Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography A, 1051 (2004) 3–17
JOURNAL OF CHROMATOGRAPHY A
www.elsevier.com/locate/chroma

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 $[^2H_0]/[^2H_3]$-acrylamide; the other one, also based on a Cys reactive compound, exploits $[^2H_0]/[^2H_4]$ 2-vinylpyridine. The latter reagent achieves 100% efficiency coupled to 100% specificity. The advantages and limitations of the various protocols are discussed.

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

Contents

* Corresponding author. Tel.: +39 045 8027901; fax: +39 045 8027929.
E-mail address: righetti@sci.univr.it (P.G. Righetti).

0021-9673/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2004.05.106

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, if 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 there from, 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 Storhman [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 electrokinetic 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,

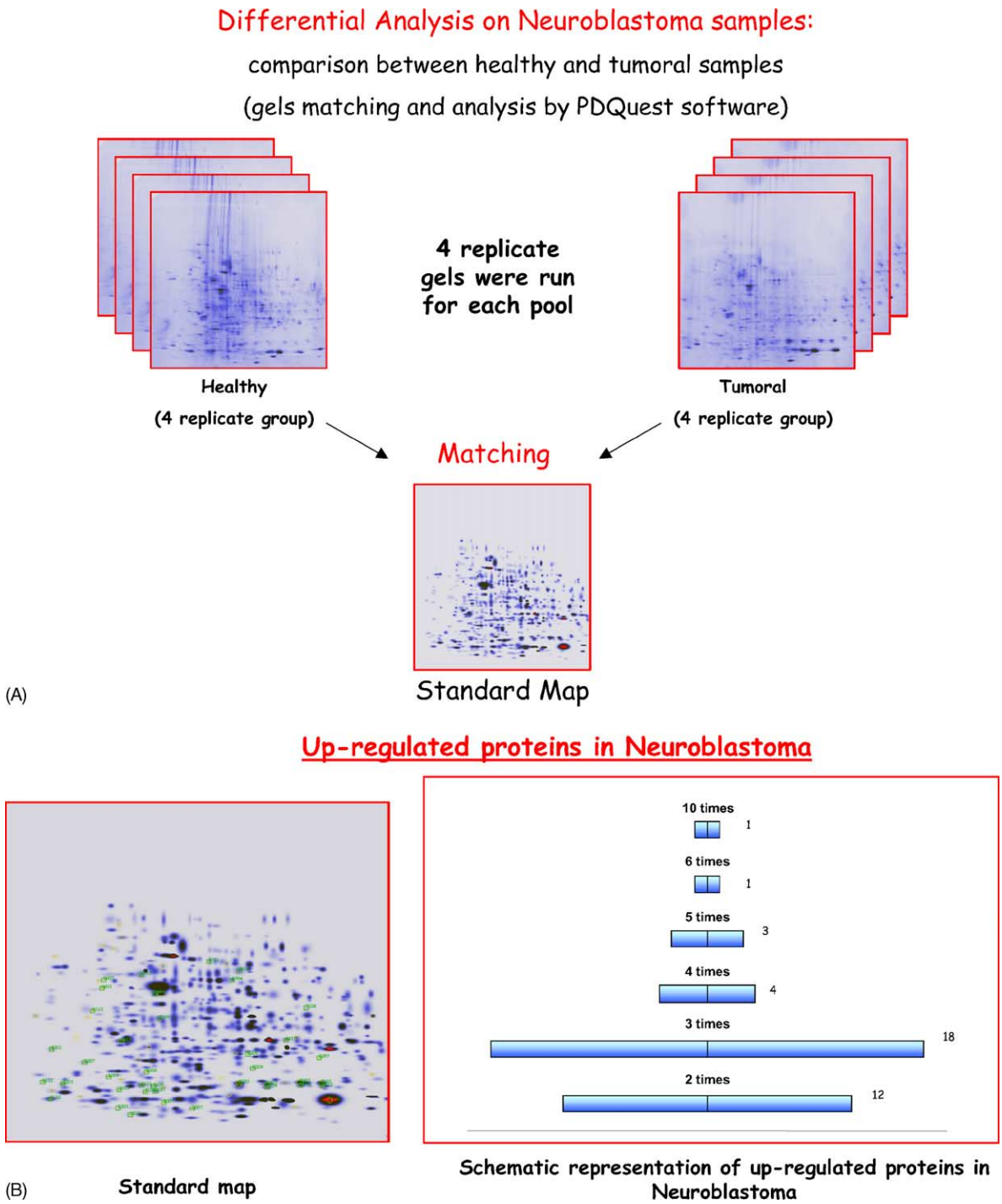
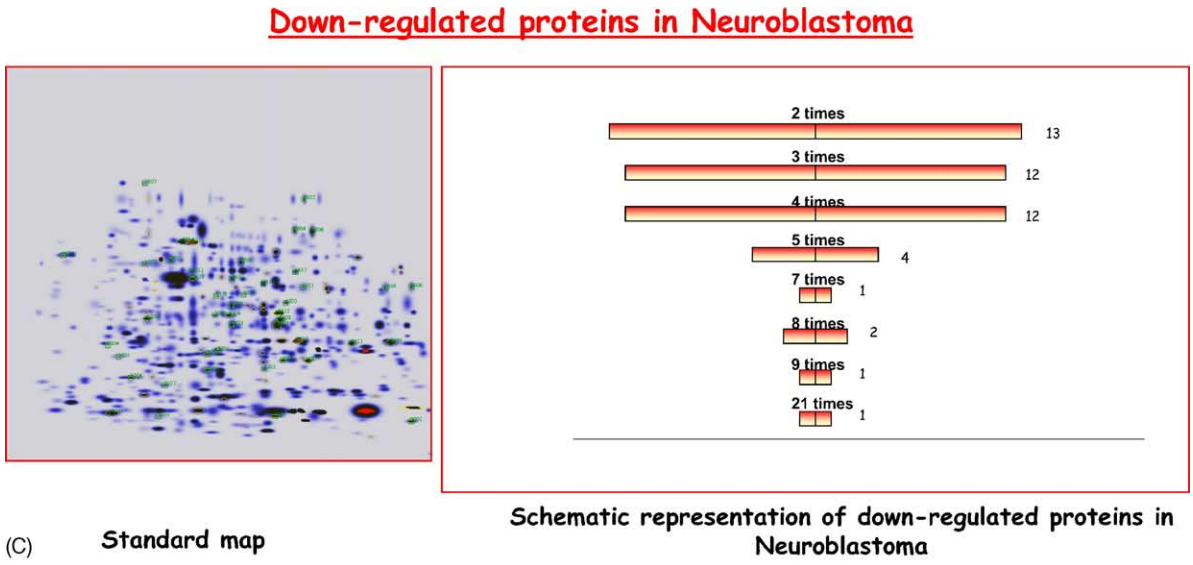


