

## Review of some recent approaches in quantitative proteomics

M. Hamdan<sup>1</sup>, P.G. Righetti<sup>2</sup>

<sup>1</sup>Computational, Analytical & Structural Sciences, GlaxoSmithKline, Verona, Italy

<sup>2</sup>Department of Agricultural and Industrial Biotechnologies, University of Verona, Italy

### Abstract

An attempt is made to review a number of approaches that appeared in recent literature which address quantitative proteome analysis. These methods can be divided into two main streams: those applicable to conventional two-dimensional map analysis, coupling orthogonally a charge-based step (isoelectric focusing) to a size-based separation (SDS-electrophoresis) and those applicable to two-dimensional chromatographic protocols. The first method, although being by and large the most popular approach, can offer differential display of paired samples with relatively few methods, the oldest one being based on statistical analysis performed on sets of gels via powerful software packages. Recent developments comprise analysis performed on a single gel containing mixed samples differentially labelled, either with fluorophors (Cy3 and Cy5) or with  $d_0/d_3$  acrylamide. Conversely, chromatographic approaches, which mostly rely on analysis not of intact proteins but of their tryptic digests, offer a panoply of differential labelling protocols, most of which rely on stable isotope tagging. Essentially all possible reactions have been described, such as those involving Lys, Asp, Glu, Cys residues, as well as a number of methods exploiting differential derivatization of amine and carboxyl groups generated during proteolysis.

### 1 Introduction

The goal of proteomics is usually defined as the identification of the proteins encoded in the human (or any other) genome followed by the determination of:

1. their range of expression across the different cell types that constitute the various (human) tissues;
2. their sub-cellular localisation;

3. their post-translational modifications;
4. their interaction with other proteins,
5. their structure-function relationships.

Proteome analysis, in all of its variegated aspects, has been recently amply described in a number of books(1-7), which trace the evolution of the technique, from the well-ingrained O'Farrell two-dimensional (2-D) maps by orthogonal charge/size fractionation (IEF/IPG SDS-PAGE) in a gel slab format (8), to recent multi-dimensional chromatographic steps coupling practically all kinds of chromatographies described up to the present time, or to hybrid methodologies, combining chromatography with electrophoresis.

A key issue in all these approaches is quantitation of all spots/peaks resolved, so as to enable a description of the evolution of physiological/pathological events in terms of up- and down-regulation of any possible expressed phenotype. Such a differential display would allow to select key proteins involved in such events as potential drug targets in an ever expanding pharmaceutical market seeking development of new, powerful and properly engineered drugs. In the conventional 2-D approach, able to display several thousand spots in a single gel slab, this was done with the aid of powerful computer algorithms, developed over the years, such as the Melanie and the PD-Quest, able to create standard maps of, e.g., physiological vs. pathological cell populations, compare them via an overlapping procedure, and detect all spots exhibiting higher or lower stain intensity (typically Coomassie Blue), on the assumption that such densitometric differences would reflect differential protein expression. Although this procedure has been amply demonstrated in innumerable publications up to the present, it suffers from two shortcomings: first of all, the extremely laborious and time-consuming set-up, requiring generation of at least five maps for each state (control vs. disease); in a second instance, the fact

that it could not possibly apply to all chromatographic approaches recently described.

New alternatives to quantitative proteomics had thus to be sought, and in fact a host of methods appeared suddenly in the scientific press in just about two-three year's period. It is the aim of this review to bring them to the lime light and give figures of merits to the various approaches.

