to compute the similarities [27] and on the way the point configuration is obtained (Shoenberg [28], Young and Householder [29], Gower [30], Shepard [31,32], Kruskal [33,34]).

In the present case the calculation of MDS from the 2D-map similarities was performed by using the Kruskal iterative method. The search for the co-ordinates is performed by the steepest descent minimisation algorithm, the target function being the so called Stress (S), which is a measure of the ability of the configuration of points to simulate the distance matrix:

$$S = \sum [d_{i,j} - \delta_{i,j}]^2 \quad (6)$$

where $d_{i,j}$ is the measured distance between objects i and j , and $\delta_{i,j}$ is the same distance approximated by the MDS algorithm.

Once the co-ordinates have been obtained, they can be used for a multivariate statistical analysis. The results are often self-evident, so that a visual inspection is sufficient for obtaining the final results.

III. EXPERIMENTAL

The dataset consists of 18 2D-maps, divided in four groups:

- 4 replicate 2D-maps of a Paca44 cell line pool;
- 5 replicate 2D-maps of a T3M4 cell line pool;
- 4 replicate 2D-maps of a Paca 44 cell pool treated for 48 hours with Trichostatin-A;
- 5 replicate 2D-maps of a T3M4 cell pool treated for 48 hours with Trichostatin-A.

4-5 replicas of 2D-maps for each sample were considered sufficient for reducing variability due to experimental errors, since pools of cell lines (grown under the same conditions) were used. This procedure is common practice in today's proteome analysis. Figure 2 present an example of 2D-maps belonging to each of the 4 groups.

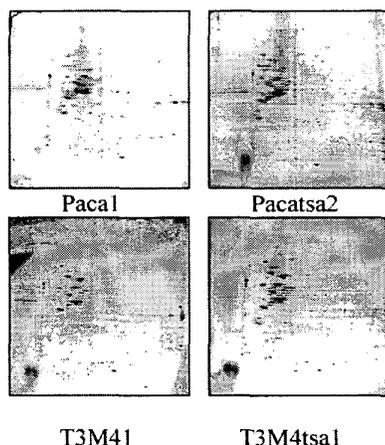


Fig.2. An example of 2D-map for each group of cells.

- *Cell treatment with TSA and lysis.* Paca44 and T3M4 cells were grown in RPMI 1640 supplemented with 20 mM glutamine and 10% (v/v) FBS (BioWhittaker, Italy) and were incubated at 37°C with 5% (v/v) CO₂. Subconfluent cells were treated with 0.2 mM TSA for 48 h. Protein extraction from cells was performed with lysis buffer (40 mM Tris, 1% v/v NP40, 1 mM Na₃VO₄, 1 mM NaF, 1 mM PMSF, protease inhibitor cocktail). Cells were left in lysis buffer for 30 min in ice. After centrifugation at 14.000 x g at 4°C the protein pellets were collected and stored at -80°C until used.

- *Two-dimensional gel electrophoresis.* Seventeen cm long, pH 3-10 immobilized pH gradient strips (IPG; Bio-Rad Labs., Hercules, CA, USA) were rehydrated for 8 h with 450 µL of 2-D solubilizing solution (7 M urea, 2 M thiourea, 5 mM tributylphosphine, 40 mM Tris and 20 mM iodoacetamide) containing 2 mg/mL of total protein from sample cells. Isoelectric focusing (IEF) was carried out with a Protean IEF Cell (Biorad, Hercules, CA, USA), with a low initial voltage and then by applying a voltage gradient up to

10000 V with a limiting current of 50 µA. The total product time x voltage applied was 70000 Vh for each strip and the temperature was set at 20°C. For the second dimension, the IPGs strips were equilibrated for 26 min by rocking in a solution of 6 M urea, 2% w/v SDS, 20% v/v glycerol, 375 mM Tris-HCl, pH 8.8. The IPG strips were then laid on an 8-18%T gradient SDS-PAGE with 0.5% w/v agarose in the cathode buffer (192 mM glycine, 0.1% w/v SDS and Tris to pH 8.3). The anodic buffer was a solution of 375 mM Tris HCl, pH 8.8. The electrophoretic run was performed by setting a current of 2 mA for each gel for 2 hours, then 5 mA/gel for 1 h, 10 mA/gel for 20 h and 20 mA/gel until the end of the run. During the whole run the temperature was set at 11°C. Gels were stained overnight with Colloidal Coomassie blue (0.1% w/v Coomassie Brilliant Blue G, 34% v/v methanol, 3% v/v phosphoric acid and 17% w/v ammonium sulphate), while destaining was performed with a solution of 5% v/v acetic acid until a clear background was achieved.

- *Software.* The 2D-PAGE maps were scanned with a GS-710 densitometer (Bio-Rad Laboratories, Hercules, CA, USA), and analyzed with the software PDQuest Version 6.2 (Bio-Rad Laboratories, Hercules, CA, USA). The fuzzy maps and the similarity matrices were calculated by an algorithm programmed in Visual Basic 6.0 (Microsoft) and Matlab 6.5 (The Mathworks Inc.). MDS calculations were performed by the STATISTICA software (Statsoft inc., Ver. 7.1). Graphical representations were performed with both STATISTICA and Matlab).

IV. RESULTS AND DISCUSSION

The effect of the changes of the fuzzyfication parameters σ_x and σ_y was studied by changing the level of the two parameters. The investigated couples of values are reported in table 1. The values of the two parameters determine the region of influence of each spot: to a larger standard deviation along one of the electrophoretic directions corresponds a larger uncertainty of the spot position, shape and size along the same direction.

TABLE 1

σ_x	σ_y
0.00	0.00
0.75	0.50
1.50	1.00
2.25	1.50
3.00	2.00

Figure 3 represents an example of fuzzyfication of one of the real samples (PACA1) at the four levels of σ_x and σ_y .

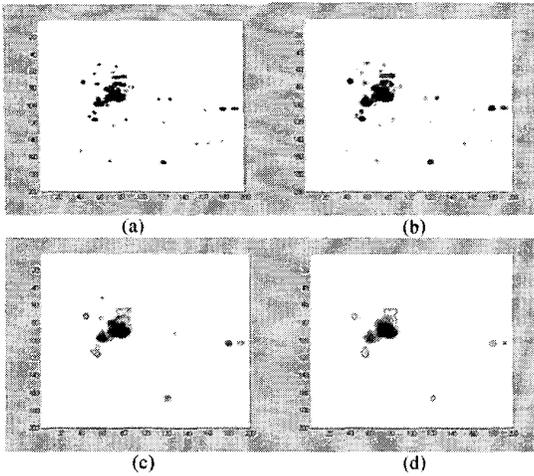


Fig. 3. Sample Pacal at the four levels of σ_x and σ_y : 0.75-0.50 (a), 1.50-1.00 (b), 2.25-1.50 (c), 3.00-2.00 (d).

The area of influence of each spot increases while moving towards larger values of the parameters, thus producing more confused maps.

The high complexity of the dataset, due to the presence of two kinds of information (cell line type and treatment with the active principle) makes it very difficult to separate the four groups according to their pattern. Here, only the results for the settings of the two parameters which gave the best results in the separation of the samples, together with the case in which no fuzzyfication is applied, are presented. The best couple resulted to be the third one ($\sigma_x = 2.25$ and $\sigma_y = 1.50$), showing that a fuzzyfication step is required in order to better identify the groups of samples, but a too large fuzzyfication finishes to mask their differences, with the nearest spots merging one into the other.

The matching of the maps and the calculation of the similarity matrices is performed on both the fuzzy maps (at the different levels of the parameters) and on the digitalised maps without the application of the fuzzyfication step. Multidimensional Scaling is then performed on the resulting similarity matrices for all cases. Figure 4 shows the final configurations in the case where no fuzzyfication is applied and in the best fuzzyfied case.

When no fuzzyfication is applied, the final configuration does not allow a clear separation between the groups of samples: the first dimension seems to enable the separation between diseased and drug-treated samples, while the second dimension seems to enable the differentiation between the two cell lines: however, many samples are present in the wrong position.

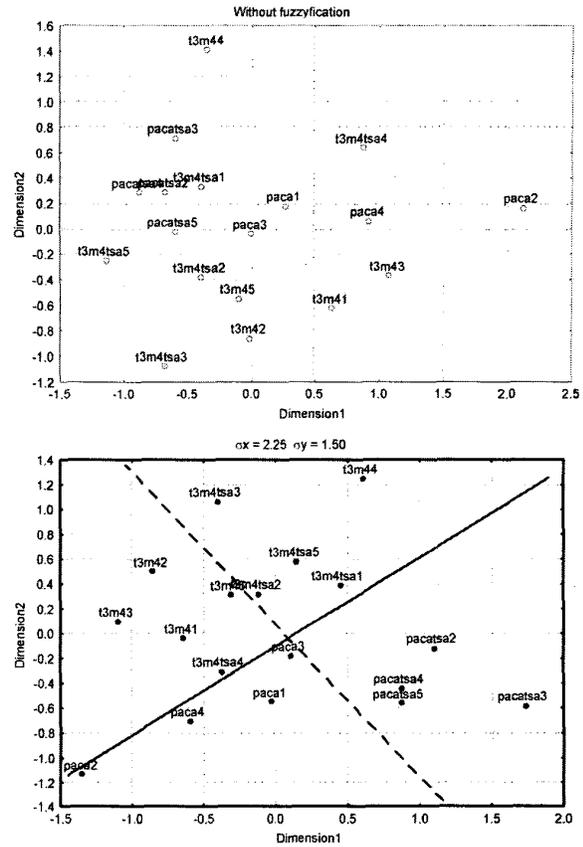


Fig. 4. Final MDS configurations for the original dataset without fuzzyfication (first graph) and with the best fuzzyfication parameters (second graph).

In the case where the best results are obtained ($\sigma_x = 2.25$ and $\sigma_y = 1.50$) instead, a clearer separation can be identified. The continuous line present in figure 4 allows the separation of the samples between the two cell lines. The dotted line instead allows the separation of the samples between diseased and drug-treated ones, with the only exception of samples T3M44 and T3M4tsa4, which present the opposite behaviour. It is also to be noticed that the samples of the PACA44 cell line and the diseased samples of the T3M4 cell line present a smaller variability, with respect to the other two groups of samples, in facts they appear quite close to the samples of the same group.

V. CONCLUSIONS

Fuzzy logic principles and MDS are applied here to a dataset constituted by 18 samples belonging to control (untreated) and drug-treated pancreatic human cancer cells; the samples belong to two different cell lines: Paca44 and T3M4. The fuzzyfication procedure applied to each 2D-PAGE map resulted to be necessary in order to distinguish

the groups of samples present in the dataset. An MDS procedure applied on the dataset where no fuzzyfication was applied did not allow the identification and separation of the four groups present. The best results were obtained with the parameters σ_x and σ_y settled at 2.25 and 1.50 respectively. With this setting, MDS allowed a quite good identification of the four groups of samples.

ACKNOWLEDGEMENTS

Prof. P.G. Righetti is supported by Grants from AIRC (Associazione Italiana Ricerca sul Cancro, Milano, Italy), from FIRB 2001 (No. RBNF01KJHT), from MIUR 2003 (COFIN), and from the European Community, Grant N° QLG2-CT-2001-01903.

REFERENCES

- [1] Righetti P.G., Stoyanov A., Zhukov M., *The Proteome revisited: Theory and Practice of All 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., (2003) *Electrophoresis* 24 (1-2): 260-275
- [4] Kim J., Kim S. H., Lee S. U., et al., (2002) *Electrophoresis* 23: 4142-4156
- [5] Dwek M. V., Rawlings S. L., (2002) *Mol Biotechnol* 22: 139-152
- [6] Kostanjevecki V., Lannoy K., Daniels A., et al., (2002) *Neurobiol Aging* 23 (Suppl. 1): 1383
- [7] Castegna A., Aksenov M., Thongboonkerd V., et al., (2002) *J. Neurochem* 82: 1524-1532
- [8] Ryan T. E., Patterson S. D., (2002) *Trends Biotechnol* 20 (Suppl. S): S45-S51
- [9] Cacabelos R., (2002) *Ann Med* 34: 357-379
- [10] Westergren-Thorsson G., Malmstrom J., Marko-Varga G., (2001) *J. Pharmaceut. Biomed. Anal.* 24: 815-824
- [11] Bhatia K., Lord R., Stanton P., (2002) *Eur. J. Cancer* 28: S156
- [12] Hoffmann F., Kriegel K., Wenk C., (1999) *Discrete Appl. Math.* 93: 75-88
- [13] Massart D. L., Vanderginste B. G. M., Deming S. N., Michotte Y., Kaufman L., *Chemometrics: a textbook*, Elsevier: Amsterdam, 1988
- [14] Box G. E. P., Hunter W. G., Hunter J. S., *Statistics for Experimenters*, Wiley: New York, 1978
- [15] Anderson N. L., Hofmann J. P., Gemmell A., Taylor J., (1984) *Clin Chem.* 30: 2031-2036
- [16] Tarroux P., Vincens P., Rabilloud T., (1987) *Electrophoresis* 8: 187-199
- [17] Kovarova H., Radzioch D., Hajdich M., Sirova M., Blaha V., Macela A., Stulik J., Hernychova L., (1998) *Electrophoresis* 19: 1325-1331
- [18] 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., (2002) *New Engl. J. Med.* 34: 662-667
- [19] Boon N., De Windt W., Verstraete W., Top E. M., (2002) *FEMS Microbiol. Ecol.* 39: 101-112
- [20] Couto M. M. B., Vogels J. T. W. E., Hofstra H., Husiintveld J. H. J., Vandervossen J. M. B. M., (1995) *J. Appl. Bacteriol.* 79: 525-535
- [21] Kovarova H., Hajdich M., Korinkova G., Halada P., Krupickova S., Gouldsworthy A., Zhelev N., Strnad M., (2000) *Electrophoresis* 21: 3757-3764
- [22] Marengo E., Robotti E., Gianotti V., Righetti P. G., (2003) *Ann. Chim. (Rome)* 93: 105-116
- [23] Marengo E., Robotti E., Righetti P. G., Antonucci F., (2003) *J. Chromatogr. A* 1004: 13-28
- [24] Marengo E., Robotti E., Gianotti V., Righetti P. G., Domenici E., Ceconi D., (2003) *Electrophoresis* 24: 225-236
- [25] Marengo E., Leardi R., Robotti E., Righetti P. G., Antonucci F., Ceconi D., (2003) *J. Proteome Research* 2: 351-360
- [26] Tulp, A., Verwoerd, D., Neeffjes, J., *J. Chromatogr B* 1999, 722, 141-51
- [27] Young, G., Householder, A. S., *Psychometrica* 1930, 3, 19-22
- [28] Cox, T. F., Cox, M. A. A., *Multidimensional Scaling*, Chapman & Hall, London 1994
- [29] Schoenberg, I. J., *Ann. Math.* 1935, 36, 724-32
- [30] Young, G., Householder, A. S., *Psychometrika* 1938, 3, 19-22
- [31] Gower, J. C., *Biometrika* 1966, 53, 325-38
- [32] Shepard, R. N., *Psychometrika* 1962, 27, 125-40
- [33] Shepard, R. N., *Psychometrika* 1962, 27, 219-46
- [34] Kruskal, J. B., *Psychometrika* 1964, 29, 1-27

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

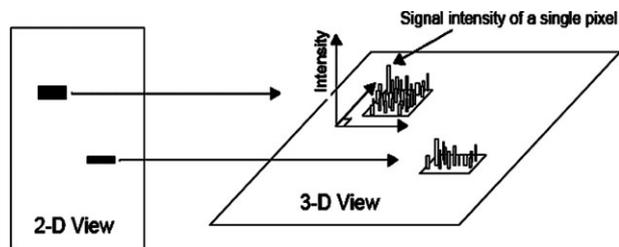


Figure 1. Representation of the pixels in two digitally imaged bands in a gel, as a 2-D (left) and three-dimensional (right) view.

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

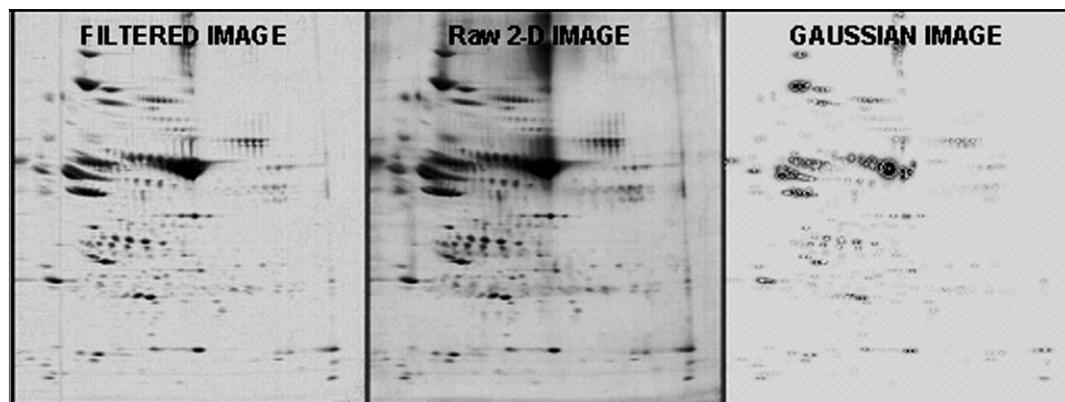


Figure 2. Representation of a filtered image, a raw 2-D scan and a Gaussian image. The data shown is from a 2-D map of rat serum.

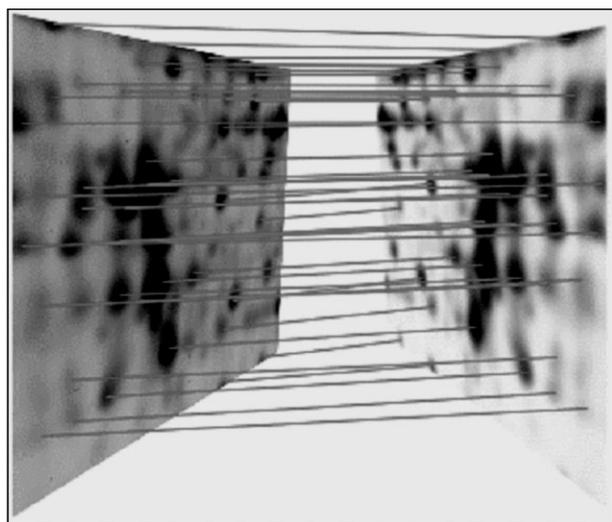


Figure 3. Schematic drawing on how protein spots from two gels are matched with PDQuest.

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 defuzzyfication, (iii) refuzzyfication.

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 defuzzyfication

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 defuzzyfication of each map, is focused on the elimination of the sensitivity to the destaining protocol.

3.3 Refuzzyfication

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

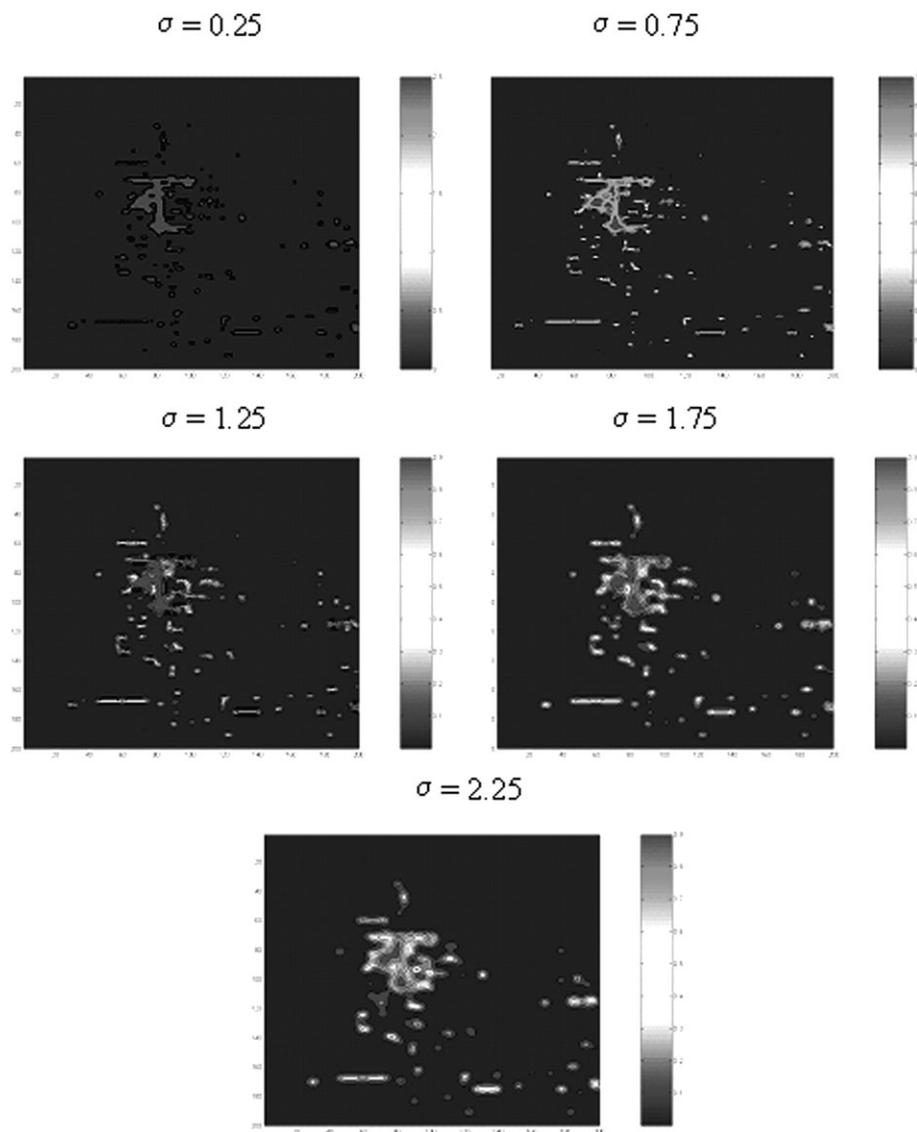


Figure 4. Example of fuzzy maps obtained for a control human lymph-node for different values of the σ parameter. Reprinted from [27] with permission.