Fig. 1. Experimental design for generating sets of 2D maps from control and treated samples and for comparing master maps, stained with colloidal Coomassie Blue, for detection of up- and down-regulated proteins in the paired samples. All comparative steps performed with the PDQuest software. (A) Creating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



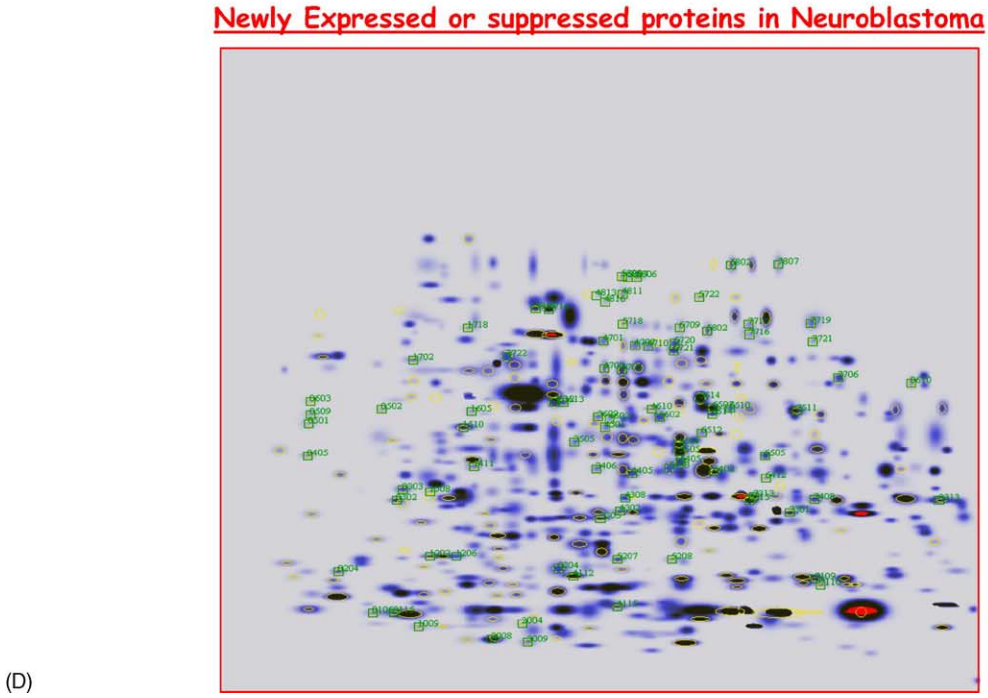


Fig. 1. (*Continued*).

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, NCBInr, 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 different 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.

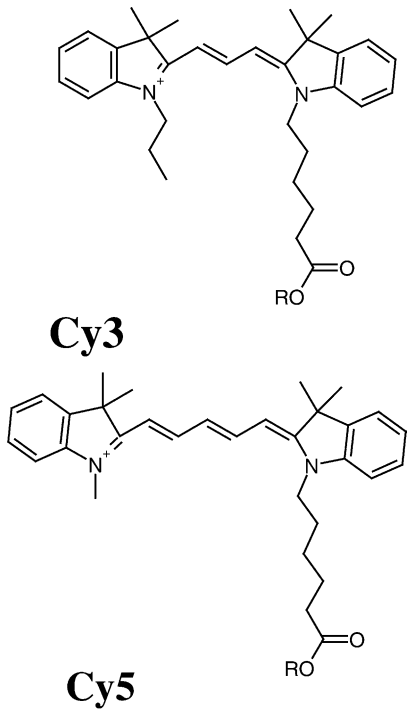


Fig. 2. Chemical formulae of the Cy3 and Cy5 dyes.