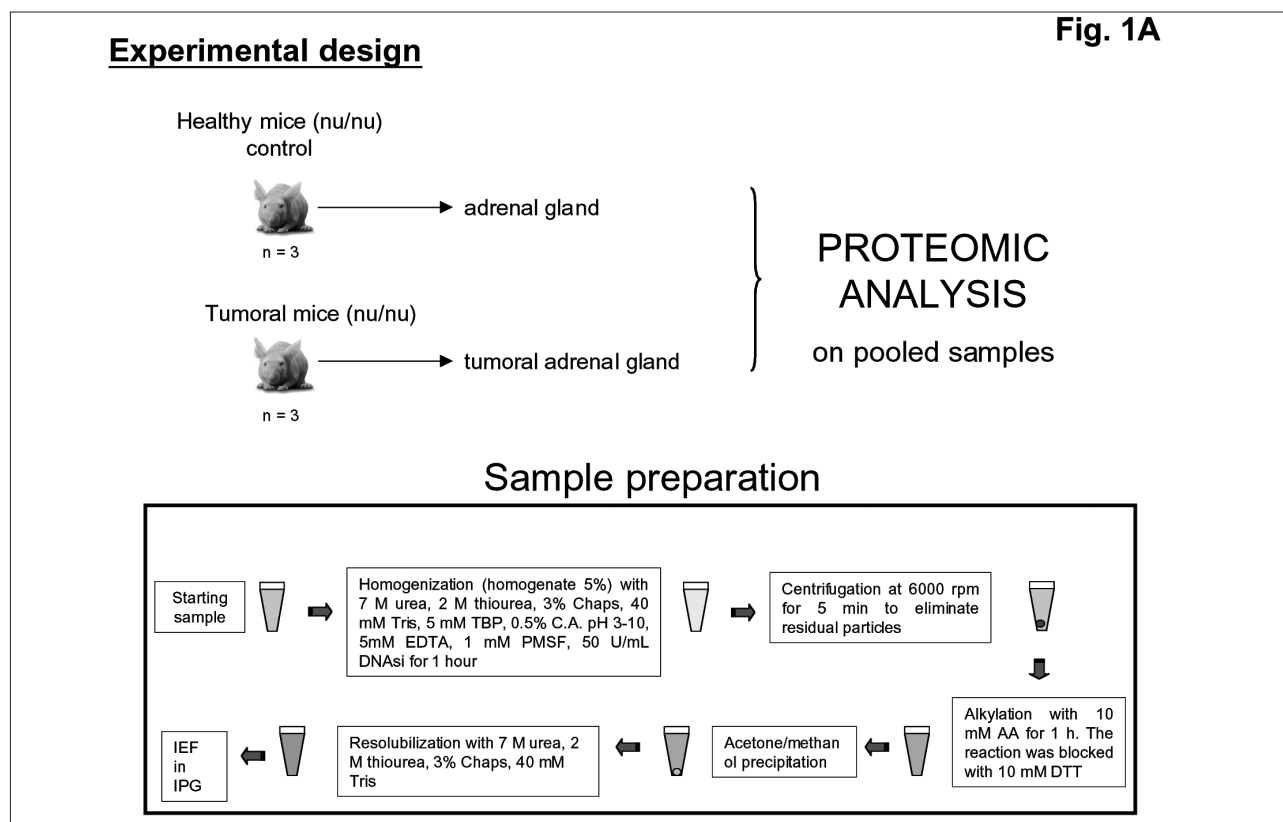
## 2 Quantitative approaches based on 2-D map analysis