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].

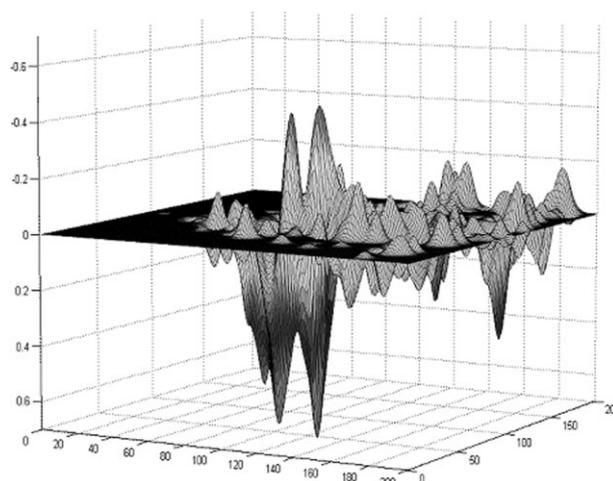


Figure 5. Display of the fuzzy map obtained as the difference of the mean healthy and pathological fuzzy maps. Reprinted from [27] with permission.

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

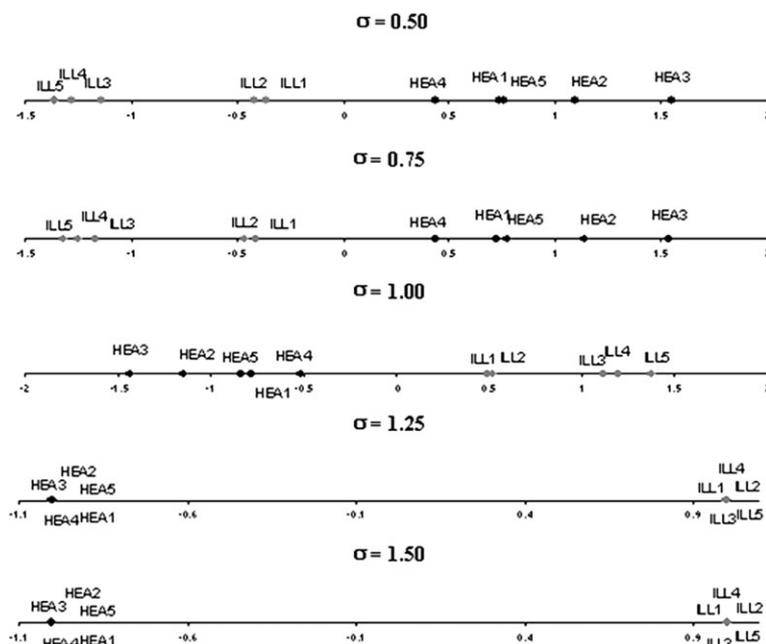


Figure 6. Rat sera dataset: MDS final configurations for different values of the σ parameter. Reprinted from [28] with permission.

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:

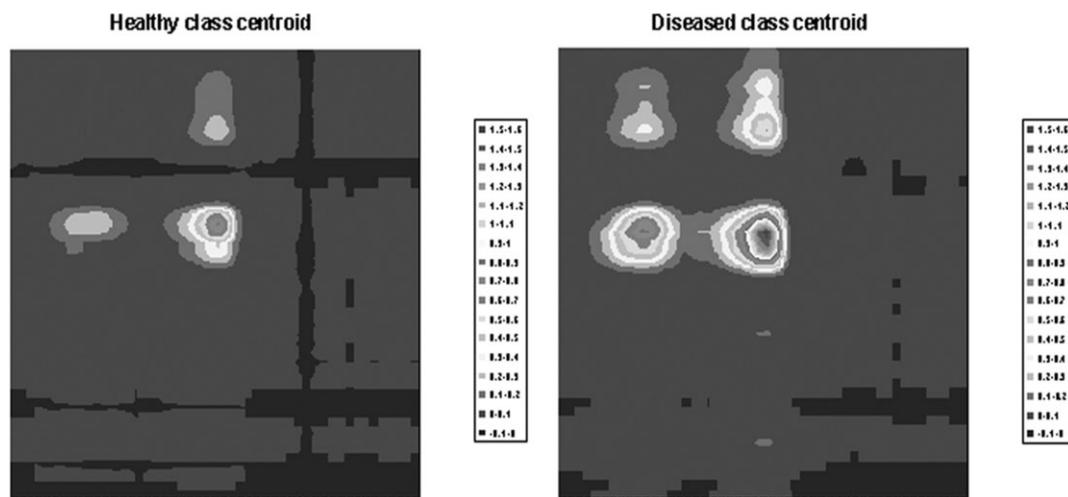
$$x_k(i,j) = \frac{X_k(i,j)}{\max(X_k)} \quad (9)$$

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.

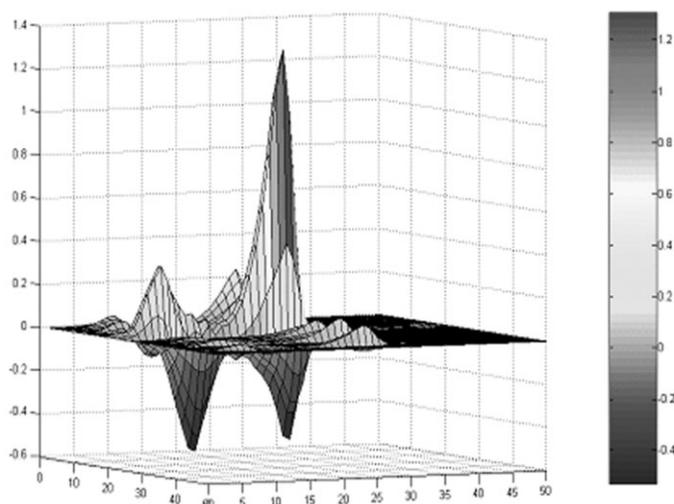
Three-way PCA, based on the Tucker-3 model [39–45], has been used for the identification of classes of samples present in the two datasets. Three-way PCA allows the three-way structure of the dataset which can be considered as a parallelepiped of size $I \times J \times K$ (conventionally defined as 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 uninformative 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

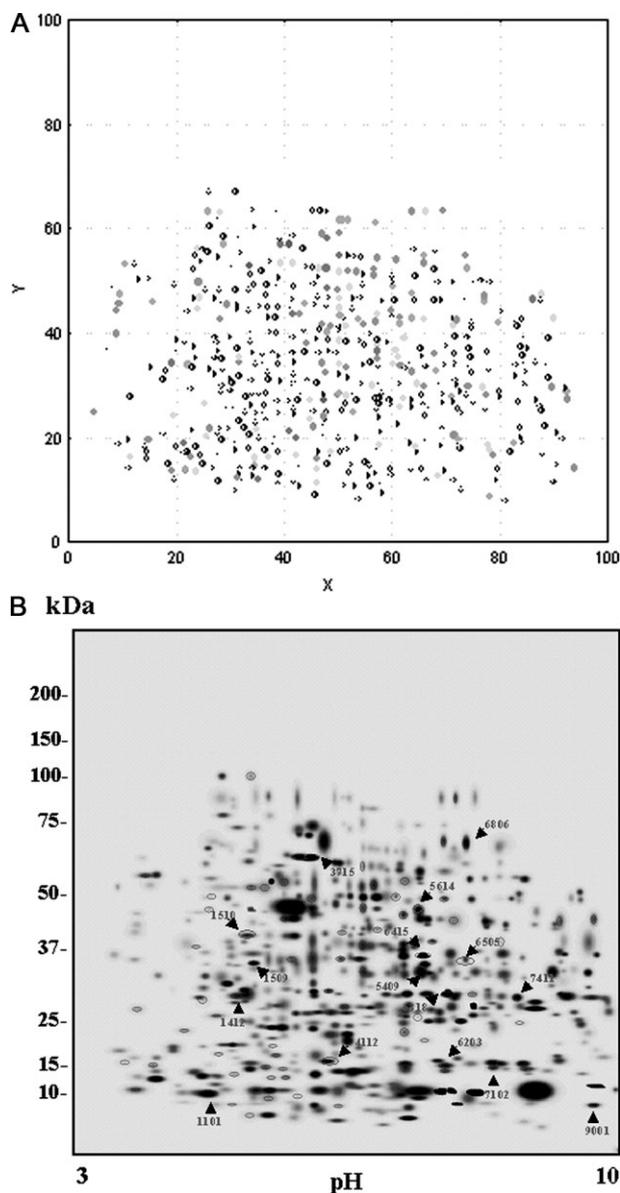


Figure 8. Data reduction of control *vs.* neuroblastoma murine samples *via* (A) combined PCA-SIMCA data handling and (B) PDQuest analysis. In (A) the red spots of different color intensity represent polypeptide chains which are differentially expressed in the two sets of samples; the same spots, in (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

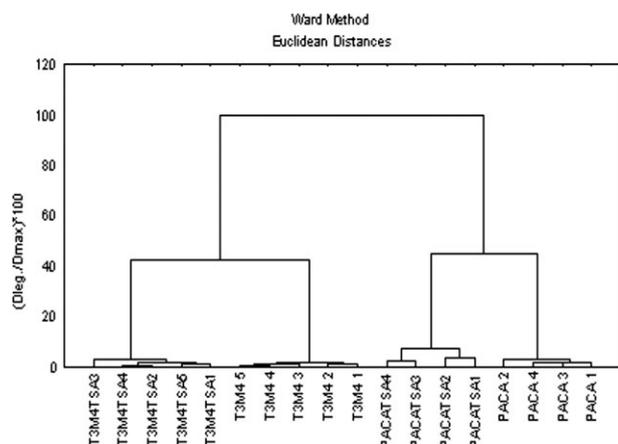


Figure 9. Dendrogram calculated on the basis of the first three PCs (Ward method, Euclidean distances) for a human pancreatic cancer cell dataset. Reprinted from [52] with permission.

[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.

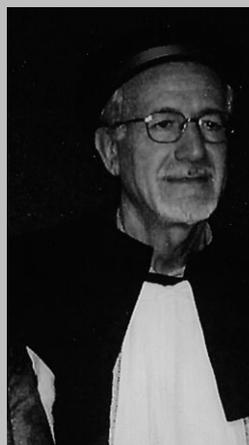
The approaches described in Section 5 show another important advantage, related to the possibility of identifying the differences between groups of samples smaller than the classical 2-fold: this is achieved by means of PCA, which also allows the identification of small differences due to its peculiar 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., Lipkinn, 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-Jakob 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

Proteomic changes in rat serum, polymorphonuclear and mononuclear leukocytes after chronic nicotine administration

Chiara Piubelli¹, Daniela Cecconi¹, Hubert Astner², Fabrizio Caldara³, Michela Tessari⁴, Lucia Carboni⁵, Mahmoud Hamdan², Pier Giorgio Righetti¹ and Enrico Domenici⁵

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

² Department of Computational, Analytical and Structural Science, GlaxoSmithKline S.p.A., Verona, Italy

³ Bioinformatics, GlaxoSmithKline S.p.A., Verona, Italy

⁴ Drug Dependency & Behavioural Neurochemistry, Psychiatry Centre of Excellence for Drug Discovery, GlaxoSmithKline S.p.A., Verona, Italy

⁵ Molecular Medicine, Psychiatry Centre of Excellence for Drug Discovery, GlaxoSmithKline S.p.A., Verona, Italy

In order to gain information about the effect triggered at the molecular level by nicotine, its neuroimmunomodulatory properties and its impact on the pathogenesis of inflammatory diseases, peripheral blood serum and leukocytes of rat submitted to passive nicotine administration were subjected to proteomic investigation. Serum, polymorphonuclear (PMN) and mononuclear (MN) leukocytes from chronically treated animals and from control animals were analysed by a two-dimensional (2-D) gel electrophoresis/mass spectrometry approach to detect differentially expressed proteins. The nicotine regimen selected is known to have a stimulatory effect on locomotor activity and to produce a sensitisation of the mesolimbic dopamine system mechanism involved in addiction development. After 2-D gel analysis and matching, 36 spots in serum, seven in PMN and five in MN were found to display a statistical difference in their expression and were subjected to matrix-assisted laser desorption/ionization-time of flight-mass spectrometry peptide fingerprinting for protein identification. Fifteen different proteins were identified. The results indicate an overall impact of nicotine on proteins involved in a variety of cellular and metabolic pathways, including acute phase response (suggesting the effect on inflammatory cascades and more in general on the immune system), oxidative stress metabolism and assembly and regulation of cytoskeleton. In particular, the observed changes imply a general reduction in the inflammatory response with a concomitant increased unbalance of the oxidative stress metabolism in the periphery and point to a number of potential noninvasive markers for the central nervous system (CNS) and non-CNS mediated activities of nicotine.

Received: April 29, 2004
Revised: July 26, 2004
Accepted: September 3, 2004

Keywords:

Leukocytes / Matrix-assisted laser desorption/ionization-time of flight-mass spectrometry / Nicotine / Serum / Two-dimensional maps

Correspondence: Dr. Enrico Domenici, Department of Molecular Medicine, Psychiatry Centre of Excellence for Drug Discovery, GlaxoSmithKline S.p.A., via Fleming, 4, 37135 Verona, Italy
E-mail: enrico.h.domenici@gsk.com
Fax: +39-0459218047

Abbreviations: MN, mononuclear leukocytes; PMN, polymorphonuclear leukocytes; SOD, superoxide dismutase

1 Introduction

Nicotine is widely accepted as the addictive constituent of tobacco endowed with reinforcing properties that is responsible for tobacco addiction [1, 2]. It acts *via* activation of nicotinic acetylcholine receptors (nAChRs) that are present in neuronal and in non-neuronal cells [3–6]. However, the

pharmacodynamics of nicotine are complex and results can be dependent on the duration of exposure to the drug; moreover tolerance to many effects of nicotine develops rapidly [7, 8]. Due to its complexity, the mechanism of action of nicotine is still not fully understood, neither in the central nervous system (CNS) nor in the peripheral system. Particularly in the periphery, several factors are involved in the pathological effects of smoking in humans, but there is less information about the pharmacological effects of nicotine itself. From the literature, we know that nicotine produces some sympathetic effects in the periphery, such as tachycardia, vasoconstriction, mydriasis, bronchiolar dilatation, decrease in gastrointestinal motility and glycogenolysis.

These effects are mediated by the nicotine stimulation of postganglionic sympathetic neurons and adrenal medulla [9], where nAChRs are expressed in animals [10] as well as humans [11]. Recently, some studies in rodents and rabbit models and cell preparations showed that nicotine, when administered acutely, induced angiogenic and arteriogenic effects, which could be mediated through non-neural nAChRs [12]. It has been also shown that nicotine could stimulate peripheral blood monocytes by increasing their adhesion, chemotaxis and transendothelial migration [13, 14], factors that are important for angiogenesis. These findings would suggest a role of nicotine in pathological (*i.e.*, inflammation, tumor and atherosclerosis) as well as therapeutic angiogenesis (*i.e.*, protection against vascular ischemia). There is increasing evidence that nicotine can affect both humoral and cell-mediated branches of the immune system. Altered changes would be characterised by a decrease in inflammation, a decrease in antibody response and reduction in T cell receptor-mediated signalling [14, 15]. A few studies showed that chronic exposure to nicotine impaired immune response in rats [16, 17]. It has been reported that continuous subcutaneous treatment with nicotine for 3–4 weeks reduced in fact the antibody-forming cell response, producing an immunosuppression [17]. Furthermore, recent evidence suggests that non-neuronal nAChRs are also involved in the regulation of vital cell functions, such as mitosis, differentiation, organisation of the cytoskeleton, cell-cell contact, locomotion, and migration [3].

The aim of this study was to establish if chronic nicotine treatment in rats could induce proteomic changes in peripheral blood serum and leukocytes, providing information about pathways involved in the nicotine pharmacological action, for example investigating the involvement of nicotine in the pathogenesis of inflammatory and immune system diseases.

To detect proteins differentially expressed after nicotine treatment, a differential proteomic technique was applied, based on a 2-D gel electrophoresis/mass spectrometry approach. Serum, polymorphonuclear (PMN) and mononuclear (MN) leukocytes were analysed from rats treated with nicotine in comparison with control rats, treated with saline solution. The chronic nicotine treatment was performed according to regimens similar to those known to

have a stimulatory effect on locomotor activity and to produce a sensitisation of the mesolimbic dopamine system, a mechanism involved in addiction development [18]. The same nicotine treatment is also known to induce axonal distribution changes in rat ventral tegmental area neurofilaments [19, 20]. In addition, from previous observations it can be extrapolated that this amount of nicotine (0.4 mg/kg/day) is comparable to that received by a 70-kg human, smoking approximately one pack of cigarettes/day [21].

After 2-D gel analysis and matching, a statistical test was performed for detecting up- and down-regulated proteins by nicotine treatment, thus eliminating sample and gel variability. Differentially expressed proteins in serum, PMN and MN samples were identified by MALDI-TOF MS peptide fingerprinting. Database query was performed by using three different software packages, comparing different algorithms to confirm protein identifications.

2 Materials and methods

2.1 Animals and nicotine treatment

Male Wistar rats (200–300 g) were purchased from Charles River (Calco, Italy). Rats were housed in three or four *per* cage and maintained with food and water *ad libitum* in a temperature-controlled environment on a 12 h/12 h light-dark cycle with light on at 6:00 a.m. Rats were treated with nicotine (0.4 mg/kg/day in saline, in a single injection) or with saline solution ($n = 5$ /group) *via* subcutaneous injection for 14 days. The animals were sacrificed under deep CO₂ anaesthesia, blood was collected intracardially and immediately heparinised for leukocyte isolation. The research complied with national legislation and with the company policy on the Care of Use of Animals and related codes of practice.

2.2 Serum sample preparation

On day 14 of treatment, 500 μ L blood samples were taken from the tail vein of rats kept fasting since the night before in order to avoid lipid interference. Samples were collected in Microtainer tubes (BD Biosciences, Erembodegen, Belgium), incubated for 30 min on an orbital shaker and centrifuged for 15 min at 1000 \times *g* at room temperature. From each sample about 200 μ L of serum (about 60 mg/mL of total protein) were recovered and stored at -20°C until analysis.

2.3 Leukocyte sample preparation

Blood from the nicotine- or saline-treated animals was pooled (5 rats/group). The PMN and MN leukocyte populations were isolated by using the polymorphprep method (Nycomed Pharma, Oslo, Norway). Rat blood was processed according to the manufacturer's protocol, as previously described [22], PMN and MN bands were retrieved and washed in PBS solution (PBS Dulbecco's; Gibco, Invitrogen,

Paisley, UK). An additional lysing step was added for PMN purification to eliminate erythrocyte contamination by Vita-Lyse™ erythrocyte lysing kit WBL1000 (BioErgonomics, St. Paul, MN, USA) [22]. In order to monitor the purity of the PMN and MN populations, an aliquot of each sample was analysed by flow cytometry (light scattering) on a Partec PAS cytofluorometer (Partec, Münster, Germany). About 9.5×10^5 PMN cells and 6×10^6 MN cells *per* mL of blood were recovered. The PMN and MN pellets were finally dissolved at a concentration of 2×10^7 cells/mL in a 2-D solubilising solution: 7 M urea (Fluka, Buchs, Switzerland), 2 M thiourea (Fluka), 1% amidosulphobetaine 14 (Calbiochem, San Diego, CA, USA), 40 mM Tris (Sigma-Aldrich, St. Louis, MO, USA), 3 mM tributyl phosphine (Sigma), 1% Pharylates 3.5–10 (Amersham Biosciences, Uppsala, Sweden), sonicated and stored at -80°C . Protein concentration was evaluated with bicinchoninic assay (Pierce, Perbio, Helsingborg, Sweden) after dialysis in 1% SDS and 1 mM EDTA in water, overnight at room temperature.