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

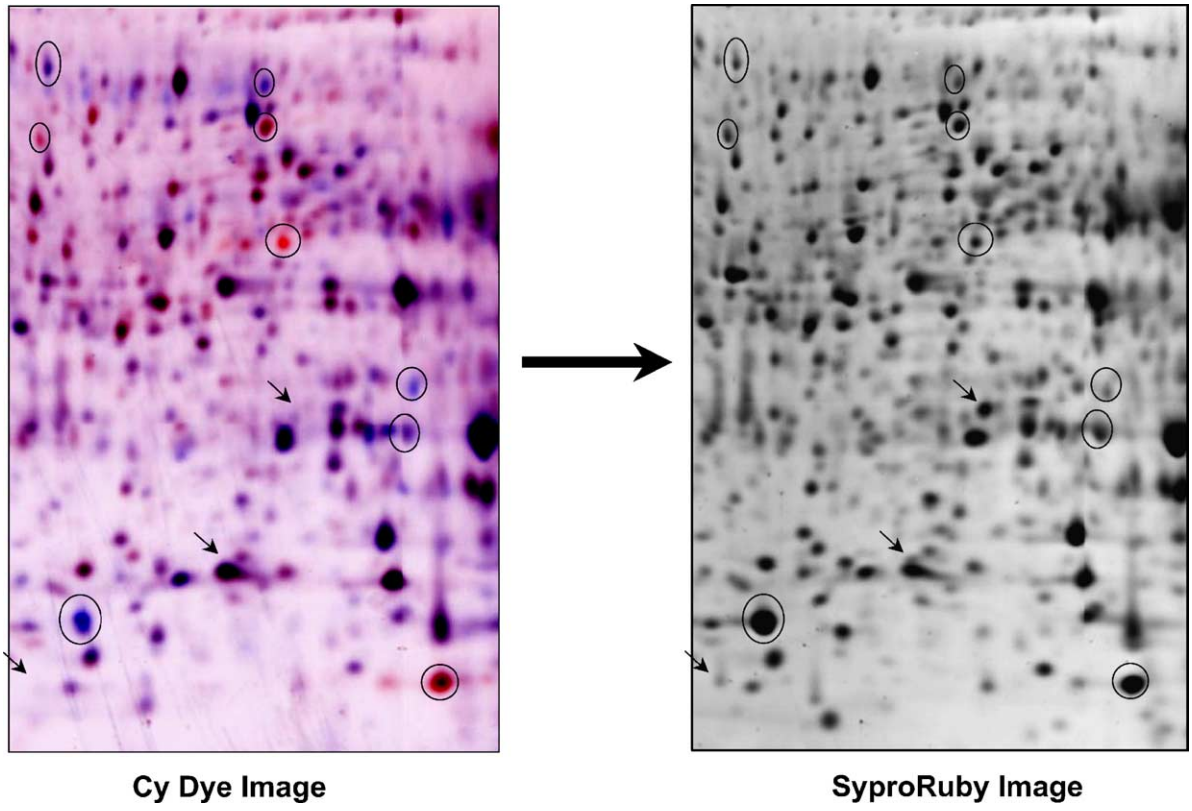


Fig. 3. Comparison of 2D-DIGE imaging and Sypro Ruby post-staining. Left panel: merged Cy dye image of HB4a lysate labelled with Cy3 (red) and HBc3.6 lysate labelled with Cy5 (blue). The same gel was post-stained with Sypro Ruby (right panel). Circles represent differentially expressed proteins detectable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

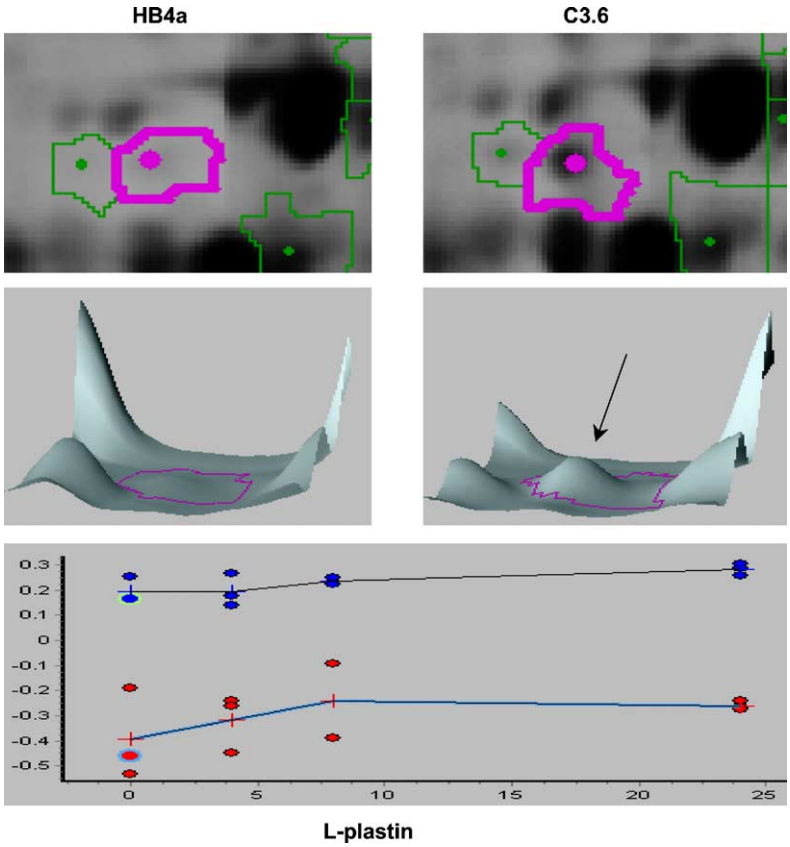


Fig. 4. Detailed gel area on a Cy3/Cy5 labelled sample pool, showing the induction of L-plastin in tumour cells. Bottom panel: kinetics of L-plastin induction. From [22] with permission.

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

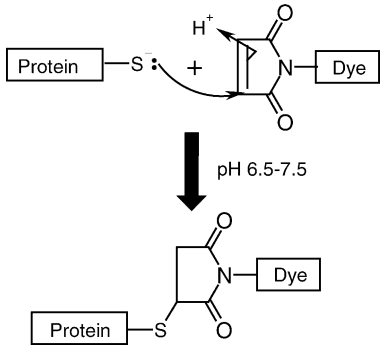


Fig. 5. Reaction scheme of maleimido cyanine dyes with the –SH group of proteins. From [26] with permission.