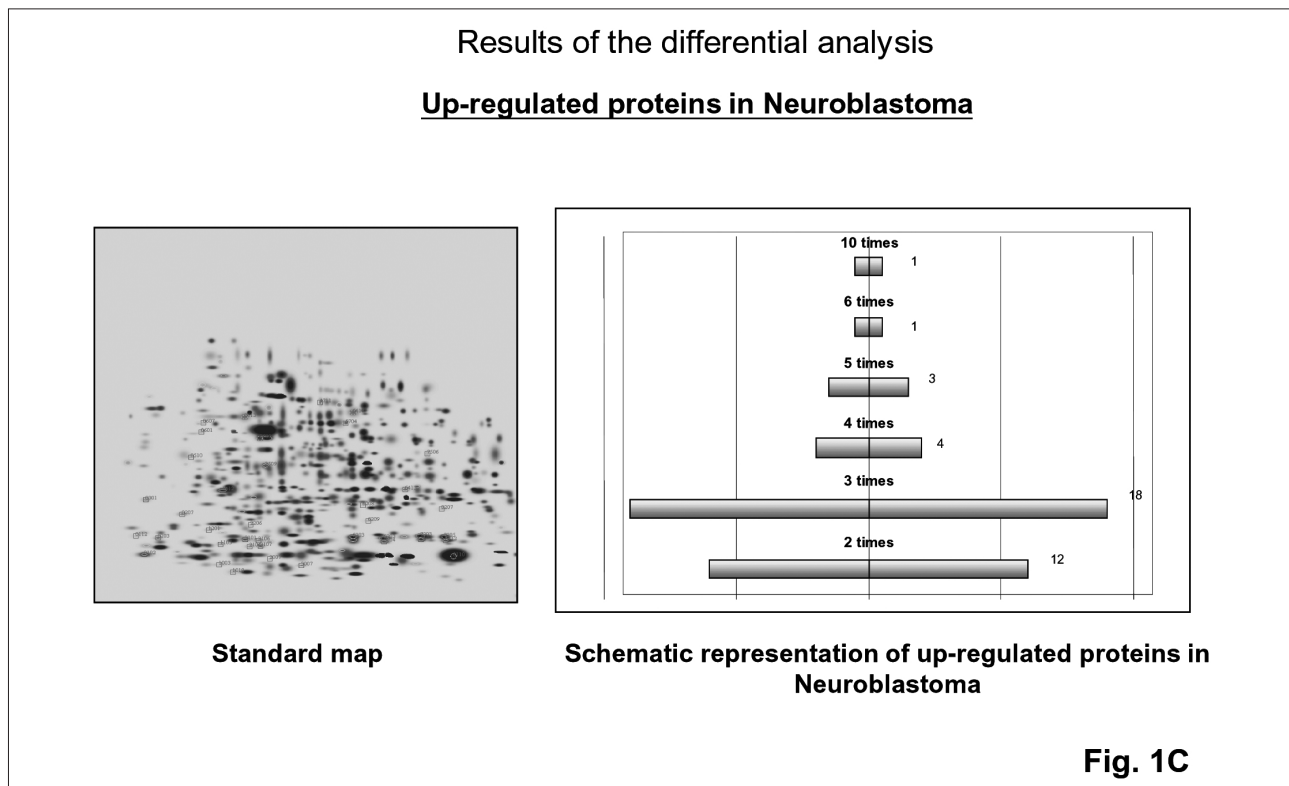
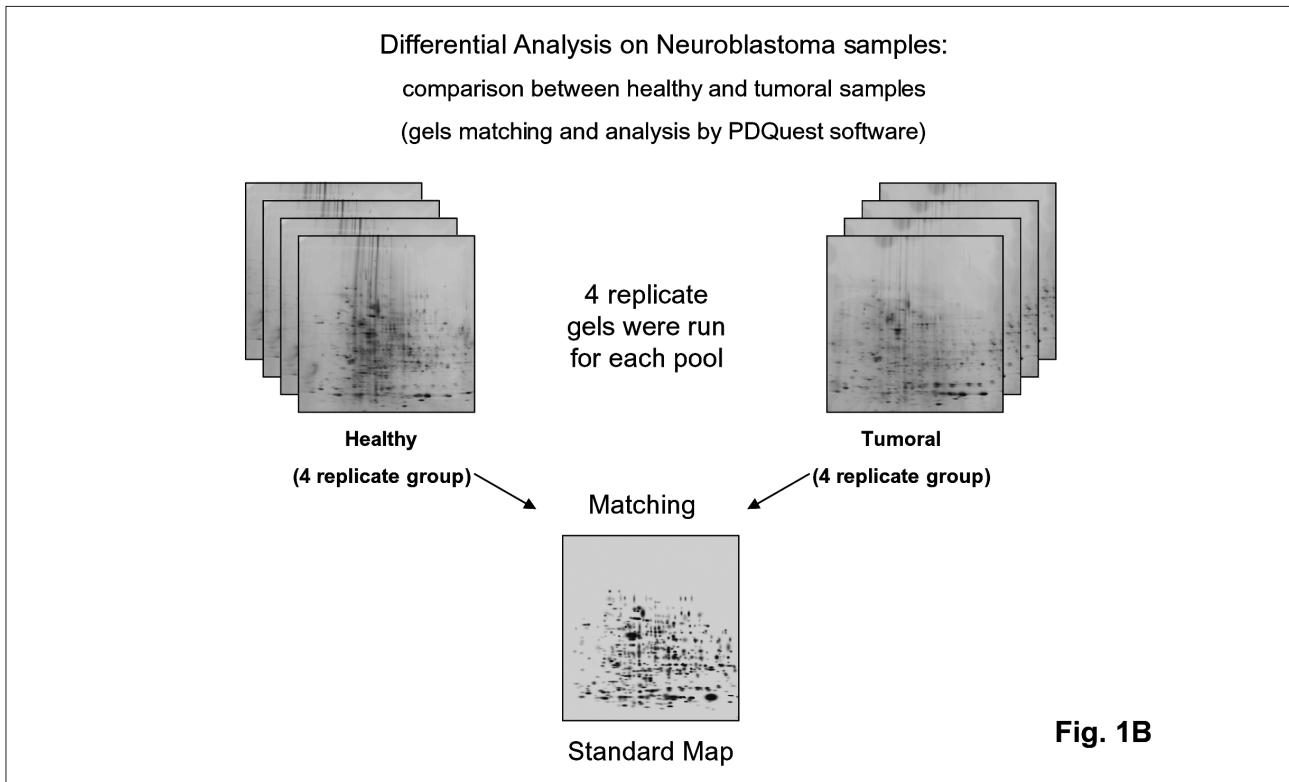
### 2.1 Statistical analysis of separately-run 2-D maps