2.4 Serum sample 2-DE

One hundred μL of rat serum (containing about 60 mg/mL total protein) were added with 0.4 mL of a 2-D solubilising solution, containing 7 M urea, 2 M thiourea, 40 mM Tris, 5 mM tributyl phosphine. Samples were incubated 90 min at room temperature to reduce protein disulphide bonds. Cysteine thiolic groups were then alkylated by 20 mM iodoacetamide (Sigma) for 90 min at room temperature. Samples were dialysed in a regenerated cellulose membrane (cut-off of 3.5 kDa) overnight in distilled water. Then chaotropes and surfactant were added to a final concentration of 7 M urea, 2 M thiourea, 20 mM Tris and 2% CHAPS (Sigma). Nonlinear immobilised pH gradient strips (18 cm long, pH 3–10; Amersham Biosciences) were rehydrated for 8 h with 450 μL of the sample solution (final total protein concentration of 6 mg/mL) containing traces of bromophenol blue to monitor the electrophoretic run. IEF was conducted at 20°C for 60 000 Vh using a Protean IEF Cell (Bio-Rad, Hercules, CA, USA) 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. 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). 7–20%T gradient SDS-PAGE were prepared in-house, using the Protean Multi-Gel Casting Chamber (Bio-Rad) and the IPG strips were then laid on top of the gels with 0.5% agarose in the cathode buffer (192 mM glycine (Sigma), 15 mM Tris, 0.1% SDS, pH 8.3). The anode buffer consisted of 375 mM Tris-HCl pH 8.8. Gels were run at 10°C and 2 mA/gel for 2 h, 5 mA/gel for 1 h and 10 mA/gel overnight in a Protean XL Multi-cells apparatus (Bio-Rad), hosting six gels simultaneously. Gels were stained 20 h with colloidal Coomassie blue G-250 (0.1% Blue G-250 (BDH, Poole, England), 34% v/v methanol, 3% v/v o-phosphoric acid, 17% w/v ammonium sulphate (Sigma)) and destained in 5%

acetic acid (Sigma). Images were captured with a GS-710 imaging densitometer (Bio-Rad) selecting a high resolution ($42.3 \mu\text{m}^2$ for pixel) scanning.

2.5 PMN and MN samples 2-DE

Seventeen cm long, immobilised pH 3–10 linear gradient strips (Bio-Rad) were rehydrated for 8 h with 450 μL of 2-D solubilising solution containing about 10×10^6 cells of PMN or 15×10^6 MN leukocytes (about 1.6 mg of proteins) and 10 mM of iodoacetamide as alkylating agent. IEF was conducted at 20°C for 65 000 Vh at a maximum of 3500 V using Multiphor II system (Amersham Biosciences). The Multiphor system was used for PMN and MN samples because it could host large pre-wet electrode wicks, which were necessary to absorb salt residues remaining from the leukocytes isolation procedure. For the second dimension, the IPG strips were equilibrated for 25 min by rocking in a solution of 6 M urea, 2% SDS, 20% glycerol, 375 mM Tris-HCl pH 8.8. Second dimension runs and image capture were performed as described in Section 2.4.

2.6 Protein pattern differential analysis

2-D gel analysis was performed by PDQuest software v6.2 (Bio-Rad). Each gel was analysed for spot detection, background subtraction and protein spot OD intensity quantification (spot quantity definition). The gel image showing the higher number of spots and the best protein pattern was chosen as a reference template, and spots in a standard gel were then matched across all gels. For serum samples an average number of 350 spots *per* gel was detected and 325 spots of the resulting reference map were matching across the whole gel set. For PMN samples an average number of 500 spots were detected *per* gel and 459 spots of the resulting reference map were matching to the gel set. For MN samples the average number of the detected spots *per* gel was 850 and 805 spots of the resulting reference map matched across the matchset. Spot quantity values were normalised in each gel dividing the raw quantity of each spot by the total quantity of all the spots included in the standard gel. Gels were divided in two separated groups (control and nicotine-treated samples) and, for each protein spot, the average spot quantity value and its coefficient of variation (CV, standard deviation/average $\times 100$) in each group were determined. For serum samples an average CV of 58% was determined (calculated for all the spots matched in the matchset, excluding missing values), with the 46% of the spots showing a CV $< 50\%$. For PMN and MN matchsets average CVs of 59% and 57% respectively were calculated, with 47% and 46% of the spots showing a CV $< 50\%$. A Student's *t*-test was performed in order to compare the two groups and identify sets of proteins that showed a statistically significant difference with a confidence level of 0.05. For the differentially expressed spots, when matched spots were missing in some gels, the statistical test analysis was per-

formed by GraphPad Prism Version 3.00 for Windows (GraphPad Software, San Diego, CA, USA), for checking the results.

2.7 MALDI-TOF MS

Each selected spot was carefully cut and destained with 2×10 min washing steps in 50% v/v acetonitrile (Sigma), 50% 5 mM Tris pH 8.5 followed by a third wash with 5 mM Tris pH 8.5 for 10 min. The gel discs were dried in a Speedvac sc110A device (Thermo Savant, NY, USA) for 1 h at room temperature and then covered with 15 μ L of sequencing grade modified trypsin (0.02 mg/mL; Promega, Madison, WI, USA) in NH_4HCO_3 buffer (40 mM, pH 8.5) and left at 37°C overnight. The gel pieces were then crushed and peptides were extracted twice in 50 μ L of 50% acetonitrile, 50% H_2O with 1% formic acid v/v and a third time with 50 μ L of acetonitrile. The extractions were conducted in an ultrasonic bath for 15 min. The three extraction solutions were mixed and evaporated to dryness in the Speedvac device and the residues were dissolved in 10 μ L of H_2O with 0.1% of TFA. For an additional purification, the samples were cleaned by using ZipTip C18 (Millipore, Bedford, MA, USA). Two μ L of the resulting solution were mixed with an equivalent volume of matrix solution, prepared fresh every day by dissolving 10 mg/mL CHCA in acetonitrile:ethanol (1:1, v/v). One μ L of the resulting mixture was loaded onto the MALDI sample plate and allowed to dry. Measurements were performed using a ToFSpec 2E MALDI-TOF instrument (Micromass, Manchester, UK), operated in reflector mode, with an accelerating voltage of 20 kV.

Two different analysis tools were used for protein identification by peptide masses in order to have always a double confirmation of the results. MASCOT software (Matrix Science, UK) [23], that incorporates a probability-based scoring, was used to search Swiss-Prot, TrEMBL and NCBI non-redundant databases with Mammalia (mammals) as taxonomic category. ProFound (V4.10.5) searches were performed using the same databases with the same taxonomic category. Database queries were carried out for monoisotopic peptide masses, using a peptide mass tolerance of ± 150 ppm, 1 as maximum number of missed tryptic cleavages, molecular mass range depending on the apparent M_r of the protein spot in the selected 2-D gel and all pI ranges. Allowed modifications included oxidation of Met and alkylation of Cys residues by iodoacetamide. The criteria for positive identification of proteins were set as follow: protein scores are considered significant when the probability of random events is less than 1 in 20 ($p < 0.05$). For MASCOT analyses, this means that proteins with a probability-based MOWSE score (the reported MASCOT score in the identification Tables 1 and 2) greater than the threshold calculated by the tool algorithm were considered significant identifications. For ProFound it roughly corresponds to a Z-score greater than 1.65.

A third tool, ProteinProbe (V3.4, BioLynx; Micromass), was used as an additional confirmation of the identified proteins and the resulting MOWSE score was considered significant if there was a difference of two orders of magnitude in the score value between the first and the second hit.

The result of a database search was significant if the protein was ranked as best hit and there was a sequence coverage of at least 20%. Positive identification of the protein was assigned only if the mass deviation of matched peptides was constant over the whole mass range. Matched peptide masses were evenly distributed throughout the complete protein amino acid sequence, and the identified protein molecular mass and pI values corresponded approximately to experimental values, with some exceptions.

3 Results

3.1 Differential proteomic analysis in serum samples

In order to obtain good focusing of protein in the first dimension, the serum samples were dialysed for desalting. Ten maps, one for each animal in the experiment (five controls, S1–S5, and five nicotine-treated rats, N1–N5), were performed, as previously described in Section 2.4. The separated protein spots were visualised on 2-D gels by colloidal Coomassie blue staining, which allows good reproducibility and protein spot quantification for comparison analysis.

The ten serum 2-D gels corresponding to the ten animals involved in the experiment were analysed and matched in the same matchset. Control and nicotine-treated samples were then divided in two groups (five gels *per* group) and statistically compared as reported in Section 3.3. A Student's *t*-test was performed in order to identify sets of proteins that are differentially expressed with a confidence level of 0.05. The results were checked again by extracting the normalised quantity value data and performing the statistical analysis with Prism software.

Thirty-six spots resulted as significantly different with a $p < 0.05$ and, among these 41% of the spots were significantly different also with a more conservative cut-off ($p < 0.01$). Seven spots were up-regulated and 29 down-regulated in nicotine-treated samples with respect to controls (Fig. 1).

3.2 Differential proteomic analysis in PMN and MN leukocytes

Polymorphonuclear and mononuclear leukocyte populations were isolated from the pooled blood of the five controls or the five nicotine-treated rats as reported in Section 2.3. As previously assessed in the development of PMN 2-D maps [22], the isolation of cells from single rat blood results in low yields, not sufficient for obtaining good maps (in terms of spot number). Pooling blood samples from five rats allowed a substantial improvement of PMN and MN recovery, giving in addition the possibility of running three gel replicates *per*

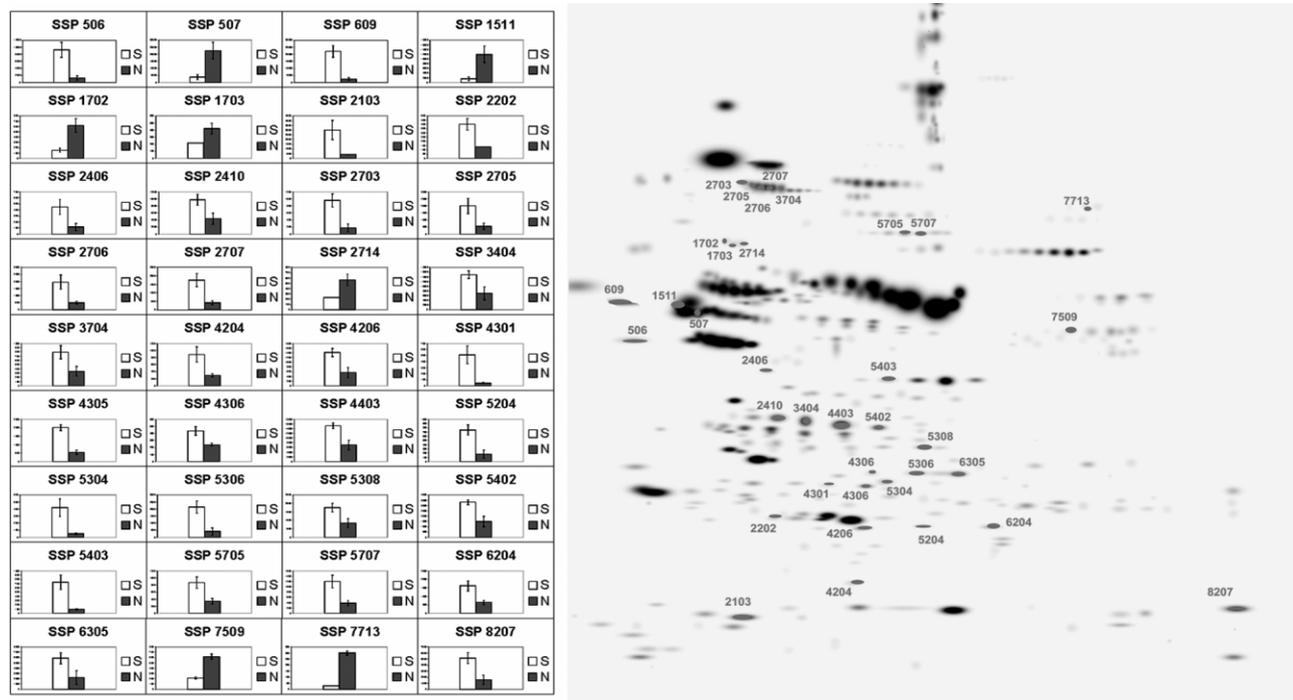


Figure 1. Reference image for serum matchset. Significant spots resulting from statistical analysis are indicated and numbered. Histograms represent the corresponding spot intensity (average value between replicates) in saline (left bar) and in nicotine (right bar) groups.

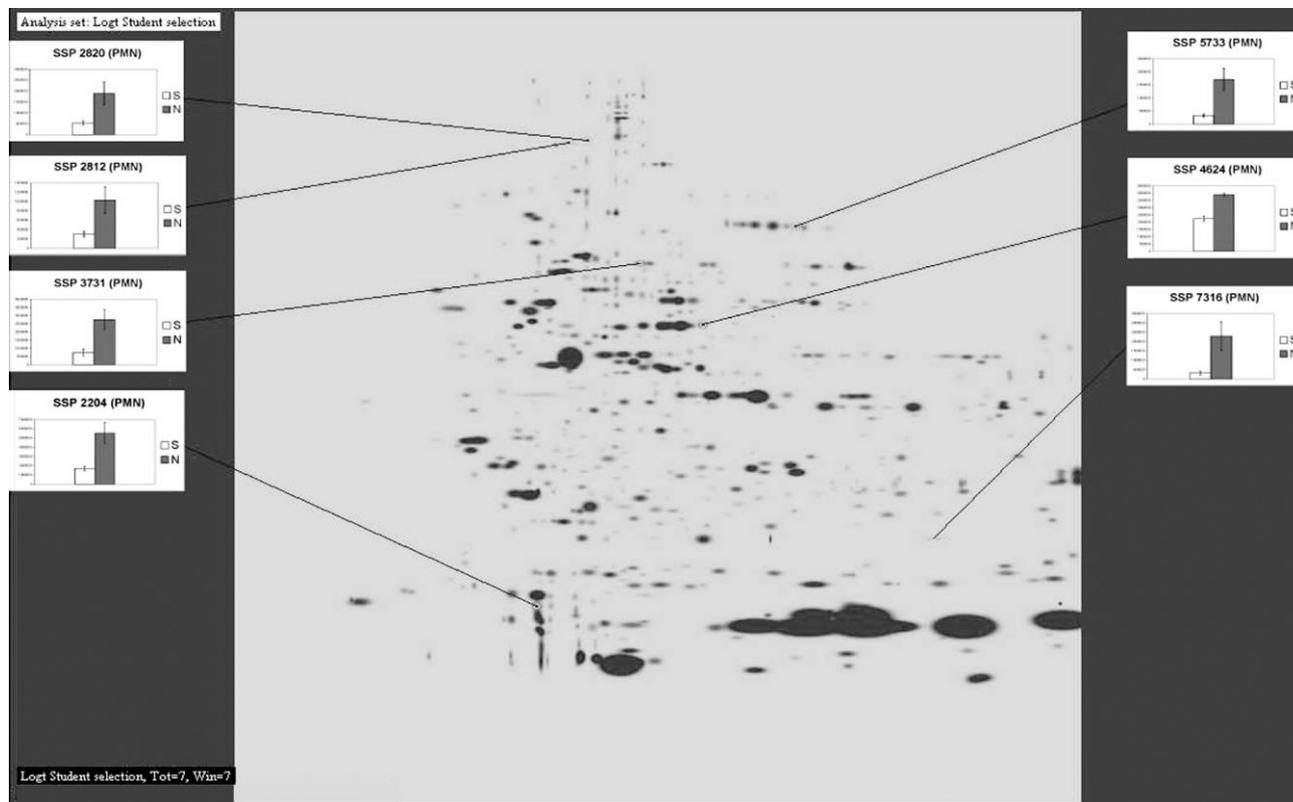


Figure 2. Reference image for PMN matchset. Significant spots resulting from statistical analysis are indicated. Histograms represent the corresponding spot intensity (average value between replicates) in saline (left bar) and in nicotine (right bar) groups.

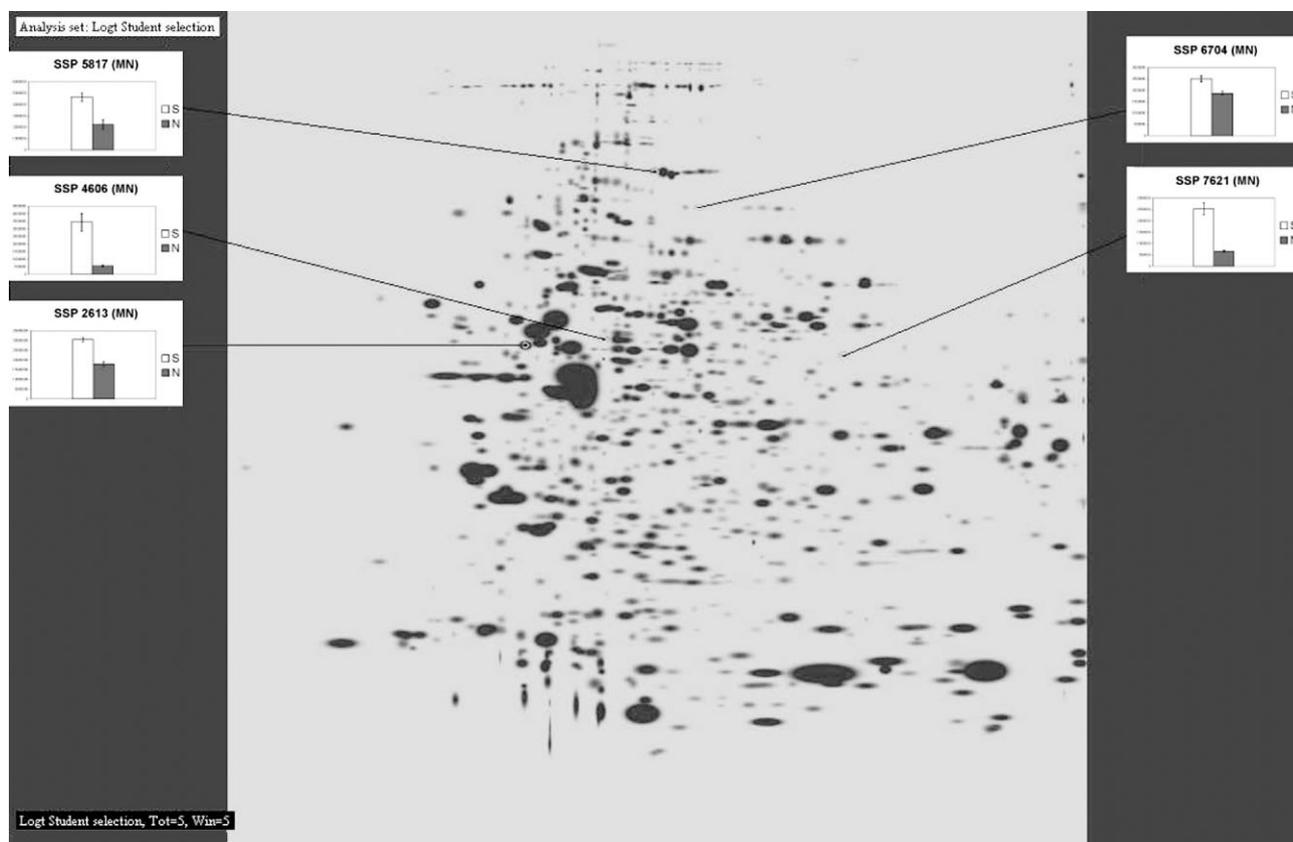


Figure 3. Reference image for the MN matchset. Significant spots resulting from statistical analysis are indicated.

experimental group. The purity of PMN and MN populations was checked according to morphological analysis (light scattering) by flow cytometry: in saline-treated samples, isolated PMN and MN populations were both 82% pure; in the nicotine-treated samples, isolated PMN and MN preparations were respectively 85% and 80% pure. Flow cytometry analysis thus indicated that the cell preparations to be compared (saline- vs. nicotine-treated) had similar purity levels in their leukocyte populations. Three 2-D gel replicates were performed for both nicotine-treated and control group PMN, and the same procedure was adopted for MN samples (for details, refer to Sections 2.3 and 2.5).

The Coomassie stained 2-D gels were divided and analysed in two separate experimental groups: the first including gels obtained from PMN of both saline- and nicotine-treated samples and the second including gels performed with MN recovered from the same samples. Comparative and statistical analyses of each experimental group were performed as previously reported for serum samples (Section 3.1). In the PMN analysis seven protein spots were up-regulated and in the MN analysis five protein spots resulted as down-regulated in the nicotine-treated samples with respect to controls (Figs. 2 and 3 with $p < 0.05$. Among these, five showed a $p < 0.02$ and one a $p < 0.005$. Due to the very similar leukocyte population percentage between the com-

pared samples, we assume that the two opposite trends of up- or down-regulation are not due to a modification of the PMN/MN ratio in the analysed samples and we are confident that they reflect the differential protein expression caused by the effect of nicotine in the examined population.

3.3 Identification of differentially expressed proteins from serum, PMN and MN samples

The serum map was compared with the existing reference map of rat serum (<http://linux.farma.unimi.it/RSPSG/index.html> and [24]) for the possible identification of the 36 modulated proteins. From this comparison, spots 507 and 1511 could be predicted as isoforms of the contrapsin-like protease inhibitor 1 and spots 3404, 4403, 2410, 5308, 5402 as different isoforms of alpha-1-macroglobulin. Since our 2-D gel protocols imply an alkylation step before the first dimension [25], that may result in different migration patterns, to confirm these predictions and achieve new identifications all spots selected from the differential analysis of serum, PMN and MN leukocytes were subjected to MS for protein identification.