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 μg 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-

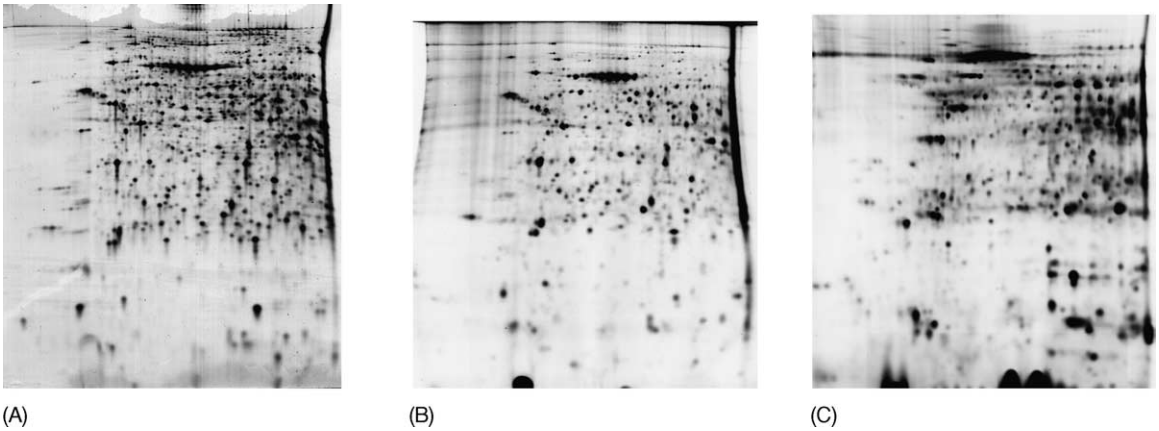


Fig. 6. 2D gel images of (A) silver stained, (B) Cy3 minimal dye-labelled and (C) Cy3 saturation dye-labelled liver homogenate (50 μg, pH 4–7 18 cm IPG strip). Whilst the silver stained image and the Cy3 minimal dye image are very similar, the Cy3 saturation dye image shows an altered spot pattern. From [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 p*I* value, due to the well-known fact that the p*I* of a protein is a point on the pH scale of minimum of solubility.

5. Isotope-coded two-dimensional maps: $[^2H_0]/[^2H_3]$ 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

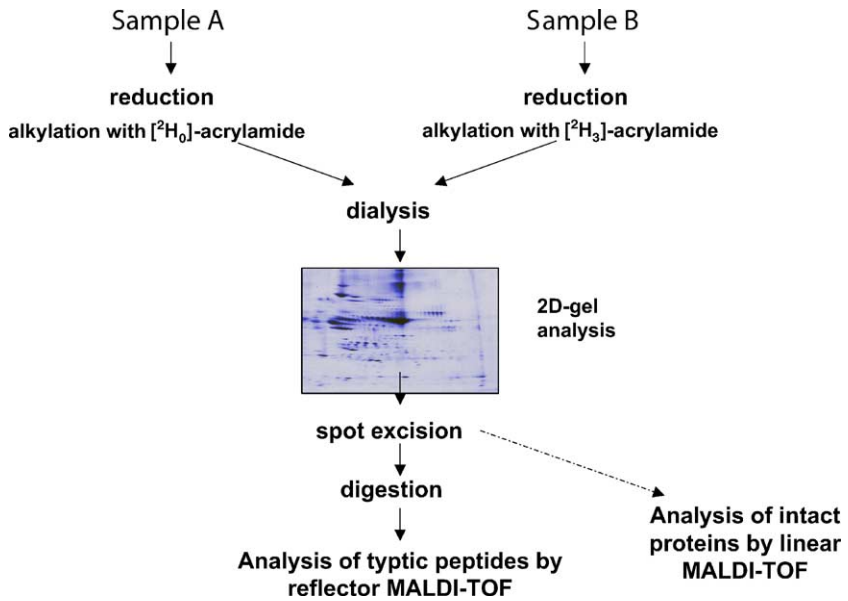


Fig. 7. Scheme for differential labelling of two samples with $[^2H_0]/[^2H_3]$-aclamide (alkylation of Cys residues). The central map refers to rat sera, labelled separately with either $[^2H_0]$- or $[^2H_3]$-acrylamide and mixed in a 30:70% ratio. From [34] with permission.