As stated above, comparison of 2-D 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 Brahe, a famous Danish astronomer of the seventeen century) (9) is one of the oldest and most popular methods in the electrophoretic approach to proteome analysis. The sequence of panels in Figure 1 gives an example of such a procedure. It refers to neuroblastomas, a type of tumour that accounts for approximately 9% of all childhood cancers, occurring once out of 8,000 live births, as analyzed in an experimental mice model. Figure 1A shows the experimental design: for statistical significance 3 to 5 animals, for each state,

should be sacrificed; the pooled tissue samples will then be processed as illustrated below, in preparation for the first dimension of a 2-D map (in general a pH 3-10, non-linear, immobilized pH gradient). As illustrated in Figure 1B, 4-5 replicas of such 2-D maps should be run simultaneously, so as to maximize spot reproducibility (in general, we prefer fairly large-size 2-D maps, 18 cm in the focusing dimension, 20 cm in the SDS dimension, although even larger sizes, e.g. 24 x 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. Figure 1C 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). Figure 1D gives an analogous scheme for down-regulated proteins in neuroblastomas. Such analysis can offer additional information too, as shown in Figure 1E. It can detect protein spots that are newly expressed in the tumoral

**Figure 1.** Experimental design for generating sets of 2-D 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: sample collection and preparation. B: creating master maps from replica gels. C: display of up-regulated proteins. D: display of down-regulated proteins. E: display of newly expressed or suppressed proteins in neuroblastoma.





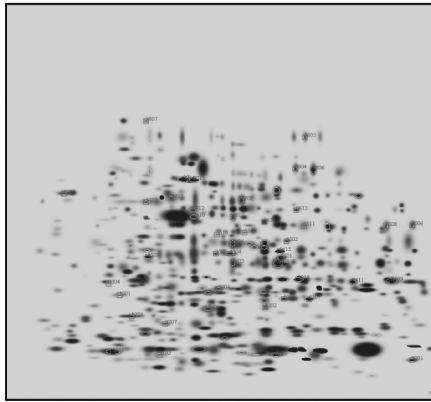
samples as regard to control ones and protein spots that are newly silenced in the tumoral samples as compared to healthy ones. These kinds of spots are highlighted in green. Once this differential analysis has been performed, all the spots of interest are excised, in gel digested and subsequently characterized by mass spectrometry (e.g., using MALDI-TOF or LC/ESI-MS). Once the precise mass of each tryptic

fragment is obtained, together with a lead sequence, interrogation of a number of databases (e.g., SwissProt, TrEMBL, NCBI nr, and the like) enables proper identification of the unknown protein, provided, of course, that it is listed in any of them.