In Tables 1 and 2 the identity of the successfully identified proteins corresponding to up- or down-regulated spots in the 2-D gel maps for serum, PMN and MN are shown,

Table 1. Summary of identified proteins from serum 2-D gels.

Spot	Exp. pI	Exp. M_r (Da)	ID	Name	Theor. pI, M_r	MASCOT	MOWSE- score	Z-score	No. of matched peptides	% sequence coverage	Function of the protein	Observed expression after nicotine administration
2 (507)	~4	~50 000	CPI1_RAT	Contrapsin-like protease inhibitor 1	5,4 46 562	154	4,34 E15	2.34	23	52	Inhibits kallikreins and trypsin but not chymotrypsin or elastase	Upregulated
4 (1511)	~4	~50 000	CPI1_RAT	Contrapsin-like protease inhibitor 1	5,4 46 562	169	3,10 E10	2.35	17	36	Inhibits kallikreins and trypsin but not chymotrypsin or elastase	Upregulate
5 (1702)	~5	~70 000	AFAM_RAT	Afamin	5,9 69 336	245	3,48 E21	2.37	37	52	Possible role in the transport of yet unknown ligand	Upregulated
6 (1703)	~5	~70 000	AFAM_RAT	Afamin	6,1 69 336	126	1,72 E14	2.38	25	31.6	Possible role in the transport of yet unknown ligands	Upregulated
11 (2703)	~5.5	~100 000	O35802	Inter-alpha inhibitor H4 heavy chain	6,4 103 607	254	4,05 E17	2.42	29	30	Plasma proteinase inhibitor	Down-regulated
12 (2705)	~5.5	~100 000	O35802	Inter-alpha inhibitor H4 heavy chain	6,4 103 607	224	7,44 E20	2.41	34	35	Plasma proteinase inhibitor	Down-regulated
14 (2707)	~5.5	~100 000	O35802	Inter-alpha inhibitor H4 heavy chain	6,4 103 607	285	7,46 E21	2.37	35	37	Plasma proteinase inhibitor	Down-regulated
16 (3404)	~6	~40 000	Q63332	Alpha-1-macroglobulin	6,8 167 160		1,14 E9	2.37	15	11	Inactivation of proteinases released in response to tissue injury, necrosis or inflammation	Down-regulated
23 (4403)	~6.5	~40 000	Q63332	Alpha-1-macroglobulin	6,8 167 160		2,29 E9	2.4	15	10	Inactivation of proteinases released in response to tissue injury, necrosis or inflammation	Down-regulated
10 (2410)	~6	~40 000	Q63332	Alpha-1-macroglobulin	6,8 167 160		3,91 E9	2.36	14	11	Inactivation of proteinases released in response to tissue injury, necrosis or inflammation	Down-regulated
27 (5308)	~7.5	~38 000	Q63332	Alpha-1-macroglobulin	6,8 167 160		8,46 E11	2.29	17	14	Inactivation of proteinases released in response to tissue injury, necrosis or inflammation	Down-regulated
28 (5402)	~7	~40 000	Q63332	Alpha-1-macroglobulin	6,8 167 160		1,25 E9	2.41	16	12	Inactivation of proteinases released in response to tissue injury, necrosis or inflammation	Down-regulated
30 (5705)	~7	~80 000	GELS_MOUSE	Geisolin	5,7 80 747	248	4,85 E10	2.34	27	23	Controls cytoskeletal assembly and disassembly	Down-regulated

Table 1. Continued.

Spot	Exp. p/ (Da)	Exp. M_r (Da)	ID	Name	Theor. pI, M_r	MASCOT score	MOWSE- score	Z-score	No. of matched peptides	% se- quence coverage	Function of the protein	Observed expression after nico- tine ad- ministration
32 (6204)	~8	~25 000	GSHP_RAT	Plasma glutathione peroxidase	8,4 25 393	167	3,29 E8	2.35	14	45	Protects cells and enzymes from oxidative damage, by catalysing the reduction of hydrogen peroxide, lipid peroxides and organic hydroperoxide, by glutathione	Down-regulated
24 (5204)	~7.5	~25 000	GSHP_RAT	Plasma glutathione peroxidase	8,4 25 393	143	8,23 E7	2.41	13	45	Protects cells and enzymes from oxidative damage, by catalysing the reduction of hydrogen peroxide, lipid peroxides and organic hydroperoxide, by glutathione	Down-regulated

For spot numbers, refer to Fig. 1. Matching peptides vs. total number of peptides submitted to database search, sequence coverage and score resulted for peptide fingerprinting matching are listed. Experimental pI and M_r values, resulting from gel separations, are indicated in the "Exp. pI and M_r " columns. Protein ID, theoretical pI and M_r and protein functions were obtained by the Swiss-Prot and NCBI databases. MASCOT, MOWSE and Z-scores (output of the identification softwares MASCOT, ProteinProbe and ProFound respectively) are measures of the statistical significance of the identification hits (see Section 2.7).

Table 2. Summary of identified proteins from PMN and MN 2-D gels

Spot	Exp. p/ (Da)	Exp. M_r (Da)	ID	Name	Theor. pI, M_r	Z-score	MASCOT score	MOWSE score	No. of matched peptides	% se- quence coverage	Function of the protein	Observed expression after nico- tine ad- ministration
1 PMN (2820)	~5.6	~200 000	MYH9_RAT	Myosin heavy chain, non-muscle type A	5.49, 226 338.10	1.97	164	286 E18	37	16	Cellular myosin appears to play a role in cytokinesis, cell shape, and specialized functions such as secretion and capping	Upregulated
3 PMN (5733)	~7.5	~85 000	TRFE_RAT	Serotransferrin	6.8, 74 469	2.42	240	2.14 E18	29	40	Transferrin is an iron-binding transport protein. It is responsible for the transport of iron from sites of absorption and heme degradation to those of storage and utilisation. Serum transferrin may also have a further role in stimulating cell proliferation	Upregulated

Table 2. Continued.

Spot	Exp. p/	Exp. M_r	ID	Name	Theor. p/, M_r	Z-score	MASCOT score	MOWSE score	No. of matched peptides	% sequence coverage	Function of the protein	Observed expression after nicotine administration
4 PMN (3731)	~6	~68 000	ALBU_RAT	Serum albumin	6.4, 68 719	2.35	138	3.70 E9	16	30	Serum albumin, the main protein of plasma, has a good binding capacity for water, Ca(2+), Na(+), K(+), fatty acids, hormones, bilirubin and drugs. Its main function is the regulation of the colloidal osmotic pressure of blood	Upregulated
5 PMN (4624)	~6.9	~45 001	ENOA_RAT	Alpha enolase	6.16, 46 985	2.23	63	2.29 E14	19	47	Glycolytic enzyme. Catalytic activity: 2-phospho-D-glycerate = phosphoenolpyruvate + H ₂ O	Upregulated
6 PMN (7316)	~9	~20 000	SODM_RAT	Superoxide dismutase	9.1, 24 674	2.28	84	7.67 E6	9	47.7	Destroys radicals which are normally produced within the cells and which are toxic to biological systems. Catalytic activity: 2 superoxide + 2 H+ = O ₂ + H ₂ O ₂	Upregulated
7 PMN (2204)	~5	~15 000		Unknown							Good MS spectrum, but no identification by peptide fingerprinting	Upregulated
1 MN (5817)	~6.2	~120 000	VINC_MOUSE	Vinculin	5.72, 116 674	2.4	102	7.55 E12	18	19	Involved in cell adhesion and in the integrin signalling pathway. May be involved in the attachment of the actin-based microfilaments to the plasma membrane	Down-regulated
2 MN (6704)	~6.8	~100 000	AAC1_RAT	Alpha-actinin 2	5.23, 102 960	2.35	88	1.31 E11	19	20	F-actin cross-linking protein which is thought to anchor actin to a variety of intracellular structures. This is a bundling protein, involved in the integrin signalling pathway	Down-regulated

Table 2. Continued.

Spot	Exp. pI	Exp. M_r	ID	Name	Theor. pI, M_r	Z-score	MASCOT score	MOWSE score	No. of matched peptides	% sequence coverage	Function of the protein	Observed expression after nicotine administration
3 MN (2613)	~5.1	~60 000	VIME_RAT	Vimentin	5.06, 53 601.54	2.41	258	8.24 E19	38	69	Vimentins are class III intermediate filaments found in various nonepithelial cells. Activated macrophages secrete vimentin into the extracellular space, in response to pro-inflammatory signalling pathways; this protein is probably involved in immune function (bacterial killing and generation of oxidative metabolites)	Down-regulated
4 MN (4606)	~6	~60 000	FIBG_RAT	Fibrinogen gamma chain	5.39, 50 632.65	2.26	109	2.10 E9	15	31	Fibrinogen has a double function: yielding monomers that polymerise into fibrin and acting as a cofactor in platelet aggregation. The plasma fibrinogen interacts with leukocyte integrins; this interaction is mediated by the C-term of the gamma chain. Fibrinogen promotes leukocyte adhesion to endothelium and leukocyte transendothelial migration	Down-regulated

For spot numbers, refer to Figs. 2 and 3. Matching peptides vs. total number of peptides submitted to database search, sequence coverage and score resulted for peptide fingerprinting matching are listed. Experimental pI and M_r values, resulting from gel separations, are indicated in the "Exp. pI and M_r " columns. Protein ID, theoretical pI, M_r and protein functions were obtained by the Swiss-Prot and NCBI databases. MASCOT, MOWSE and Z-scores (output of the identification softwares MASCOT, ProteinProbe and ProFound respectively) are measures of the statistical significance of the identification hits (see Section 2.7).

together with the identification parameters and the indication of the main role(s) of the protein in cellular pathway and metabolism. Not all the differentially expressed spots were identified, because some of them did not give a good MS spectrum after the tryptic digest of the protein: 15 spots were identified in serum samples, five in PMN and four in MN. One PMN spot gave a good MS spectrum, but it was not

possible to identify the protein by data-base search. A few different spots in serum resulted in the same identified protein, thus indicating that the modulation of expression occurs in different isoforms of the same protein, often with the same M_r but different pI. All the reported identifications had significant score values resulting from the three selected identification software packages. Alpha-1-macroglobulin was

an exception as only a small fragment of the protein was identified by two of the three algorithms (ProFound and ProteinProbe software). MASCOT software did not recognise the fragment as a part of the protein, probably due to the low coverage percentage of the identified sequence with respect to the total protein. The spots identified by differential proteomics in PMN and MN are proteins known to be expressed by leukocytes, with the exception of serum transferrin and albumin. However, these proteins (and their change upon treatment) should not necessarily be ascribed to contaminating serum in the leukocyte preparation, as leukocytes are known to be able to bind and internalise serum proteins under some conditions [26]. For PMNs, two of the identified proteins (myosin and alpha enolase) have been shown previously to belong to the proteomic complement detectable by 2-DE [22].

Overall, the main impact of chronic treatment with nicotine can be observed in acute-phase proteins (indicating an overall effect on inflammatory cascades and more in general on the immune system), in proteins belonging to the oxidative stress metabolism and in proteins involved in the assembly and regulation of cytoskeleton structures. The indication arising from the analysis of the individual changes observed and their potential implications are discussed below.

4 Discussion

4.1 Effects of chronic nicotine on inflammation

The proteomic analysis of serum has highlighted a main impact of chronic nicotine treatment on the inflammatory response, in agreement with the known anti-inflammatory properties of nicotine [27, 28]. Accordingly, we have shown upon chronic nicotine administration an increase in the serum content of contrapsin-like protease inhibitor 1 (or kallikrein binding protein), a negative acute-phase protein that has a protective effect during acute phase inflammation [29]. The expression of additional proteins involved in the acute phase inflammatory response was also modified by the treatment. As an example, we identified inter-alpha trypsin inhibitor (ITIH4), a protein found to behave as a so-called positive acute phase protein in animals [30], and a type II acute phase protein in humans [31]. ITIH4 significantly decreased in serum, whilst its concentration was found to be increased in patients suffering from acute phase processes [31]. Alpha-1 macroglobulin, that is required to bind and inactivate proteinases released in response to tissue injury and inflammation, is decreased, indicating that the acute phase synthetic response of the liver is down-regulated by chronic exposure to nicotine.

The impact of nicotine administration on the immune system can also be observed by examining the changes in the proteomic complement of mononuclear cells. As an example, vimentin, a class II filament located in nonepithelial

cells, was found to be decreased upon nicotine exposure. Interestingly, this protein is usually secreted by activated macrophages in response to pro-inflammatory stimuli, and its secretion is enhanced by tumour necrosis factor and decreased by anti-inflammatory cytokines such as interleukin-10 [32]. Moreover, we have observed a decrease in fibrinogen gamma chain, possibly bound to the integrin receptor on the leukocyte surface. The interaction of fibrinogen with leukocyte integrins constitutes an important component of the inflammatory response [33]. In addition to its role in haemostasis, this protein is usually required to mount competent inflammatory response *in vivo* and it is responsible for the mediation of leukocyte adhesion and migration.

4.2 Effects on cytoskeleton reorganisation

Other proteins found to be decreased in mononuclear cells are vinculin and alpha actinin, both of them belonging to the class of cytoskeleton proteins. The above observed changes seem to suggest a cytoskeleton reorganisation in lymphocyte cells that could possibly be related to the altered immune function observed.

The analysis of the protein changes observed in the 2-D gel proteomic analysis of PMN cells also indicates changes occurred at the cytoskeleton level, that could probably be related to altered inflammatory response and adhesion as well. The up-regulation of myosin heavy chain in PMN for example indicates changes in the cytoskeleton structure that could alter crucial functions such as cytokinesis and chemotaxis. Indeed [34] it has been shown that nicotine is the cigarette smoke component responsible for increased chemotaxis in PMN, that is partially responsible for the increase of PMN in the lung of smokers. Also, it has been shown that nicotine can induce significant leukocyte rolling and adhesion in the cerebral microcirculation of the mouse [35].

4.3 Effects on oxidative stress balance

Additionally, in PMN we can observe changes in proteins belonging to the pathway involved in the response to oxidative stress that, together with additional changes observed in serum, could also modulate leukocyte activation. In answering the question of whether nicotine is the major causal agent of oxidative damage triggered by cigarette smoke, a number of investigations have shown a strong relationship between nicotine and oxidative stress, both *in vitro* and *in vivo* [36–38]. In our context, changes observed in serum for gelsolin, plasma glutathione peroxidase and afamin, and in the PMN themselves for superoxide dismutase (SOD) (an enzyme that scavenges the superoxide anion), seem to indicate a clear disturbance of the oxidative stress balance and a series of potential compensatory mechanisms occurring in the organism.

Gelsolin is an actin modulating protein that exists both as cytoplasmic and plasma secreted forms, which is known to be reduced in pulmonary failures. In a murine model of

oxidant injury [39], the degree of hyperoxic injury was inversely related to gelsolin levels suggesting an association between gelsolin depletion and the pathogenesis of acute oxidant lung injury in patients. Decrease in the activity of blood and erythrocyte glutathione peroxidase (a key enzyme secreted in plasma that catalyses reduction of hydrogen peroxide) has been also observed in studies conducted on smokers, where oxidative stress is clearly exacerbated [40–42]. In two of the above studies [40, 41], a parallel reduction in the plasma level of vitamin E has been shown. Interestingly, in our study we have observed a concomitant increase in the vitamin E binding protein afamin that has been shown to be able to increase the survival of cortical neurons under oxidative stress *in vitro* that can be read as a compensatory mechanism. In the third study [42] a significant increase of the erythrocyte activity of SOD was concomitantly found in smokers, interpreted as an adaptation process counteracting the pro-oxidant/antioxidant imbalance in the blood of smokers, that was also paralleled in old smokers by a change in the leukocyte stimulation index. Similarly, the observed increase in SOD of PMN could also be seen as a parallel mechanism resulting in a compensation for the increased oxidative stress as well as for the increased leukocyte rolling and adhesion. Indeed, in murine models [35] it was observed that nicotine-induced rolling and adhesion can be effectively inhibited by SOD, suggesting a key role of superoxide anion in this event.

4.4 Direct and indirect action of nicotine in the periphery

All of the changes highlighted in the blood by our proteomic approach could be a direct reflection of the interaction with non-neuronal binding sites present in peripheral tissues, including leukocytes themselves, or an indirect impact on the periphery mediated through the central nervous system (CNS). Indeed, there is a large body of evidence that indicates a bidirectional communication between the CNS and the immune system, and, in addition to the potential direct effects on nicotine on the leukocyte response, a CNS-mediated strong modulation of the neuroendocrine and immune systems can be expected. A series of studies on morphine and nicotine have shown that drugs of abuse can modulate both immune cell activities and hypothalamic-pituitary-adrenal (HPA) axis function by acting through specific receptors. Morphine for instance suppresses the immune system *via* activation of central opioid receptors, whilst nicotine and nicotinic agonists appear to act predominantly at peripheral receptors, presumably at the autonomic ganglia [43].

Evidence of the involvement of nicotinic receptors in linking inflammation and leukocyte functions with oxidative stress arises from studies on the effects of nicotine agonists on microcirculation. As an example, the nicotine-induced leukocyte rolling and adhesion can be blocked in a murine model by mecamylamine, as well as by anti-P-selectin anti-

bodies, and by SOD, suggesting the action through a ganglionic-type nicotinic receptor and *via* P-selectin and reactive oxygen radical-dependent mechanisms [35].

The overall effect of nicotine on oxidative stress can be mediated through different mechanisms and result in opposite effects (pro-oxidant and antioxidant properties) depending on the system being investigated [44]. Indeed, nicotine administration can result in oxidative stress upon induction of the generation of reactive oxygen species in the periphery and in CNS. In parallel, it has been shown *in vitro* that nicotine displays intracellular antioxidant properties through the activation of nicotinic receptors [45, 46] but also extracellularly by acting as a iron-binding radical scavenger [47], therefore suggesting at the CNS level a potential role for nicotine in neuroprotection. In the light of all the above indications, our results point the existence of a number of potentially different mechanisms triggered by nicotine, both mediated by binding to specific central and peripheral receptors, or by additional chemical or pharmacological properties of nicotine, that can result in physiologically opposite actions.

5 Concluding remarks

This is the first investigation on a proteomic scale of the changes caused by nicotine in the peripheral blood of laboratory animals. With our global approach, we have chosen not to dissect out a single mechanism, or to focus on a specific target district (that could result in an more precise but less comprehensive observation), but to provide a wider picture of the overall effects of a “physiologically” relevant nicotine dose regimen in rodents. By selecting plasma and its leukocyte components we aimed at identifying potential noninvasive markers for the CNS and non-CNS mediated activities of nicotine.

We have shown by a large-scale proteomic analysis that chronic treatment with nicotine has an overall impact on a series of proteins involved in a variety of cellular and metabolic pathways. The main effect seems to be a general reduction in the inflammatory response with a concomitantly increased imbalance of the oxidative stress metabolism in the periphery. Both imply a potential impact on leukocyte activation in different directions that is reflected also by changes in their cytoskeleton organisation. This approach has therefore highlighted a series of potential changes reflecting *in vivo* centrally and peripherally mediated nicotine effects that deserve to be further tested in additional models, such as for instance inflammation models, or by using nicotine antagonists and in dose-response studies, to better understand the physiological mechanisms underlying the observed changes. In perspective, similar studies in humans can open up an unprecedented source of protein changes to be validated as noninvasive markers for the neuroimmunomodulatory activities exerted by the psychoactive component of smoke.

We thank the GSK Laboratory Animal Sciences group for help with experimental animals and Mr. Claudio Righetti for the technical support in cytofluorometer analysis. PGR is supported by GSK grant N° GEC-556-2000, by FIRB 2001 (contract No. RBNE01KJHT) and PRIN 2003 (MURST, Rome).