"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-

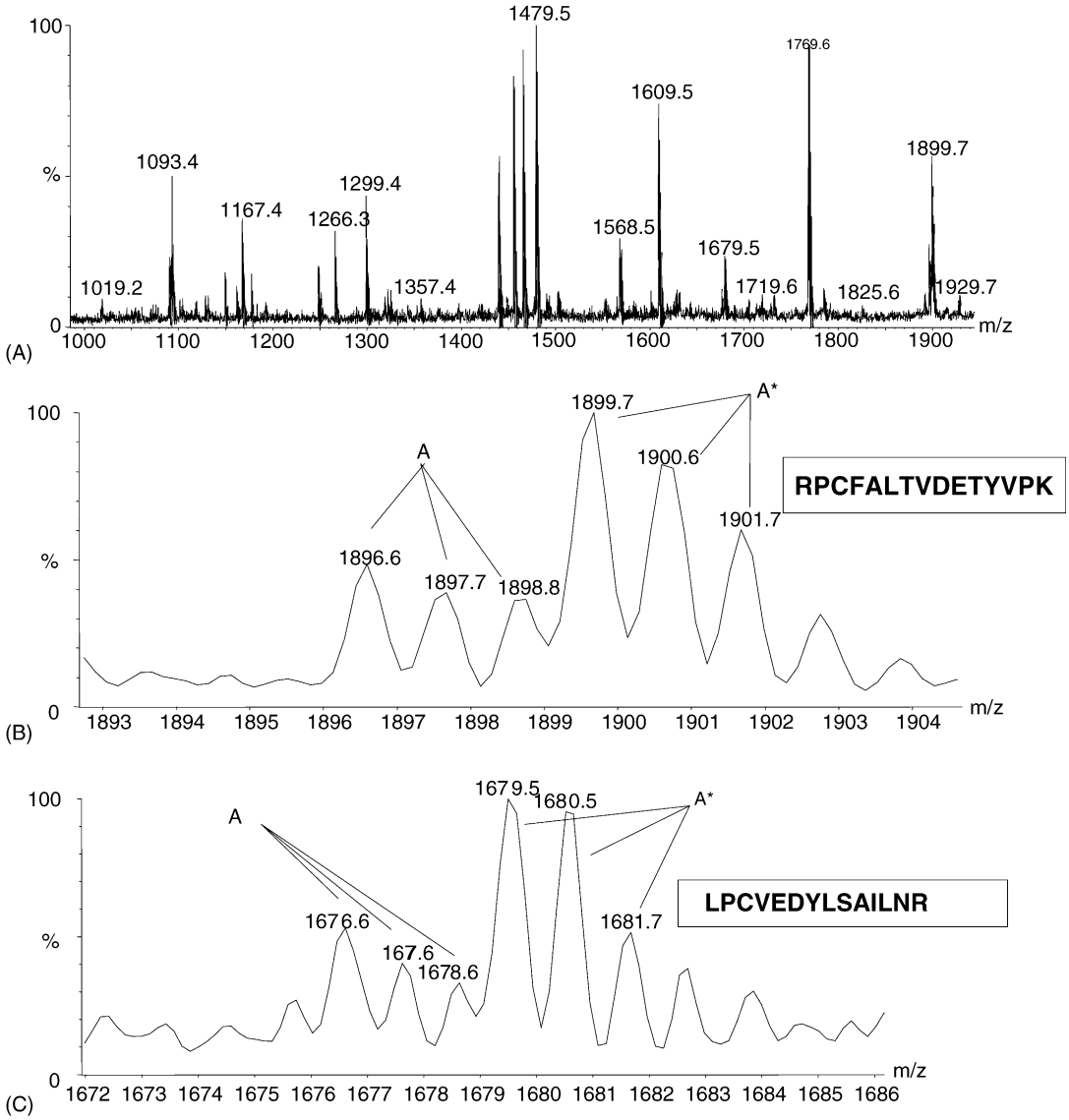


Fig. 8. (A) Reflector MALDI mass spectrum of an in-situ digest of apo-transferrin taken from the 2D map of rat sera displayed in Fig. 7, that were alkylated with [$^2H_0$]- and [$^2H_3$]-acrylamide and mixed in a 30%/70% ratio. (B) and (C) are two short intervals taken from (A), and are associated with the two indicated peptide sequences. From [34] with permission.

gies. An example of such an approach could be the use of $[^2H_0]/[^2H_3]$-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 identification 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 $[^2H_0]$-acrylamide, and the second with its $[^2H_3]$ 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,