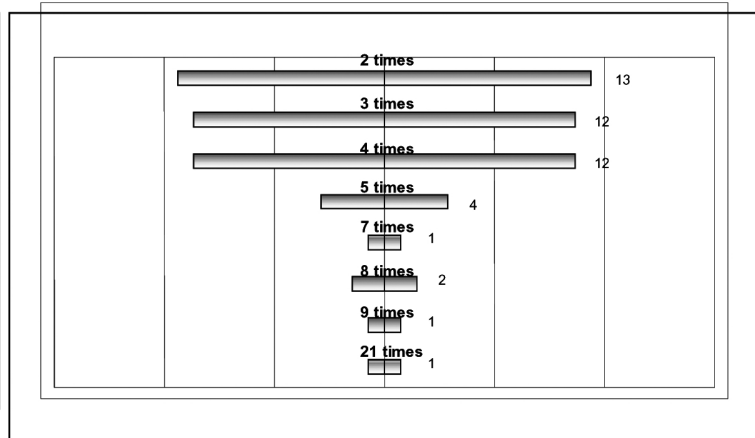
*2.2 Differential, in-gel electrophoresis*

This approach, amply used to the present, gives

### Down-regulated proteins in Neuroblastoma



Standard map



Schematic representation of down-regulated proteins in Neuroblastoma

Fig. 1D

### Newly Expressed or suppressed proteins in Neuroblastoma

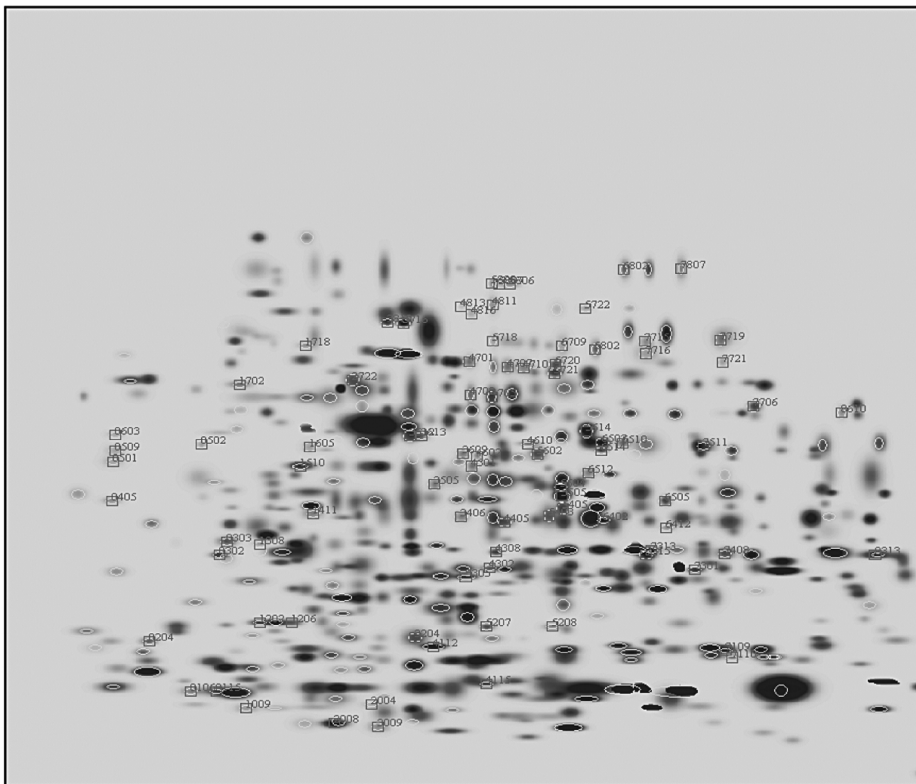
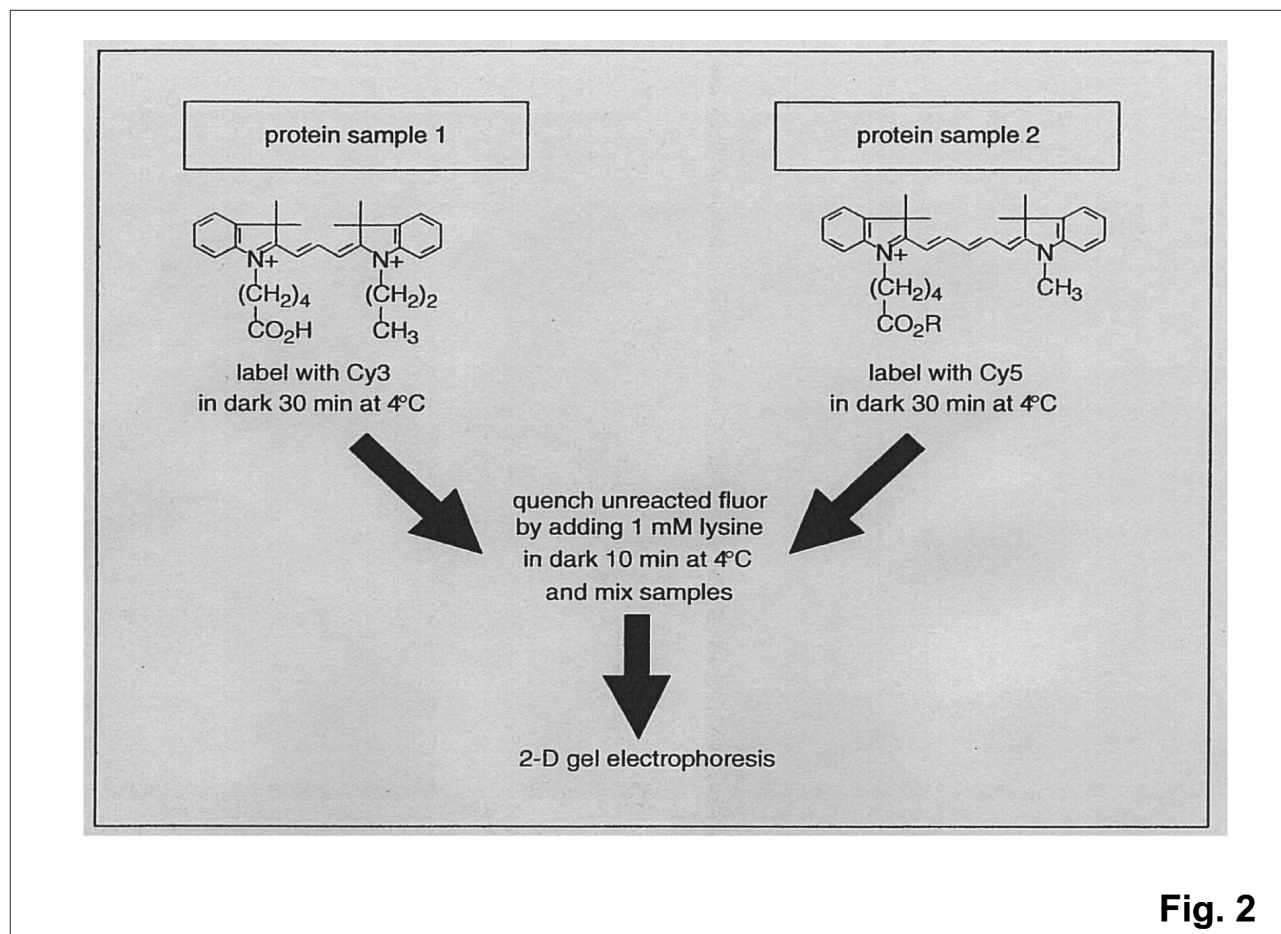