6 References

- [1] Benowitz, N. L., *Bri. J. Addict.* 1991, 86, 495–499.
- [2] Stolerman, I. P., Jarvis, M. J., *Psychopharmacology* 1995, 117, 2–10.
- [3] Wessler, I., Kirkpatrick, C. J., Racke, K., *Clin. Exp. Pharmacol. Physiol.* 1999, 26, 198–205.
- [4] Grando, S. A., Horton, R. M., Pereira, E. F., Diethelm-Okita, B. M. *et al.*, *J. Invest. Dermatol.* 1995, 105, 774–781.
- [5] Maus, A. D., Pereira, E. F., Karachunski, P. I., Horton, R. M. *et al.*, *Mol. Pharmacol.* 1998, 54, 779–788.
- [6] Macklin, K. D., Maus, A. D., Pereira, E. F., Albuquerque, E. X., Conti-Fine, B. M., *J. Pharm. Exp. Ther.* 1998, 287, 435–439.
- [7] Benowitz, N. L., *New Engl. J. Med.* 1988, 319, 1318–1330.
- [8] Fattering, K., Verotta, D., Benowitz, N. L., *J. Pharm. Exp. Ther.* 1997, 281, 1238–1246.
- [9] Haass, M., Kubler, W., *Cardiovasc. Drugs Ther.* 1997, 10, 657–665.
- [10] Rogers, K. R., Fernando, J. C., Thompson, R. G., Valdes, J. J., Eldefrawi, M. E., *Anal. Biochem.* 1992, 202, 111–116.
- [11] Mousavi, M., Hellstrom-Lindhahl, E., Guan, Z. Z., Bednar, I., Nordberg, A., *Life Sci.* 2001, 70, 577–590.
- [12] Heesch, C., Weis, M., Aicher, A., Dimmeler, S., Cooke, J. P., *J. Clin. Invest.* 2002, 110, 527–536.
- [13] Heesch, C., Weis, M., Cooke, J. P., *J. Am. Coll. Cardiol.* 2003, 41, 489–496.
- [14] Sopori, M. L., Kozak, W., Savage, S. M., Geng, Y., Kluger, M. J., *Adv. Exp. Med. Biol.* 1998, 437, 279–289.
- [15] Sopori, M. L., Kozak, W., *J. Neuroimmunol.* 1998, 83, 148–156.
- [16] Kalra, R., Singh, S. P., Kracko, D., Matta, S. G. *et al.*, *J. Pharm. Exp. Ther.* 2002, 302, 935–939.
- [17] Geng, Y., Savage, S. M., Razani-Boroujerdi, S., Sopori, M. L., *J. Immunol.* 1996, 156, 2384–2390.
- [18] Balfour, D. J., Benwell, M. E., Birrell, C. E., Kelly, R. J., Al Aloul, M., *Pharmacol. Biochem. Behavior* 1998, 59, 1021–1030.
- [19] Bunnemann, B., Terron, A., Zantedeschi, V., Merlo-Pick, E., Chiamulera, C., *Eur. J. Pharmacol.* 2000, 393, 249–253.
- [20] Sbarbati, A., Bunnemann, B., Cristofori, P., Terron, A. *et al.*, *Eur. J. Neurosci.* 2002, 16, 877–882.
- [21] Geng, Y., Savage, S. M., Johnson, L. J., Seagrave, J., Sopori, M. L., *Toxicol. Appl. Pharmacol.* 1995, 135, 268–278.
- [22] Piubelli, C., Galvani, M., Hamdan, M., Domenici, E., Righetti, P. G., *Electrophoresis* 2002, 23, 298–310.
- [23] Perkins, D. N., Pappin, D. J., Creasy, D. M., Cottrell, J. S., *Electrophoresis* 1999, 20, 3551–3567.
- [24] Eberini, I., Miller, I., Gemeiner, M., Haynes, P. *et al.*, *Electrophoresis* 1999, 20, 3599–3602.
- [25] Righetti, P. G. in: Righetti, P. G., Stoyanov, A., Zhukov, M. (Eds.), *The Proteome Revisited: Theory and Practice of the Relevant Electrophoresis Steps*, Elsevier, Amsterdam 2001, pp. 300–389.
- [26] Geuskens, M., Torres, J. M., Esteban, C., Uriel, J., *Microsc. Res. Tech.* 1994, 28, 297–307.
- [27] Singh, S. P., Kalra, R., Puttfarcken, P., Kozak, A. *et al.*, *Toxicol. Appl. Pharmacol.* 2000, 164, 65–72.
- [28] Sopori, M. in: Arneric, S. P., Brioni, J. D. (Eds.) *Neuronal Nicotinic Receptors*, Wiley-Liss, New York 1999, pp. 197–209.
- [29] Chao, J., Miao, R. Q., Chen, V., Chen, L. M., Chao, L., *Biol. Chem.* 2001, 382, 15–21.
- [30] Soury, E., Olivier, E., Daveau, M., Hiron, M. *et al.*, *Biochem. Biophys. Res. Commun.* 1998, 243, 522–530.
- [31] Pineiro, M., Alava, M. A., Gonzalez-Ramon, N., Osada, J. *et al.*, *Biochem. Res. Commun.* 1999, 263, 224–229.
- [32] Mor-Vaknin, N., Punturieri, A., Sitwala, K., Markovitz, D. M., *Nat. Cell Biol.* 2003, 5, 59–63.
- [33] Languino, L. R., Duperray, A., Joganic, K. J., Fornaro, M. *et al.*, *Proc. Nat. Acad. Sci. USA* 1995, 92, 1505–1509.
- [34] Nowak, D., Ruta, U., Piasecka, G., *Exp. Pathol.* 1990, 39, 37–43.
- [35] Yong, T., Zheng, M. Q., Linthicum, D. S., *J. Neuroimmunol.* 1997, 80, 158–164.
- [36] Gillespie, M. N., Owasoyo, J. O., Kojima, S., Jay, M., *Toxicology* 1987, 45, 45–52.
- [37] Yildiz, D., Ercal, N., Armstrong, D. W., *Toxicology* 1998, 130, 155–165.
- [38] Ashakumary, L., Vijayammal, P. L., *Pharmacology* 1996, 52, 153–158.
- [39] Christofidou-Solomidou, M., Scherpereel, A., Solomides, C. C., Muzykantov, V. *et al.*, *Lung* 2002, 180, 91–104.
- [40] Zhou, J. F., Yan, X. F., Guo, F. Z., Sun, N. Y. *et al.*, *Biomed. Environ. Sci.* 2000, 13, 44–55.
- [41] Abou-Seif, M. A., *J. Biochem. Toxicol.* 1996, 11, 133–138.
- [42] Hulea, S. A., Olinescu, R., Nita, S., Crocnan, D., Kummerow, F. A., *J. Environ. Pathol. Toxicol. Oncol.* 1995, 14, 173–180.
- [43] Mellon, R. D., Bayer, B. M., *J. Pharm. Exp. Ther.* 1999, 288, 635–642.
- [44] Newman, M. B., Arendash, G. W., Shytle, R. D., Bickford, P. C. *et al.*, *Life Sci.* 2002, 71, 2807–2820.
- [45] Guan, Z. Z., Yu, W. F., Nordberg, A., *Neurochem. Int.* 2003, 43, 243–249.
- [46] Liu, Q., Zhao, B., *Br. J. Pharmacol.* 2004, 141, 746–754.
- [47] Ferger, B., Spratt, C., Earl, C. D., Teismann, P. *et al.*, *Naunyn-Schmiedeberg's Arch. Pharmacol.* 1998, 358, 351–359.

ORIGINAL ARTICLE

Characterization of the Signaling Pathway Downstream p75 Neurotrophin Receptor Involved in β -Amyloid Peptide-Dependent Cell Death

Claudio Costantini,^{*1} Filippo Rossi,¹ Elena Formaggio,² Roberto Bernardoni,³ Daniela Cecconi,⁴ and Vittorina Della-Bianca¹

¹Department of Pathology, Section of General Pathology, University of Verona, Strada Le Grazie 8, 37134 Verona, Italy; ²Department of Medicine and Public Health, Section of Pharmacology, University of Verona, P. le Scuro 10, 37134 Verona, Italy; ³Department of Biology, University of Bologna, via Selmi 3, 40126 Bologna, Italy; and ⁴Department of Agricultural and Industrial Biotechnologies, University of Verona, Strada Le Grazie 15, 37134 Verona, Italy

Received April 23, 2004; Accepted August 14, 2004

Abstract

The accumulation of β -amyloid (A β) peptide is a key pathogenic event in Alzheimer's disease. Previous studies have shown that A β peptide can damage neurons by activating the p75 neurotrophin receptor (p75^{NTR}). However, the signaling pathway leading to neuronal cell death is not completely understood. By using a neuroblastoma cell line devoid of neurotrophin receptors and engineered to express either a full-length or a death domain (DD)-truncated form of p75^{NTR}, we demonstrated that A β peptide activates the mitogen-activated protein kinases (MAPKs) p38 and c-Jun N-terminal kinase (JNK). We also found that A β peptide induces the translocation of nuclear factor- κ B (NF- κ B). These events depend on the DD of p75^{NTR}. β -Amyloid (A β) peptide was found not to be toxic when the above interactors were inhibited, indicating that they are required for A β -induced neuronal cell death. p75 neurotrophin receptor (p75^{NTR})-expressing cells became resistant to A β toxicity when transfected with dominant-negative mutants of MAPK kinases 3, 4, or 6 (MKK3, MKK4, or MKK6), the inhibitor of κ B α , or when treated with chemical inhibitors of p38 and JNK. Furthermore, p75^{NTR}-expressing cells became resistant to A β peptide upon transfection with a dominant-negative mutant of p53. These results were obtained in the presence of normal p38 and JNK activation, indicating that p53 acts downstream of p38 and JNK. Finally, we demonstrated that NF- κ B activation is dependent on p38 and JNK activation. Therefore, our data suggest a signaling pathway in which A β peptide binds to p75^{NTR} and activates p38 and JNK in a DD-dependent manner, followed by NF- κ B translocation and p53 activation.

Index Entries: A β ; p75^{NTR}; p38; JNK; NF- κ B; p53.

Introduction

Alzheimer's disease (AD) is a neurodegenerative disorder, characterized by progressive loss of neuronal plasticity in the hippocampus and cortical areas of the brain (Dawbarn and Allen, 2001).

Accumulation of β -amyloid (A β) peptide, as consequence of increased production from the precursor amyloid precursor protein and/or reduced removal, is one of the key pathogenic events for the neuronal loss (Dawbarn and Allen, 2001; Hardy and Selkoe, 2002).

*Author to whom all correspondence and reprint requests should be addressed. E-mail: claudio.costantini@medicina.univr.it

Several data using primary neurons or neuronal cell lines have shown that A β peptide is neurotoxic in vitro, and this cytotoxicity correlates with its β -sheet structure and fibrillar state (Iversen et al., 1995). The molecular mechanism(s) by which extracellular A β peptide damages neurons is still mostly unknown. Besides disrupting membrane integrity (McLaurin and Chakrabartty, 1996), A β peptide binds to a variety of cell-surface receptors (Boland et al., 1996; El Khoury et al., 1996; Yan et al., 1996; Lorenzo et al., 2000; Wang et al., 2000; Kajkowski et al., 2001; Tiffany et al., 2001), including p75 neurotrophin receptor (p75^{NTR}) (Yaar et al., 1997; Kuner et al., 1998). p75 neurotrophin receptor (p75^{NTR}) belongs to the family of death receptors; it binds to nerve growth factor (NGF) and other neurotrophins and is involved in several different biological activities, including apoptosis (Dechant and Barde, 2002). In a previous paper (Perini et al., 2002), we confirmed and extended the importance of p75^{NTR} in A β -mediated neurotoxicity demonstrated previously by Yaar et al. (1997) and Kuner et al. (1998).

The molecular mechanism by which p75^{NTR} induces neuronal damage is still elusive (Rabizadeh and Bredesen, 2003). Studies performed on several models of p75^{NTR}-dependent cell death have demonstrated the involvement of c-Jun N-terminal kinase (JNK) activity (Casaccia-Bonnet et al., 1996; Aloyz et al., 1998; Yoon et al., 1998; Friedman, 2000; Brann et al., 2002; Harrington et al., 2002; Salehi et al., 2002). The JNK pathway is also activated when p75^{NTR}-expressing neurons are treated with A β peptide, suggesting a link between JNK activation and cell death (Yaar et al., 2002). More recently, chemical inhibition of JNK was shown to protect neuronal hybrid cells from p75^{NTR}-dependent neurotoxicity induced by A β (Tsukamoto et al., 2003). Finally, mitogen-activated protein kinase (MAPK) phosphorylation has been documented in neurons from both AD patients and animal models of AD (Pei et al., 2001; Savage et al., 2002; Zhu et al., 2003), and in vitro studies have shown that A β -induced cell death is associated with an increase in MAPK activity (Behrens et al., 1999; Bozyczko-Coyne et al., 2001; Daniels et al., 2001; Morishima et al., 2001; Troy et al., 2001; Jang and Surh, 2002; Wei et al., 2002; Fogarty et al., 2003; Minogue et al., 2003; Tamagno et al., 2003a, 2003b).

The role of nuclear factor- κ B (NF- κ B) in the p75^{NTR}-activated cell death pathway is less clear. Though several models indicate a survival function for NF- κ B (Carter et al., 1996; Ladiwala et al., 1998; Yoon et al., 1998; Hamanoue et al., 1999; Khursigara et al., 1999; Foehr et al., 2000; Gentry et al., 2000; Khursigara et al.,

2001; Bui et al., 2002), Kuner et al. (1998) have demonstrated that A β can induce apoptosis in a neuroblastoma cell line by activating NF- κ B through p75^{NTR}.

Here, we used a neuroblastoma cell line devoid of all neurotrophin receptors and stably transfected with either full-length or a death domain (DD)-truncated form of p75^{NTR} to investigate the signaling pathway activated during the p75^{NTR}-mediated neuronal death induced by A β peptide. Our findings are as follows: (1) A β peptide induces activation of p38 and JNK, and translocation of NF- κ B via the DD of the intracellular region of p75^{NTR}; (2) translocation of NF- κ B requires activation of p38 and JNK; and (3) p53 is required for p38- and JNK-mediated neuronal death.

Materials and Methods

Materials

A β (25–35) and A β (35–25) were from Bachem AG. A β (25–35) was dissolved at 1.5 mM in PBS. A β aggregation in fibrils was monitored by using the thioflavine test (Perini et al., 2002). SB203580 and SP600125 were from Calbiochem. Anti-hemagglutinin (HA) mouse monoclonal antibody (MAb) (clone 12CA5) was purchased from Roche Molecular Biochemicals. Anti-FLAG M2 MAb was purchased from Sigma. Anti-phospho-p38 MAPK, anti-p38 MAPK, anti-phospho-stress activated protein kinase (SAPK)/JNK, and anti-SAPK/JNK antibodies were purchased from Cell Signalling, Inc. Anti-p53 (Bp53-12) mouse MAb was purchased from Santa Cruz Biotechnology.

Cell Clones

Human neuroblastoma SK-N-BE cells, which express neither p75^{NTR} nor tropomyosin-related kinase A (BENTR-free), and the derived cell lines expressing either full-length p75^{NTR} (BEp75) or a DD-truncated form of p75^{NTR} (lacking amino acids 352–427) (BEp75 Δ DD), have been described previously (Perini et al., 2002). BEp75 cells were further transfected with the following plasmids: dominant-negative pcDNA3-HA-MKK3dn and pcDNA3-HA-MKK6dn (a generous gift from Dr. T. Matsumoto, Uppsala University, Sweden), dominant-negative pcDNA3-FLAG-MKK4dn (a generous gift from Dr. A. Atfi, Centre INSERM, Hopital Saint-Antoine, Paris, France), and pCMV-neo-BamHI-p53R175H (a generous gift from Dr. G. Blandino, Regina Elena Cancer Institute, Rome, Italy). As control, BEp75 cells were also transfected with the empty vectors described above.

Au: "As control, BEp75 cells..." Punctuation OK? Otherwise, sentence not complete.

Transfected cells were selected in RPMI medium supplemented with fetal bovine serum (FBS) (15% v/v), 2 mM glutamine, 50 μ g/mL gentamycin, 150 μ g/mL hygromycin, and 300 μ g/mL G418 (Roche Molecular Biochemicals). Stably transfected cells were characterized for expression of the transfected plasmid, as described in the Western blot analysis section. BEp75 cells stably expressing the FLAG-inhibitor of κ B α (IkB α) super-repressor (IkB α SR)-internal ribosomal entry site (IRES)-Neo cassette (BEp75/IkB α SR) with G418 resistance (500 μ g/mL) were a generous gift from Dr. K. Marcu (State University of New York, Stony Brook, New York).

Experimental Protocol

Cells were plated at a density of 10,000 cells/cm² for assessment of cell damage and at 0.6×10^5 /cm² for Western blot analysis of p38 and JNK phosphorylation and for electrophoretic mobility shift assay (EMSA). Regular growth medium was replaced with fresh RPMI medium containing FBS (1% v/v) and A β (25–35) at the beginning of each treatment. When required, inhibitors of p38 and JNK activity were added 1 h before A β (25–35) treatment.

Preparation of Cell Lysates

For analysis of phospho-p38 and phospho-JNK, cells were collected, washed with cold PBS containing 1 mM sodium orthovanadate and 50 mM sodium fluoride, and lysed in 25–50 μ L of buffer (20 mM HEPES at pH 7.9, 420 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1% NP-40, 20% glycerol, 20 μ M penylarsine oxide (PAO), 1 mM sodium orthovanadate, 50 mM sodium fluoride, and 1 mM DTT), containing protease inhibitors 5 μ g/mL leupeptin, (5 μ g/mL pepstatin, and 1 mM PMSF). For detection of HA- and FLAG-tagged proteins, cells were collected, washed with PBS, and lysed in buffer (20 mM Tris-HCl at pH 7.4, 200 mM NaCl, and 1% NP-40) containing protease inhibitors (Complete Mini, Roche Molecular Biochemicals). After 20 min incubation on ice, lysates were centrifuged at 12,000g for 20 min at 4°C. Supernatants were collected, and an aliquot was taken for protein determination using a protein assay kit (Bio-Rad).

Preparation of Nuclear Extracts

Cells were collected, washed with cold PBS containing 1 mM sodium orthovanadate and 50 mM sodium fluoride, and lysed in 50 μ L of low-salt buffer (10 mM HEPES at pH 8.0, 50 mM NaCl, 1 mM EDTA, 1 mM DTT, and 0.5 mM PMSF) containing protease inhibitors. After 10 min incubation on ice, 50 μ L of low-salt buffer, supplemented with 0.2% NP-40, was added

to each sample, followed by an additional 5 min incubation on ice. Samples were centrifuged at 3,000g for 5 min at 4°C. Nuclear proteins were extracted by resuspending nuclei in 25 μ L of high-salt buffer (20 mM HEPES at pH 8.0, 420 mM NaCl, 1 mM EDTA, 1 mM DTT, 0.5 mM PMSF, and 10% glycerol) containing protease inhibitors. After centrifugation (12,000g for 10 min at 4°C), nuclear extracts were recovered and an aliquot was taken for protein determination.

Western Blot Analysis

Samples were run on 12% Tris/glycine gels and transferred to Hybond ECL nitrocellulose membranes (Amersham Pharmacia Biotech). Membranes were blocked in 5% nonfat dry milk/TBS-Tween (TBS-T) for 1 h, followed by overnight incubation with the primary antibody in the appropriate dilution buffer. Specifically, anti-phospho-p38 was used in 5% BSA/TBS-T, whereas anti-phospho-JNK, anti-HA, anti-FLAG, and anti-p53 in 5% nonfat dry milk/TBS-T. Membranes were then washed in TBS-T and incubated with horseradish peroxidase (HRP)-conjugated anti-mouse or anti-rabbit IgG diluted in 5% nonfat dry milk/TBS-T or 5% BSA/TBS-T for 2 h. Bound HRP-conjugated secondary antibody was detected with the ECL system (Amersham), according to the manufacturer's instructions.