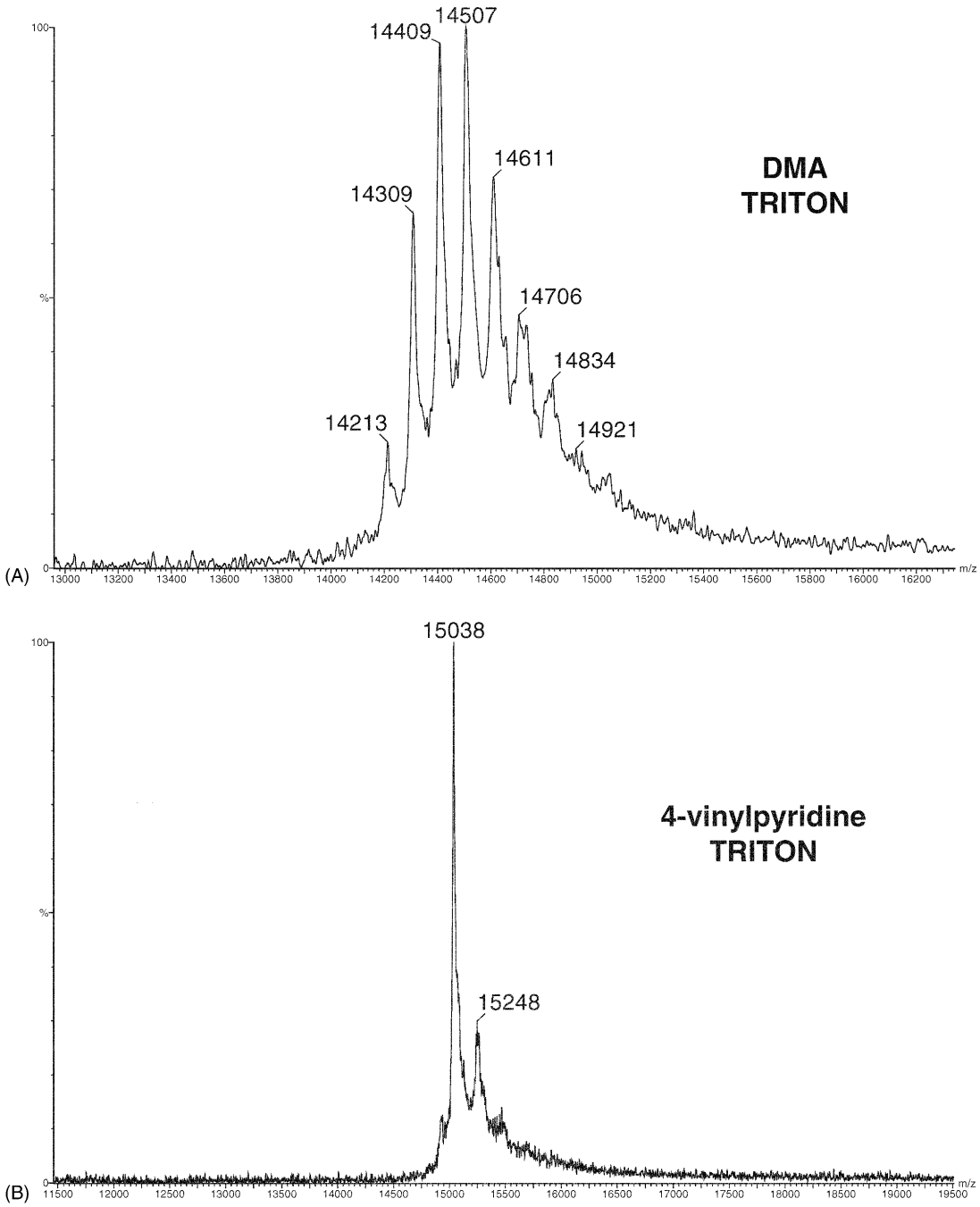


Fig. 9. MALDI-TOF mass spectra of bovine α-lactalbumin after 1 h incubation with DMA (A) or 4-vinylpyridine (B), both in presence of the surfactant 2% Triton X-100. Note that, in panel B, the peak at *m/z* 15 248 represents an adduct of LCA with the MALDI matrix, sinapinic acid. From [36] with permission.