EMSA

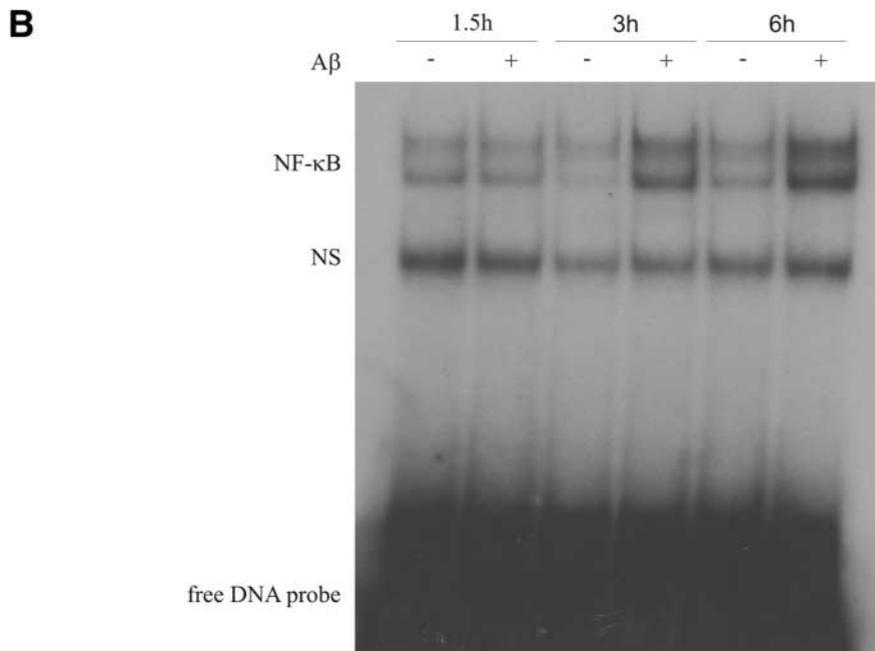
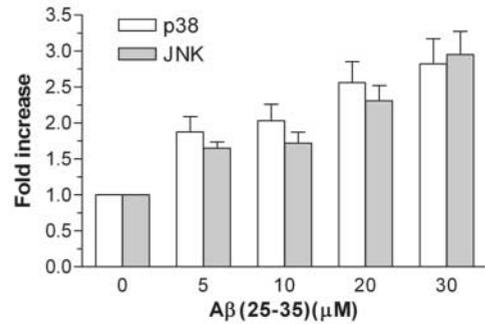
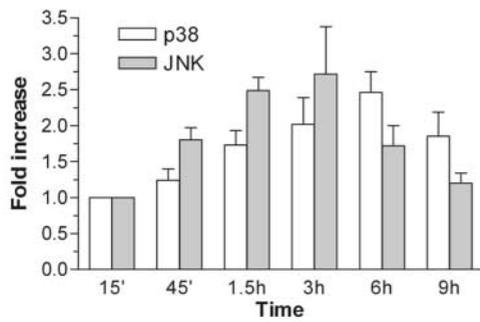
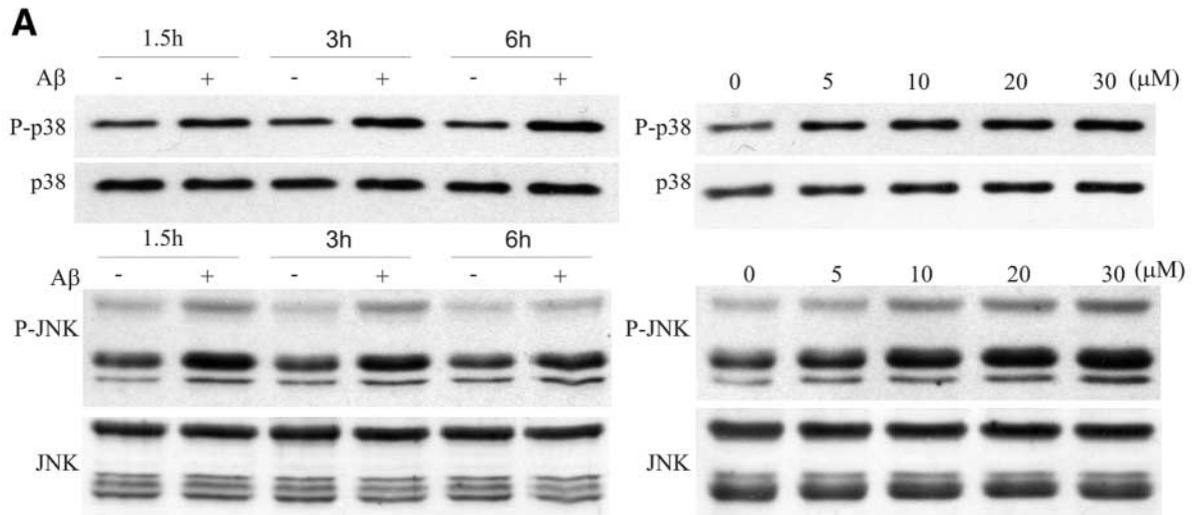
Nuclear extracts (5–10 μ g) were incubated on ice for 15 min in a buffer containing 20 mM HEPES at pH 8.0, 50 mM KCl, 0.2 mM EDTA, 0.5 mM DTT, 0.5 mM PMSF, 10% glycerol, 1 μ g of poly(dI-dC) (Roche Molecular Biochemicals), and 5×10^4 cpm of ³²P-labeled DNA probe corresponding to the consensus binding site for NF- κ B/c-Rel homodimeric and heterodimeric complexes (Santa Cruz Biotechnology). Reaction products were analyzed on a 5% polyacrylamide, 0.5 \times TBE, gel.

Assessment of Cell Damage

Cell damage was analyzed by epifluorescence microscopy after staining cells with a 1:1 (v/v) solution of acridine orange (filter setting for FITC) and ethidium bromide (filter setting for rhodamine) (both at 0.1 mg/mL in PBS; Molecular Probes), a procedure that reveals both apoptosis and necrosis (Perini et al., 2002).

Expression of p75^{NTR}

BEp75 cells and derived cell clones in suspension were first treated for 1 h at 4°C with primary MAb 8211 anti-p75^{NTR} (Chemicon Int., Inc.). After cell washing, the secondary phycoerythrin-conjugated



goat F(ab')₂ anti-mouse immunoglobulin (Southern Biotechnology Associates) was added for 1 h at 4°C. Cytofluorographic analysis was performed on a FACScan™ (Becton Dickinson) using CELLQuest™ software.

Densitometric Analysis

Band intensity was determined with a GS-710 densitometer (Bio-Rad) using the software Quantity One (Bio-Rad).

Statistical Analysis

One-way ANOVA with Dunnett's post-test was performed using GraphPad Prism (v 4 for Windows) GraphPad Software.

Results

A β Activates p38 and JNK in BEp75 Cells and Induces Translocation of NF- κ B

We have already shown (Perini et al., 2002) that the p75^{NTR}-dependent induction of cell death is similar for both the physiologically produced A β (1–40 and 1–42) and the nonphysiological, but active motif-containing, A β (25–35). Therefore, all of the experiments were performed by using only A β (25–35).

Neuroblastoma cell lines stably transfected with full-length p75^{NTR} (BEp75) were treated with A β peptide for different times, and the levels of phosphorylated p38 were then determined in cell lysates. The results (Fig. 1A) showed a sustained activation of p38 between 1.5 and 6 h, which persisted for ~9 h. p38 activation was dose-dependent (Fig. 1A, right panel), with a plateau at 20 μ M A β . Because no difference was observed between 20 and 30 μ M A β , all subsequent experiments were performed at 20 μ M. Similar results were observed when we analyzed JNK activation. The time course indicates a peak of activation between 1.5 and 3 h (Fig. 1A), which was also dose-dependent (Fig. 1A).

To analyze whether A β peptide also activates NF- κ B, nuclear extracts were tested for NF- κ B/DNA complex formation following A β treatment. Induction of NF- κ B translocation was observed after 3 h stimulation (Fig. 1B), confirming that A β can stimulate NF- κ B activity (Kuner et al., 1998). The reverse-order A β peptide, used as control, was unable to activate either p38 or JNK, or to induce NF- κ B translocation (data not shown).

The DD of p75^{NTR} Is Necessary for p38 and JNK Activation, and for NF- κ B Translocation

To confirm the role of p75^{NTR} in the A β -mediated activation of p38 and JNK, and translocation of NF- κ B, we used neuroblastoma cell lines devoid of all neurotrophin receptors (BENTR-free) or stably transfected with a p75^{NTR} construct lacking the DD (BEp75 Δ DD).

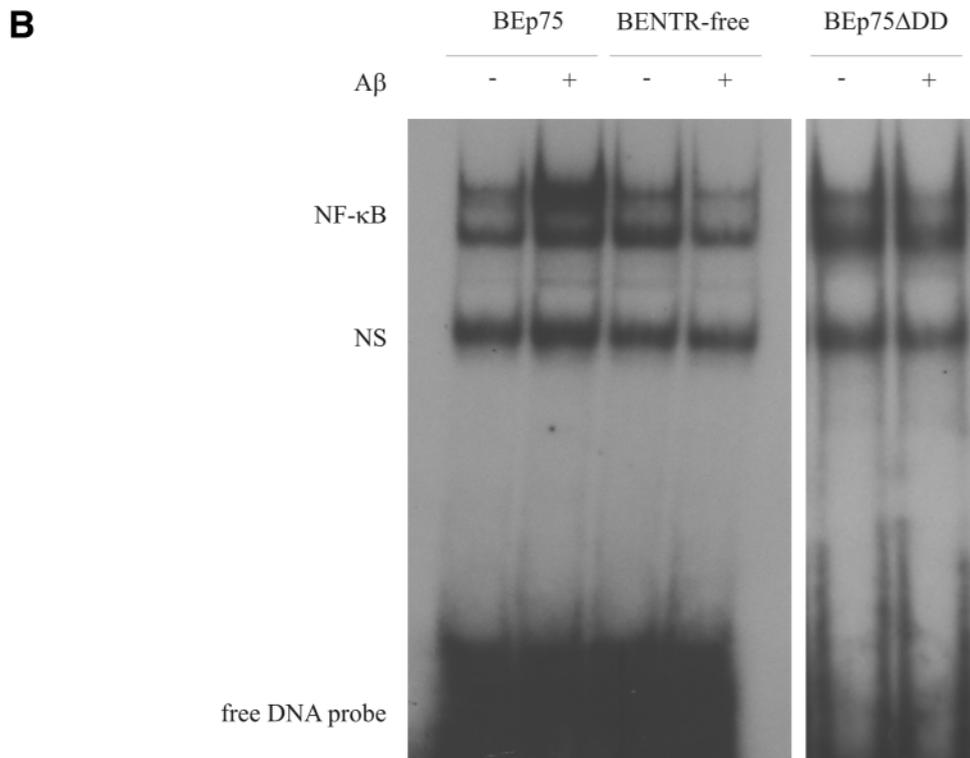
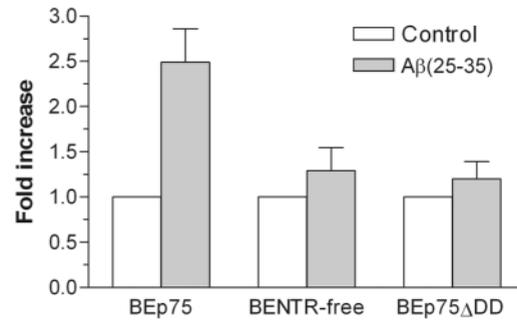
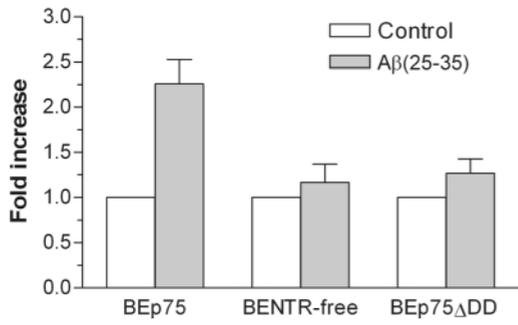
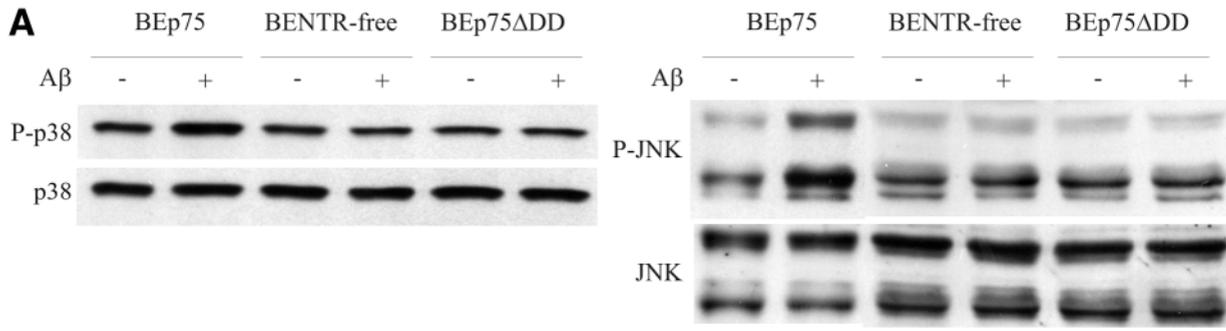
As shown in Fig. 2A, A β was able to activate p38 and JNK in cells expressing full-length p75^{NTR} (BEp75) but not in those devoid of neurotrophin receptors (BENTR-free). In addition, A β failed to activate p38 and JNK in cells transfected with p75^{NTR} lacking the DD (BEp75 Δ DD). Similar results were obtained for NF- κ B translocation: A β peptide was able to activate NF- κ B in BEp75 cells but not in BENTR-free and BEp75 Δ DD cells (Fig. 2B).

Taken together, these results indicate that p75^{NTR} is necessary for p38 and JNK activation, as well as for NF- κ B nuclear translocation, under A β -stimulation. They also suggest that such events require a functional DD.

p38 and JNK Activation, and NF- κ B Translocation Are Involved in A β -Induced Cell Death

We next investigated whether the above events are directly involved in p75^{NTR}-dependent cell death. For this purpose, we first analyzed the effect of A β peptide on BEp75 cells under conditions in which p38 and JNK activation are blocked. It is worth

Fig. 1. A β -induced p38 and JNK activation, and NF- κ B translocation in BEp75 cells. **(A)** BEp75 cells were treated with 20 μ M A β for the indicated times (left panel) or with different doses of A β (right panel) for 6 h (phospho [P]-p38 dose response) or 3 h (P-JNK dose response). Cells were lysed and subjected to Western blot analysis with anti-phospho-p38 and anti-phospho-JNK antibodies. Membranes were stripped and reprobbed with antibodies against total p38 and JNK to normalize the samples for the total amount of p38 and JNK kinases (normalized P-p38 and P-JNK are calculated by the ratio between P-p38 or P-JNK and the corresponding p38 or JNK densitometric values). The results are indicated at the bottom of both panels as mean values \pm S.D. of two to six experiments of the ratio between normalized P-p38 or P-JNK in treated samples (indicated with + or with a dose value [top panels]) and normalized P-p38 or P-JNK in control samples (indicated with – or 0 [top panels]). **(B)** BEp75 cells were treated with 20 μ M A β for the indicated times and lysed, and equal amounts of nuclear protein extracts were subjected to EMSA. An autoradiograph of a typical experiment representative of three is shown. The positions of NF- κ B-DNA complexes (NF- κ B), nonspecific protein-DNA complexes (NS), and unbound DNA oligonucleotides (free DNA probe) are indicated.



remembering that the MAPK signaling system is composed of three kinases that establish a sequential activation pathway (Johnson and Lapadat, 2002). p38 is mainly activated by MAPK kinase 3 (MKK3) and MKK6, whereas JNK is activated by MKK4. To inhibit p38 activation, we stably transfected BEp75 cells with HA-tagged, dominant-negative mutants of MKK3 (BEp75/MKK3dn) and MKK6 (BEp75/MKK6dn); to inhibit JNK activation, we stably transfected BEp75 cells with a dominant-negative FLAG-tagged version of MKK4 (BEp75/MKK4dn). Exposure of cells transfected with the above dominant-negative mutants to A β peptide resulted in a decrease of cell death, whereas no difference was observed in cells transfected with the empty vector (Fig. 3C). Results were validated by confirming that (1) the transfected plasmids were effectively transcribed and translated (Fig. 3A, left panel); (2) the expression of p75^{NTR} was not modified as a consequence of the transfection and expression of the constructs (Fig. 3A, right panel); and (3) transfected dominant-negative MKK3/6 and MKK4 depressed A β -induced phosphorylation of p38 and JNK, respectively (Fig. 3B).

To further confirm the above results, we analyzed the effect of A β on BEp75 cells in the presence of either SB203580, a specific inhibitor of p38 activity, or SP600125, a specific inhibitor of JNK activity. The results (Fig. 3D) show that both SB203580 and SP600125 pretreatments caused a marked decrease of A β -induced cell death.

Finally, we confirmed the involvement of NF- κ B in p75^{NTR}-dependent cell death under A β treatment by using BEp75 cells stably expressing the I κ B α SR-IRES-Neo cassette (BEp75/I κ B α SR). Inhibitor of κ B α (I κ B α) (Ser32Ala/Ser36Ala) functions as a dominant-negative protein that remains tethered to NF- κ B subunits even in response to extracellular NF- κ B-activating stimuli (Hamanoue et al., 1999). Expression of the mutant protein was verified by Western blot (Fig. 4A, upper panel), whereas its ability to abrogate NF- κ B translocation was demonstrated by EMSA (Fig. 4B). Exposure of BEp75/I κ B α SR cells to A β resulted in decreased activation of cell death (Fig. 4D), when compared with BEp75 cells having

similar levels of p75^{NTR} expression (Fig. 4A, lower panel). Activation of p38 and JNK by A β in BEp75/I κ B α SR cells proved to be at levels comparable with those observed in nontransfected BEp75 cells (Fig. 4C), thus indicating that the decrease of cell death by A β was only attributable to lack of NF- κ B activation.

p53 Activation Is Involved in A β -Induced Cell Death

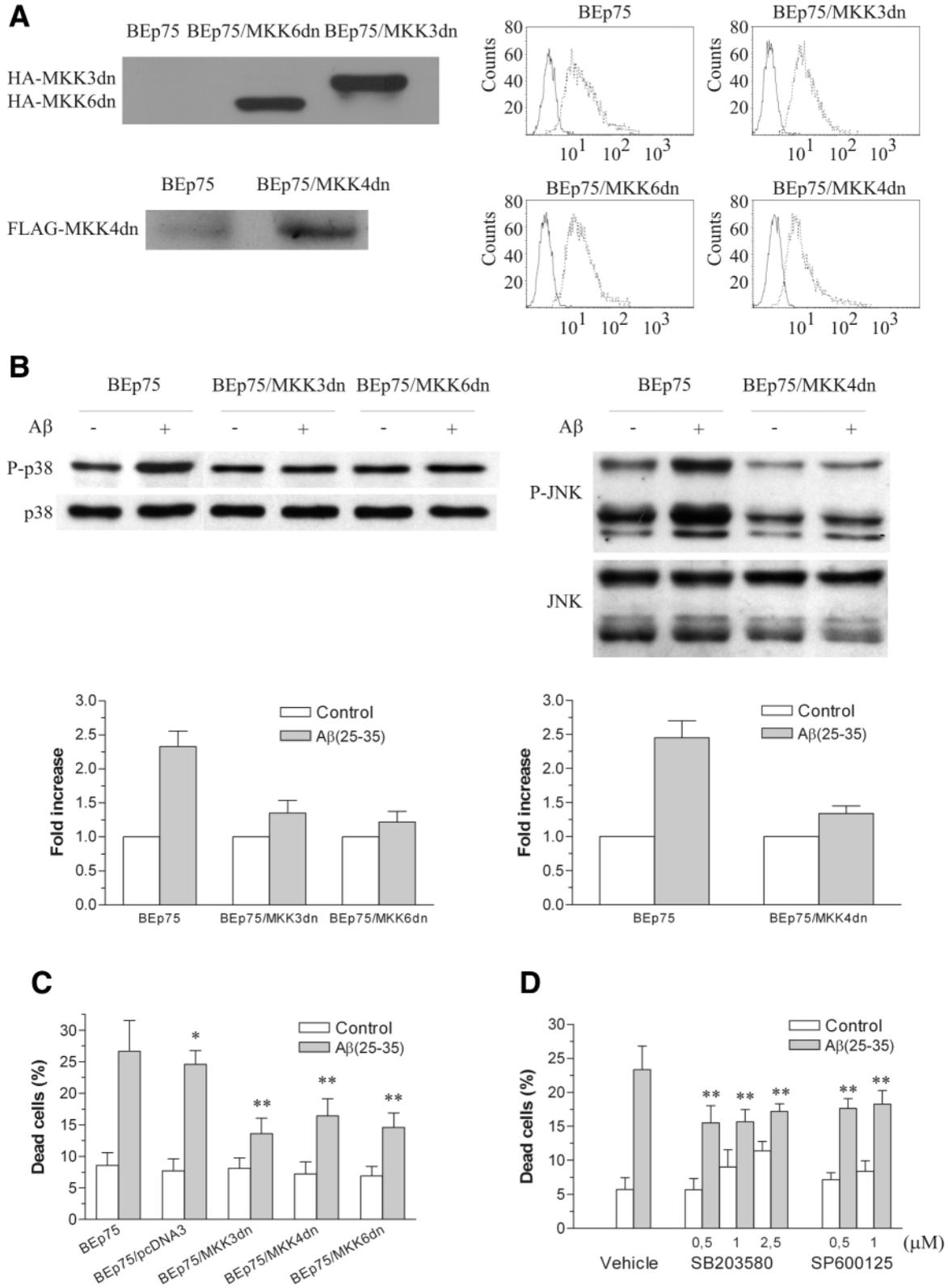
After demonstrating the involvement of p38 and JNK in A β -induced cell death, we analyzed the mechanism by which they mediate cell death. It is known that p53, a possible substrate for p38 and JNK, is involved in neuronal apoptosis under various stimuli (Aloyz et al., 1998; Kim et al., 2002; Zhu et al., 2002; Shimada et al., 2003). To investigate the relationship between p53 and A β -induced p75^{NTR}-dependent cell death, BEp75 cells were stably transfected with a dominant-negative version of p53 (BEp75/p53R175H). Figure 5A,B shows that (1) the transfected plasmid was effectively transcribed and translated (Fig. 5A, left panel), (2) the expression of p75^{NTR} was not modified as a consequence of the transfection (Fig. 5A, right panel), and (3) the activation of p38 and JNK by A β in the transfected cells was similar to BEp75 cells (Fig. 5B). Exposure of cells transfected with the above dominant-negative mutant to A β resulted in a marked decrease of cell death, whereas no difference was observed in cells transfected with the empty vector (Fig. 5C).

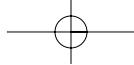
These results indicate that p53 is involved in p75^{NTR}-mediated cell death under A β treatment and that the activity of p53 requires activation of p38 and JNK.

NF- κ B Translocation Depends on p38 and JNK Activation

Because activation of p38 and JNK, as well as translocation of NF- κ B, are involved in p75^{NTR}-dependent cell death induced by A β peptide, we decided to analyze whether they are part of the same signaling pathway or they represent two distinct pathways leading to apoptosis. For this purpose,

Fig. 2. p75^{NTR} death domain is essential for A β -induced p38 and JNK activation, as well as for NF- κ B nuclear translocation. (A) BEp75, BENTR-free, and BEp75 Δ DD cells were treated with 20 μ M A β for 6 h (left panel) and 3 h (right panel) for the detection of the activation of p38 and JNK, respectively. Cell extracts were prepared, and activation of p38 and JNK was visualized by Western blot, as described in Fig. 1. The degree of activation resulting from densitometric analysis is shown at the bottom of both panels as mean values \pm S.D. of at least three experiments. (B) BEp75, BENTR-free, and BEp75 Δ DD cells were treated with 20 μ M A β for 6 h, and translocation of NF- κ B was analyzed by EMSA. An autoradiograph of a typical experiment representative of four is shown (see Fig. 1 for legend).





NF- κ B translocation under A β treatment was analyzed in BEp75 cells stably transfected with dominant-negative mutants of MKK3, MKK4, and MKK6, which cannot activate p38 or JNK (Fig. 3B). NF- κ B translocation was blocked by the lack of activation of p38 and JNK (Fig. 6), indicating that the translocation of NF- κ B is dependent on the function of p38 and JNK.

Discussion

Previous reports have shown that p75^{NTR} is involved in neuronal death induced by fibrillar A β (Yaar et al., 1997; Kuner et al., 1998). Our group has demonstrated that the DD of the intracellular region of this receptor plays an essential role in transducing the death signal (Perini et al., 2002). These results have been confirmed further in neuronal hybrid cells transfected with either full-length or a mutated form (L401K) of p75^{NTR} (Tsukamoto et al., 2003). The signaling pathway by which the binding of A β to p75^{NTR} induces neuronal death has been variously investigated, and the available data focus on two molecules: JNK (Yaar et al., 2002; Tsukamoto et al., 2003) and NF- κ B (Kuner et al., 1998).

A possible role for JNK in p75^{NTR}-induced cell death has been widely demonstrated in different models. The involvement of JNK has also been proposed in naturally occurring sympathetic neuronal death (Aloyz et al., 1998) and in neurotrophin-induced cell death of cultured oligodendrocytes (Casaccia-Bonnet et al., 1996; Yoon et al., 1998; Harrington et al., 2002) and hippocampal neurons (Friedman, 2000; Brann et al., 2002). In some of these studies, ceramide (Casaccia-Bonnet et al., 1996; Brann et al., 2002) and Rac (Harrington et al., 2002) have been hypothesized as a link between p75^{NTR}

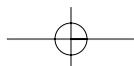
and JNK. Recently, the overexpression of the p75^{NTR} adaptor protein NRAGE was demonstrated to induce apoptosis in a JNK-dependent manner (Salehi et al., 2002).

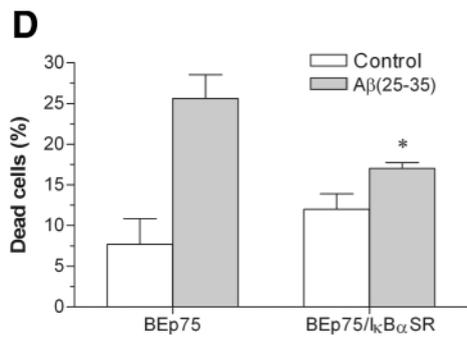
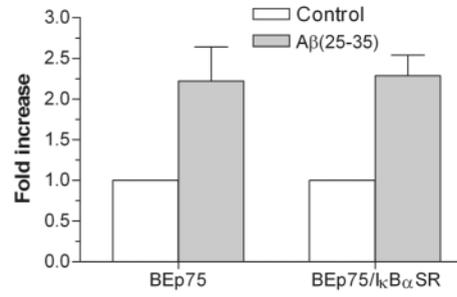
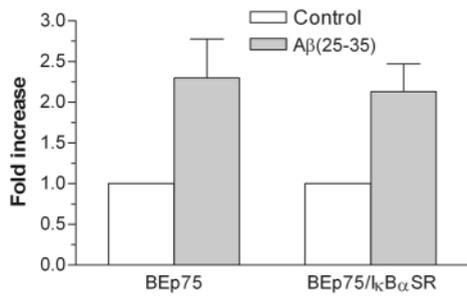
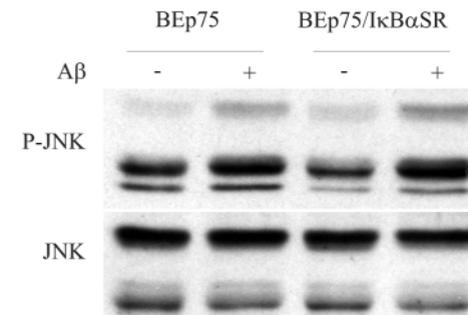
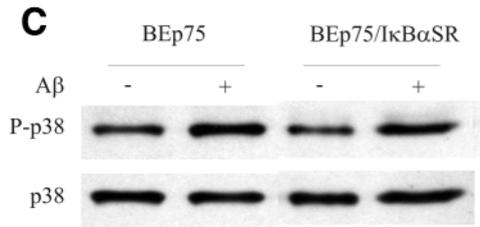
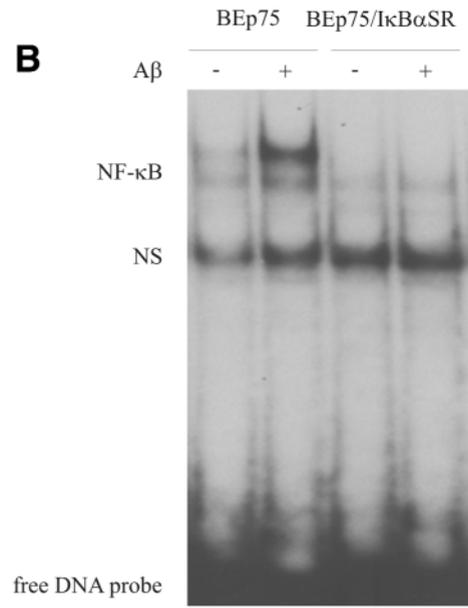
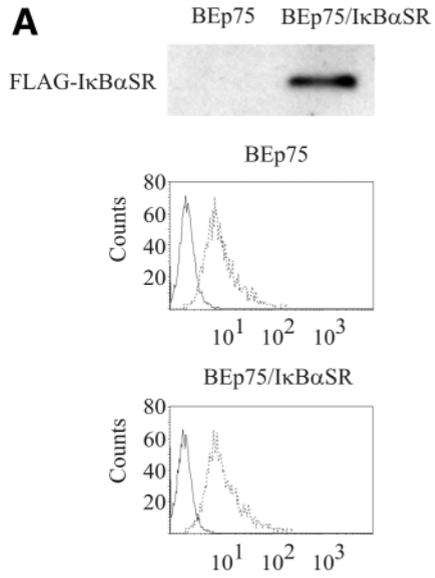
The role of NF- κ B in the p75^{NTR}-activated signaling pathway is less clear. NF- κ B activation contributes to the NGF-dependent survival of developing sensory neurons (Hamanoue et al., 1999) and a schwannoma cell line (Gentry et al., 2000); it mediates the NGF-dependent increase in the expression of the survival factor Bcl-xL (Bui et al., 2002) and induces a survival pathway in NGF-stimulated PC12 cells (Foehr et al., 2000). A possible survival function for NF- κ B has also been shown in oligodendrocytes (Ladiwala et al., 1998; Yoon et al., 1998) and Schwann cells (Carter et al., 1996; Khursigara et al., 1999; Khursigara et al., 2001). However, Kuner et al. (1998) have demonstrated that A β peptide can induce cell death in neuroblastoma cells by activating NF- κ B.

Here, we have demonstrated that fibrillar A β induces activation of p38 and JNK, as well as translocation of NF- κ B. The activation of the above kinases and NF- κ B depends on the presence of p75^{NTR}.

Mitogen-activated protein kinases (MAPKs) have been involved in the regulation of different cellular activities ranging from gene expression, mitosis, movement, metabolism, and programmed death (Johnson and Lapadat, 2002). Increasing evidence suggests that p38 and JNK play important roles in many different physiological and pathological conditions such as neuronal cell death induced by nitric oxide (Kim et al., 2002), hypoxia (Zhu et al., 2002), 2-methoxyestradiol (Shimada et al., 2003), ceramide (Willaime-Morawek et al., 2003), and in NGF-withdrawal (Aloyz et al., 1998), whereas the Erk pathway is mainly involved in the regulation of cell

Fig. 3. p38 and JNK activation is required for A β -induced p75^{NTR}-dependent cell death. **(A)** BEp75/MKK3dn, BEp75/MKK4dn, and BEp75/MKK6dn cells were lysed and subjected to Western blot analysis for the detection of HA or FLAG tags (left panel). BEp75 cells were used as a control. The position of the tagged proteins is indicated. Flow cytometric analysis of cell-surface expression of p75^{NTR} is shown in the right panel. **(B)** The left panel shows the activation of p38 after a 6-h treatment with 20 μ M A β in BEp75, BEp75/MKK3dn, and BEp75/MKK6dn cells. The right panel shows the activation of JNK after a 3-h treatment with 20 μ M A β in BEp75 and BEp75/MKK4dn cells. Cell extracts were prepared, and activation of p38 and JNK was visualized by Western blot as described in Fig. 1. The degree of activation resulting from densitometric analysis is shown at the bottom of both panels as mean values \pm S.D. of three experiments. **(C)** BEp75 and the derived cells stably expressing the empty vector pcDNA3 or each of the dominant-negative mutants of MKK3, MKK4, and MKK6 were treated with 20 μ M A β for 24 h. Cell damage was analyzed by epifluorescence microscopy, as described in Materials and Methods. The results are reported as mean values \pm S.D. of six experiments. (*) $p > 0.05$; (**) $p < 0.01$. **(D)** BEp75 cells were pretreated with vehicle or each of the two chemical compounds, SB203580 (p38 activity inhibitor) and SP600125 (JNK activity inhibitor), for 1 h at the indicated concentrations before a 24-h exposure to 20 μ M A β . Cell damage was analyzed by epifluorescence microscopy. The results are reported as mean values \pm S.D. of at least four experiments. (**) $p < 0.01$.





growth and differentiation. These different functional roles of MAPK are evident in our experimental conditions. p38 and JNK are crucial for amyloid-induced cell death mediated by p75^{NTR}. This is supported by the following findings: (1) dominant-negative forms of MKK3/6 and MKK4, which impede activation of p38 and JNK respectively, inhibit cellular damage by A β peptide; and (2) p38 and JNK inhibitors are effective in protecting cells against A β -mediated cell death. On the contrary, the Erk pathway, although activated by A β peptide, is not involved in cell death, as PD98059, a MEK inhibitor that prevents Erk1 and Erk2 activation, was ineffective in reducing amyloid-induced cell death in BEp75 cells (data not shown). These results are consistent with a previously published paper (Wei et al., 2002), showing that both JNK and Erk are activated by A β in a neuroblastoma cell line but that only JNK has a role in amyloid-induced cell death.

NF- κ B is a critical regulator of the immediate early pathogen response, playing an important role in promoting inflammation and in the regulation of cell proliferation and survival (Santoro et al., 2003). Although NF- κ B is known to protect cells from apoptosis in most cases, it is also known to contribute to apoptosis depending on cell types and extracellular stimuli (Kim et al., 2002; Shimada et al., 2003). In our experimental model, the translocation of NF- κ B is essential for amyloid-induced cell death mediated by p75^{NTR}, as demonstrated by the fact that the transfection with I κ B α super-repressor decreases cell death. This result is in agreement with previously published data (Kuner et al., 1998).

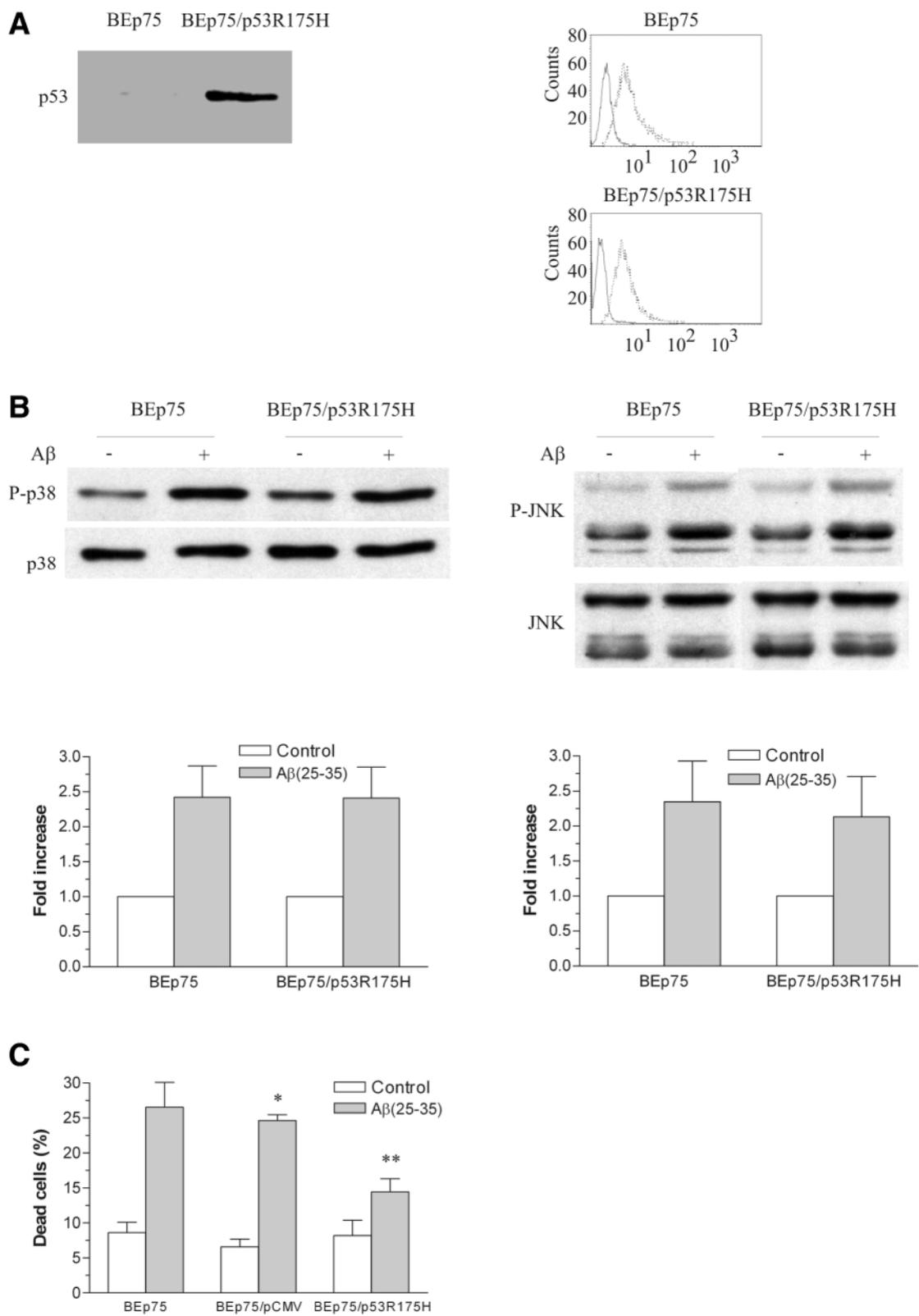
MAPKs and NF- κ B are activated through specific signaling pathways by a wide array of stimuli, including stress-inducing agents, mitogens, growth factors, and cytokines (Johnson and Lapadat, 2002; Santoro et al., 2003). In our case, both the activation of p38 and JNK, as well as the translocation of NF- κ B, require the function of the DD. The

functional relationship between p38, JNK, and NF- κ B pathways represents an intriguing issue. In some cases, p75^{NTR} can transduce a dichotomous signal with two different functions, one being proapoptotic and the other antiapoptotic (Yoon et al., 1998; Gentry et al., 2000). In our experimental model, p38/JNK activation and NF- κ B translocation have the same apoptotic function. It is thus possible that they are part of the same signaling pathway. It is known that MAPK can directly activate NF- κ B (Kim et al., 2002; Shimada et al., 2003). Our results show that NF- κ B translocation is dependent on MAPK activation; in cells where p38 and JNK activation was blocked by MKK dominant-negative mutants, NF- κ B did not translocate in the nucleus after stimulation with A β . MAPK-dependent activation of NF- κ B could function to enhance p53 transcription as in nitric oxide-induced apoptosis of chondrocytes (Kim et al., 2002) or in 2-methoxyestradiol-induced apoptosis of a human prostate cancer cell line (Shimada et al., 2003). Alternatively, NF- κ B might be induced by p53 as a required component of p53-mediated cell death, as demonstrated by Ryan et al. (2000). Finally, p53 and NF- κ B might functionally cooperate in induction of cell death, as described for HIV-1 envelope glycoprotein complex-induced apoptosis of CD4-expressing cells (Perfettini et al., 2004).

We have previously demonstrated (Perini et al., 2002) that amyloid-induced cell death mediated by p75^{NTR} requires the activation of caspase-3, a known component of amyloid-induced cytotoxicity (Takuma et al., 2004). By demonstrating the involvement of p53, we now provide the possible link between p75^{NTR} activated by A β and caspase-3 activation through modulation of the levels of Bcl-2 family members and release of cytochrome c (Morrison et al., 2003).

In summary, the results presented in this paper suggest the following death-signaling pathway: A β -p75^{NTR} complex \rightarrow activation of death domain \rightarrow acti-

Fig. 4. NF- κ B translocation is required for A β -induced p75^{NTR}-dependent cell death. (A) BEp75 and BEp75/I κ B α SR cells were lysed and subjected to Western blot analysis for the detection of the mutant version of FLAG-tagged I κ B α (top panel). The position of FLAG-I κ B α is indicated. Flow cytometric analysis of cell-surface expression of p75^{NTR} is shown in the bottom panel. (B) BEp75 and BEp75/I κ B α SR were treated with 20 μ M A β for 6 h, and the translocation of NF- κ B was evaluated by EMSA. An autoradiograph of a typical experiment representative of three is shown (see Fig. 1 for legend). (C) BEp75 and BEp75/I κ B α SR cells were treated with 20 μ M A β for 6 h (left panel) and 3 h (right panel) for detection of activation of p38 and JNK, respectively. Cell lysates were analyzed by Western blot, as described in Fig. 1. The results quantified by densitometric analysis are reported as mean values \pm S.D. of four experiments. (D) BEp75 and BEp75/I κ B α SR cells were treated with 20 μ M A β for 24 h, and cell damage was evaluated by epifluorescence microscopic analysis. The results are reported as mean values \pm S.D. of four experiments. (*) $p < 0.01$.



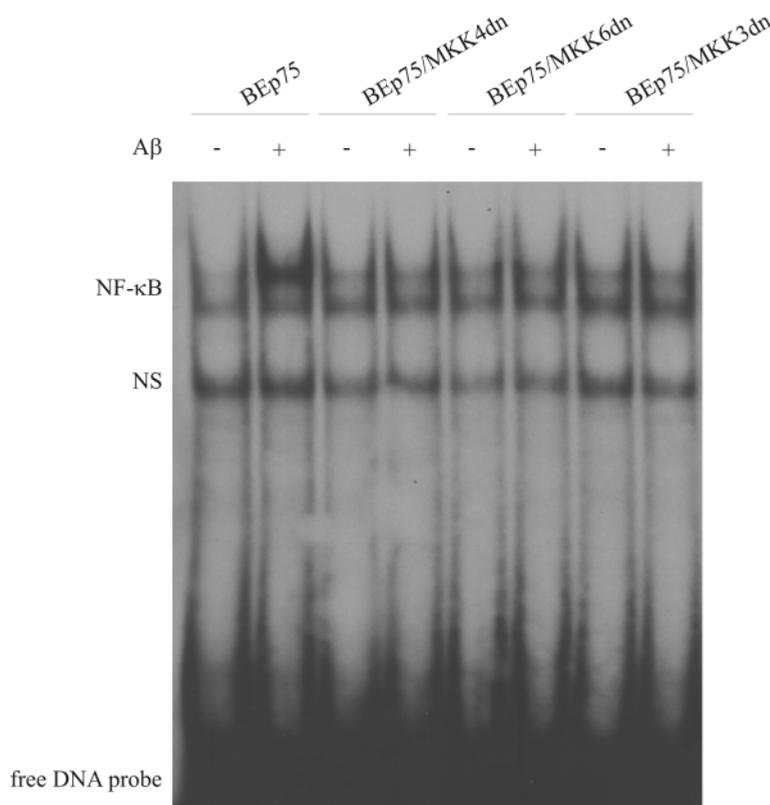


Fig. 6. NF- κ B translocation is dependent on p38 and JNK activation. BEp75, BEp75/MKK3dn, BEp75/MKK4dn, and BEp75/MKK6dn cells were treated with 20 μ M A β for 6 h. Nuclear extracts were prepared and subjected to EMSA. An autoradiograph of a typical experiment representative of three is shown (see Fig. 1 for legend).

vation of p38 and JNK \rightarrow NF- κ B translocation, and activation of p53 \rightarrow caspase-3 activation. These data have been obtained in neuroblastoma cells in which the expression of p75^{NTR} was shown to be essential for A β to exert an apoptotic effect. The involvement of p75^{NTR} was also demonstrated in previous papers in which the expression of this receptor enhanced A β toxicity (Yaar et al., 1997; Kuner et al., 1998; Tsukamoto et al., 2003). These results contrast with a recently published paper showing that p75^{NTR} has a neuroprotective role in primary cultures of human neurons against extracellular A β (Zhang et al., 2003). These opposite conclusions underline the complexity of p75^{NTR},

whose role in cell death and survival is still debated and cannot be simply reconciled based on the differences between primary cultures and tumor cell lines. The different pattern of expression of p75^{NTR} and other neurotrophin receptors in the different brain regions in physiological and pathological conditions and the different availability of adaptor proteins that transduce p75^{NTR} activation might determine the outcomes of the interaction with extracellular A β and the relative susceptibility of neurons. For this reason, further studies are required to establish the specific role of p75^{NTR} and the signaling pathway described in this paper in AD.