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 [$^2H_0$]-acrylamide, and the second (70%) reacted with [$^2H_3$]-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 [$^2H_0$] or [$^2H_3$]-acrylamide effectively co-migrate in the IEF dimension (i.e., there is no isotope effect shifting the p*K*, thus the p*I* 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 iodacetamide), 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).

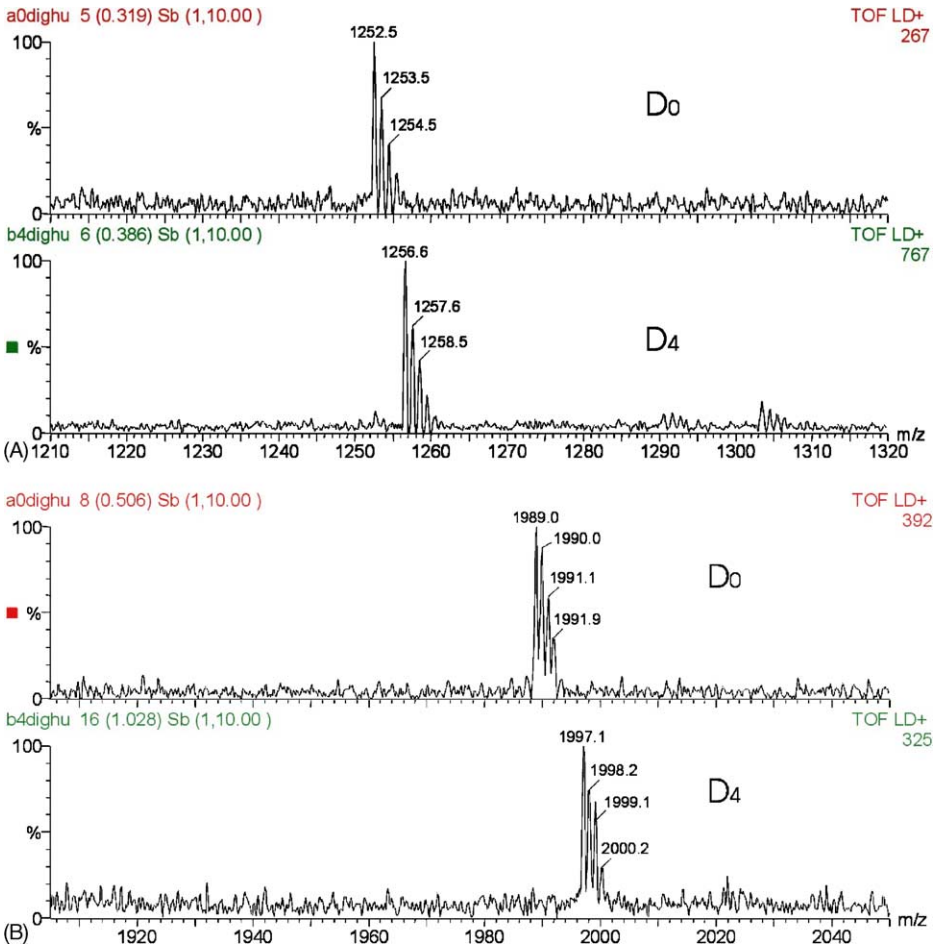


Fig. 10. (A) Zoom-in to monoisotopic distributions of the *m/z* 1252.5 peptide, obtained by tryptic digestion of α-lactalbumin, labelled with [$^2H_0$] 2-VP (D_0) and the corresponding *m/z* 1256.6 peptide labelled with [$^2H_4$] 2-VP (D_4). The peptide contains a single Cys residue. (B) Zoom-in to monoisotopic distributions of the *m/z* 1989.0 peptide labelled with [$^2H_0$] 2-VP and the corresponding *m/z* 1997.1 peptide labelled with [$^2H_4$] 2-VP. Note that this peptide contains 2 Cys residues. From [36] with permission.

6. Isotope-coded two-dimensional maps: $[^2H_0]/[^2H_4]$ 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 synthesized 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 $[^2H_0]/[^2H_4]$-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, were protein spots identification via MS appears to be problematic. Perhaps a good compromise might be alkylation with the stable isotope approach, namely with $[^2H_0]/[^2H_4]$ 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 (were 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. Garrels, 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, Protoemics 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. Schrattenholz, 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.