Fig. 5. p38 and JNK apoptotic pathway could be mediated by p53. (A) BEp75 and BEp75/p53R175H cells were lysed, and nuclear extracts were analyzed for p53 accumulation by Western blot (left panel). BEp75 and BEp75/p53R175H cells were analyzed for p75^{NTR} expression by flow cytometry (right panel). (B) BEp75 and BEp75/p53R175H cells were treated with 20 μ M A β for 6 h (left panel) and 3 h (right panel) for the detection of the activation of p38 and JNK, respectively. Cells were lysed and subjected to Western blot analysis for detection of p38 and JNK phosphorylation, as described in Fig. 1. At the bottom of both panels, densitometric analysis is reported as mean values \pm S.D. of three experiments. (C) BEp75 and BEp75/p53R175H cells were treated with 20 μ M A β for 24 h, and cell damage was evaluated by epifluorescence microscopic analysis. The results are reported as mean values \pm S.D. of five experiments. (*) $p > 0.05$; (**) $p < 0.01$.

Acknowledgments

This work was supported by grants from Fondazione Cariverona (Bando 2001 "Ambiente e Sviluppo Sostenibile"), Fondazione Cariverona (Bando 2002, "Malattie neurodegenerative: approcci innovativi allo studio dei meccanismi patogenetici, alla diagnosi e al trattamento"), and Ministero Istruzione, Università e Ricerca (COFIN 2001 and 60%). R. B. was supported by grants from Ministero Istruzione, Università e Ricerca, FIRB Projects.

References

- Aloyz R. S., Bamji S. X., Pozniak C. D., Toma J. G., Atwal J., Kaplan D. R., and Miller F. D. (1998) p53 is essential for developmental neuron death as regulated by the TrkA and p75 neurotrophin receptors. *J. Cell. Biol.* **143**, 1691–1703.
- Behrens M. M., Strasser U., Koh J. Y., Gwag B. J., and Choi D. W. (1999) Prevention of neuronal apoptosis by phorbol ester-induced activation of protein kinase C: blockade of p38 mitogen-activated protein kinase. *Neuroscience* **94**, 917–927.
- Boland K., Behrens M., Choi D., Manias K., and Perlmutter D. H. (1996) The serpin-enzyme complex receptor recognizes soluble, nontoxic amyloid-beta peptide but not aggregated, cytotoxic amyloid-beta peptide. *J. Biol. Chem.* **271**, 18,032–18,044.
- Bozyczko-Coyne D., O'Kane T. M., Wu Z. L., Dobrzanski P., Murthy S., Vaught J. L., and Scott R. W. (2001) CEP-1347/KT-7515, an inhibitor of SAPK/JNK pathway activation, promotes survival and blocks multiple events associated with Abeta-induced cortical neuron apoptosis. *J. Neurochem.* **77**, 849–863.
- Brann A. B., Tcherpakov M., Williams I. M., Futerman A. H., and Fainzilber M. (2002) Nerve growth factor-induced p75-mediated death of cultured hippocampal neurons is age-dependent and transduced through ceramide generated by neutral sphingomyelinase. *J. Biol. Chem.* **277**, 9812–9818.
- Bui N. T., König H. G., Culmsee C., Bauerbach E., Poppe M., Krieglstein J., and Prehn J. H. (2002) p75 neurotrophin receptor is required for constitutive and NGF-induced survival signalling in PC12 cells and rat hippocampal neurones. *J. Neurochem.* **81**, 594–605.
- Carter B. D., Kaltschmidt C., Kaltschmidt B., Offenhauser N., Bohm-Matthaei R., Baeuerle P. A., and Barde Y. A. (1996) Selective activation of NF-kappa B by nerve growth factor through the neurotrophin receptor p75. *Science* **272**, 542–545.
- Casaccia-Bonnet P., Carter B. D., Dobrowsky R. T., and Chao M. V. (1996) Death of oligodendrocytes mediated by the interaction of nerve growth factor with its receptor p75. *Nature* **383**, 716–719.
- Daniels W. M., Hendricks J., Salie R., and Taljaard J. J. (2001) The role of the MAP-kinase superfamily in beta-amyloid toxicity. *Metab. Brain. Dis.* **16**, 175–185.
- Dawbarn D. and Allen S. J. (2001) *Neurobiology of Alzheimer's Disease*, 2nd ed., Oxford University Press, New York.
- Dechant G. and Barde Y. A. (2002) The neurotrophin receptor p75(NTR): novel functions and implications for diseases of the nervous system. *Nat. Neurosci.* **5**, 1131–1136.
- El Khoury J., Hickman S. E., Thomas C. A., Cao L., Silverstein S. C., and Loike J. D. (1996) Scavenger receptor-mediated adhesion of microglia to beta-amyloid fibrils. *Nature* **382**, 716–719.
- Foehr E. D., Lin X., O'Mahony A., Gelezianus R., Bradshaw R. A., and Greene W. C. (2000) NF-kappa B signaling promotes both cell survival and neurite process formation in nerve growth factor-stimulated PC12 cells. *J. Neurosci.* **20**, 7556–7563.
- Fogarty M. P., Downer E. J., and Campbell V. (2003) A role for c-Jun N-terminal kinase 1 (JNK1), but not JNK2, in the beta-amyloid-mediated stabilization of protein p53 and induction of the apoptotic cascade in cultured cortical neurons. *Biochem. J.* **371**, 789–798.
- Friedman W. J. (2000) Neurotrophins induce death of hippocampal neurons via the p75 receptor. *J. Neurosci.* **20**, 6340–6346.
- Gentry J. J., Casaccia-Bonnet P., and Carter B. D. (2000) Nerve growth factor activation of nuclear factor kappaB through its p75 receptor is an anti-apoptotic signal in RN22 schwannoma cells. *J. Biol. Chem.* **275**, 7558–7565.
- Hamanoue M., Middleton G., Wyatt S., Jaffray E., Hay R. T., and Davies A. M. (1999) p75-mediated NF-kappaB activation enhances the survival response of developing sensory neurons to nerve growth factor. *Mol. Cell. Neurosci.* **14**, 28–40.
- Hardy J. and Selkoe D. J. (2002) The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science* **297**, 353–356.
- Harrington A. W., Kim J. Y., and Yoon S. O. (2002) Activation of Rac GTPase by p75 is necessary for c-jun N-terminal kinase-mediated apoptosis. *J. Neurosci.* **22**, 156–166.
- Iversen L. L., Mortishire-Smith R. J., Pollack S. J., and Shearman M. S. (1995) The toxicity in vitro of beta-amyloid protein. *Biochem. J.* **311** (Pt. 1), 1–16.
- Jang J. H. and Surh Y. J. (2002) beta-Amyloid induces oxidative DNA damage and cell death through activation of c-Jun N terminal kinase. *Ann. N. Y. Acad. Sci.* **973**, 228–236.
- Johnson G. L. and Lapadat R. (2002) Mitogen-activated protein kinase pathways mediated by ERK, JNK, and p38 protein kinases. *Science* **298**, 1911, 1912.
- Kajkowski E. M., Lo C. F., Ning X., Walker S., Sofia H. J., Wang W., et al. (2001) Beta-Amyloid peptide-induced apoptosis regulated by a novel protein containing a g protein activation module. *J. Biol. Chem.* **276**, 18,748–18,756.
- Khursigara G., Bertin J., Yano H., Moffett H., DiStefano P. S., and Chao M. V. (2001) A pro-survival function for the p75 receptor death domain mediated via the caspase recruitment domain receptor-interacting protein 2. *J. Neurosci.* **21**, 5854–5863.

- Khursigara G., Orlinick J. R., and Chao M. V. (1999) Association of the p75 neurotrophin receptor with TRAF6. *J. Biol. Chem.* **274**, 2597–2600.
- Kim S. J., Hwang S. G., Shin D. Y., Kang S. S., and Chun J. S. (2002) p38 kinase regulates nitric oxide-induced apoptosis of articular chondrocytes by accumulating p53 via NF-kappa B-dependent transcription and stabilization by serine 15 phosphorylation. *J. Biol. Chem.* **277**, 33501–33508.
- Kuner P., Schubengel R., and Hertel C. (1998) Beta-amyloid binds to p57NTR and activates NF-kappaB in human neuroblastoma cells. *J. Neurosci. Res.* **54**, 798–804.
- Ladiwala U., Lachance C., Simoneau S. J., Bhakar A., Barker P. A., and Antel J. P. (1998) p75 neurotrophin receptor expression on adult human oligodendrocytes: signaling without cell death in response to NGF. *J. Neurosci.* **18**, 1297–1304.
- Lorenzo A., Yuan M., Zhang Z., Paganetti P. A., Sturchler-Pierrat C., Staufenbiel M., et al. (2000) Amyloid beta interacts with the amyloid precursor protein: a potential toxic mechanism in Alzheimer's disease. *Nat. Neurosci.* **3**, 460–464.
- McLaurin J. and Chakrabarty A. (1996) Membrane disruption by Alzheimer beta-amyloid peptides mediated through specific binding to either phospholipids or gangliosides. Implications for neurotoxicity. *J. Biol. Chem.* **271**, 26482–26489.
- Minogue A. M., Schmid A. W., Fogarty M. P., Moore A. C., Campbell V. A., Herron C. E., and Lynch M. A. (2003) Activation of the c-Jun N-terminal kinase signaling cascade mediates the effect of amyloid-beta on long term potentiation and cell death in hippocampus: a role for interleukin-1beta? *J. Biol. Chem.* **278**, 27,971–27,980.
- Morishima Y., Gotoh Y., Zieg J., Barrett T., Takano H., Flavell R., et al. (2001) Beta-amyloid induces neuronal apoptosis via a mechanism that involves the c-Jun N-terminal kinase pathway and the induction of Fas ligand. *J. Neurosci.* **21**, 7551–7560.
- Morrison R. S., Kinoshita Y., Johnson M. D., Guo W., and Garden G. A. (2003) p53-dependent cell death signaling in neurons. *Neurochem. Res.* **28**, 15–27.
- Pei J. J., Braak E., Braak H., Grundke-Iqbal I., Iqbal K., Winblad B., and Cowburn R. F. (2001) Localization of active forms of C-jun kinase (JNK) and p38 kinase in Alzheimer's disease brains at different stages of neurofibrillary degeneration. *J. Alzheimers Dis.* **3**, 41–48.
- Perfettini J. L., Roumier T., Castedo M., Larochette N., Boya P., Raynal B., et al. (2004) NF-kappa B and p53 are the dominant apoptosis-inducing transcription factors elicited by the HIV-1 envelope. *J. Exp. Med.* **199**, 629–640.
- Perini G., Della-Bianca V., Politi V., Della Valle G., Dal-Pra I., et al. (2002) Role of p75 neurotrophin receptor in the neurotoxicity by beta-amyloid peptides and synergistic effect of inflammatory cytokines. *J. Exp. Med.* **195**, 907–918.
- Rabizadeh S. and Bredesen D. E. (2003) Ten years on: mediation of cell death by the common neurotrophin receptor p75(NTR). *Cytokine Growth Factor Rev.* **14**, 225–239.
- Ryan K. M., Ernst M. K., Rice N. R., and Vousden K. H. (2000) Role of NF-kappaB in p53-mediated programmed cell death. *Nature* **404**, 892–897.
- Salehi A. H., Xanthoudakis S., and Barker P. A. (2002) NRAGE, a p75 neurotrophin receptor-interacting protein, induces caspase activation and cell death through a JNK-dependent mitochondrial pathway. *J. Biol. Chem.* **277**, 48043–48050.
- Santoro M. G., Rossi A., and Amici C. (2003) NF-kappaB and virus infection: who controls whom. *EMBO J.* **22**, 2552–2560.
- Savage M. J., Lin Y. G., Ciallella J. R., Flood D. G., and Scott R. W. (2002) Activation of c-Jun N-terminal kinase and p38 in an Alzheimer's disease model is associated with amyloid deposition. *J. Neurosci.* **22**, 3376–3385.
- Shimada K., Nakamura M., Ishida E., Kishi M., and Konishi N. (2003) Roles of p38- and c-jun NH2-terminal kinase-mediated pathways in 2-methoxyestradiol-induced p53 induction and apoptosis. *Carcinogenesis* **24**, 1067–1075.
- Takuma H., Tomiyama T., Kuida K., and Mori H. (2004) Amyloid beta peptide-induced cerebral neuronal loss is mediated by caspase-3 in vivo. *J. Neuropathol. Exp. Neurol.* **63**, 255–261.
- Tamagno E., Robino G., Obbili A., Bardini P., Aragno M., Parola M., and Danni O. (2003a) H2O2 and 4-hydroxynonenal mediate amyloid beta-induced neuronal apoptosis by activating JNKs and p38MAPK. *Exp. Neurol.* **180**, 144–155.
- Tamagno E., Parola M., Guglielmotto M., Santoro G., Bardini P., Marra L., et al. (2003b) Multiple signaling events in amyloid beta-induced, oxidative stress-dependent neuronal apoptosis. *Free Radic. Biol. Med.* **35**, 45–58.
- Tiffany H. L., Lavigne M. C., Cui Y. H., Wang J. M., Leto T. L., Gao J. L., and Murphy P. M. (2001) Amyloid-beta induces chemotaxis and oxidant stress by acting at formylpeptide receptor 2, a G protein-coupled receptor expressed in phagocytes and brain. *J. Biol. Chem.* **276**, 23,645–23,652.
- Troy C. M., Rabacchi S. A., Xu Z., Maroney A. C., Connors T. J., Shelanski M. L., and Greene L. A. (2001) Beta-Amyloid-induced neuronal apoptosis requires c-Jun N-terminal kinase activation. *J. Neurochem.* **77**, 157–164.
- Tsukamoto E., Hashimoto Y., Kanekura K., Niikura T., Aiso S., and Nishimoto I. (2003) Characterization of the toxic mechanism triggered by Alzheimer's amyloid-beta peptides via p75 neurotrophin receptor in neuronal hybrid cells. *J. Neurosci. Res.* **73**, 627–636.
- Wang H. Y., Lee D. H., D'Andrea M. R., Peterson P. A., Shank R. P., and Reitz A. B. (2000) Beta-Amyloid(1–42) binds to alpha7 nicotinic acetylcholine receptor with high affinity. Implications for Alzheimer's disease pathology. *J. Biol. Chem.* **275**, 5626–5632.
- Wei W., Wang X., and Kusiak J. W. (2002) Signaling events in amyloid beta-peptide-induced neuronal death and insulin-like growth factor I protection. *J. Biol. Chem.* **277**, 17649–17656.
- Willaime-Morawek S., Brami-Cherrier K., Mariani J., Caboche J., and Brugg B. (2003) C-Jun N-terminal

- kinases/c-Jun and p38 pathways cooperate in ceramide-induced neuronal apoptosis. *Neuroscience* **119**, 387–397.
- Yaar M., Zhai S., Fine R. E., Eisenhauer P. B., Arble B. L., Stewart K. B., and Gilchrest B. A. (2002) Amyloid beta binds trimers as well as monomers of the 75-kDa neurotrophin receptor and activates receptor signaling. *J. Biol. Chem.* **277**, 7720–7725.
- Yaar M., Zhai S., Pilch P. F., Doyle S. M., Eisenhauer P. B., Fine R. E., and Gilchrest B. A. (1997) Binding of beta-amyloid to the p75 neurotrophin receptor induces apoptosis. A possible mechanism for Alzheimer's disease. *J. Clin. Invest.* **100**, 2333–2340.
- Yan S. D., Chen X., Fu J., Chen M., Zhu H., Roher A., et al. (1996) RAGE and amyloid-beta peptide neurotoxicity in Alzheimer's disease. *Nature* **382**, 685–691.
- Yoon S. O., Casaccia-Bonnel P., Carter B., and Chao M. V. (1998) Competitive signaling between TrkA and p75 nerve growth factor receptors determines cell survival. *J. Neurosci.* **18**, 3273–3281.
- Zhang Y., Hong Y., Bounhar Y., Blacker M., Roucou X., Tounekti O., et al. (2003) p75 neurotrophin receptor protects primary cultures of human neurons against extracellular amyloid beta peptide cytotoxicity. *J. Neurosci.* **23**, 7385–7394.
- Zhu Y., Mao X. O., Sun Y., Xia Z., and Greenberg D. A. (2002) p38 Mitogen-activated protein kinase mediates hypoxic regulation of Mdm2 and p53 in neurons. *J. Biol. Chem.* **277**, 22,909–22,914.
- Zhu X., Ogawa O., Wang Y., Perry G., and Smith M. A. (2003) JKK1, an upstream activator of JNK/SAPK, is activated in Alzheimer's disease. *J. Neurochem.* **85**, 87–93.



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 thus 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 per se a clear proof of their role in drug resistance, 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, so as to bring to the surface low-abundance proteins that might be responsible for the onset of drug resistance).

Expert Rev. Proteomics 2(2), xxx–xxx (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 this disease.

Tumor markers presently used 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 as 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, 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 N-methyl transferase	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 gradients.

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 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 much 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 many 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, which 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]. The 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 in fact 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, while engaging different intracellular targets, raises the possibility that

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	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]).

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 the kinase 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 2A (PP2A), 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

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	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	Threefold decreased in A431 treated
Stathmin	17,171 – 6	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]).

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	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]).

complex. The 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 the 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 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 the pH 2.8–5.0, the other in the 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 already significant and accepted for further analysis).

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 isoforms 14-3-3 γ 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, especially 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 ¹⁸O (resistant cells) or with ¹⁶O (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 ¹⁸O/¹⁶O 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 has 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)
- Ca²⁺-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)

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 parenthesis refer to the table in which the various proteins are listed.
HSP: Heat shock protein.

- Ion channels (e.g., voltage-dependent anion-selective channel [VDAC]-1)
- Cytoskeletal proteins and their regulators (e.g., stathmin and EB1)

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 (e.g., 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, we suspect 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 a 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 on the 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 in 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-ordinate exchange of information and collaborative efforts among several research groups will be operative. Additionally, biochemists, who are the ones 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 classical 2D map analysis will have 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 have also 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 has 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 bedside 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, Pawletz 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, Misk 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, 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 Biotech.* 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. *Patol. 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).
- *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
- *Nataschia Camprostrini, 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
camprostrini@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

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

RINGRAZIAMENTI

Al Prof. Pier Giorgio Righetti va il mio sincero ringraziamento per avermi condotto con competenza, comprensione e fiducia attraverso l'entusiasmante percorso dello svolgimento della presente Tesi di dottorato.

Desidero ringraziare il Dott. Enrico Domenici, la Dott.ssa Lucia Carboni e la Dott.ssa Silvia Mion (del laboratorio di Molecular Medicine del Centro Ricerche GlaxoSmithKline di Verona) per la collaborazione, i consigli, le discussioni e la passione contagiosa, che sono stati per me un preziosissimo aiuto per questo progetto di dottorato.

Questa Tesi si è avvalsa inoltre della prezioso contributo di numerose persone, alle quali invio il più sentito ringraziamento per il loro supporto: il Prof. Aldo Scarpa, la Prof.ssa Marta Palmieri ed il Dott. Massimo Donadelli (dei Dipartimenti di Patologia e di Scienze Neurologiche e della Visione dell'Università di Verona); il Prof. Emilio Marengo e la Dott.ssa Elisa Robotti (del Dipartimento di Scienze e Tecnologie Avanzate dell'Università del Piemonte Orientale); il Dott. Mahmoud Hamdan ed il Dott. Hubert Astner (del Laboratorio di Mass Spectrometry del Centro Ricerche GlaxoSmithKline di Verona).

Ringrazio inoltre tutte "le Righette" e tutti i componenti (presenti e passati) del Laboratorio, per essere stati, prima che colleghi, dei veri amici, per avere condiviso con me momenti belli e meno belli, ed avermi accompagnato con affetto nel corso di questi anni.

Desidero ringraziare anche tutti gli amici di sempre, per il loro sostegno e la loro impareggiabile amicizia.

Per finire, un ringraziamento di tutto cuore per la fiducia, il sostegno e l'interesse che ho sempre sentito attorno a me, a tutta la mia famiglia, in particolare a mia sorella Sabrina ed ai miei genitori, ai quali dedico questo mio lavoro.

Infine, *last but not the least*, a me stessa, per una volta...