Fig. 1E

**Figure 2.** Scheme for labelling two samples whose protein profiles are to be compared by the differential in-gel electrophoresis (Cy3/Cy5 fluorophors) (from Lilley (45), by permission).



**Fig. 2**

highly reliable results, but it is terribly time consuming and truly labour intensive. An alternative to this protocol, could be the one depicted in Figure 2, known under the acronym of DIGE, differential in gel electrophoresis, as first described in 1997 by Unlu *et al.* (10). It is based on differential labelling with N-hydroxy-succinimide ester-modified cyanine fluors, the most popular couple being named Cy3 and Cy5. 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  $\epsilon$ -amino group of Lys, and the masses of the dyes are similar to each other. The reaction is carried out so as to label only a few Lys residues per macromolecule. As long as the extent of the reaction is similar between the samples to be compared, the mass shift will be uniform and the 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 2-D gel. The differences between the quantities of the individual proteins from each sample can then be determined using specialized 2-D image analysis software. Since both samples to be compared are se-

parated 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 (11), in the DIGE technique (12). 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 (13,14). One potential limitation of this method is that excision of spots of interest for identification by MS might be inaccurate, unlike in conventional 2-D gels. Coomassie or silver-stained gels can be excised manually or using a spot cutting robot because protein spots are visible. Robots enabled to "see" fluorescent spots do not exist as yet. Even so, excision still remains 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.

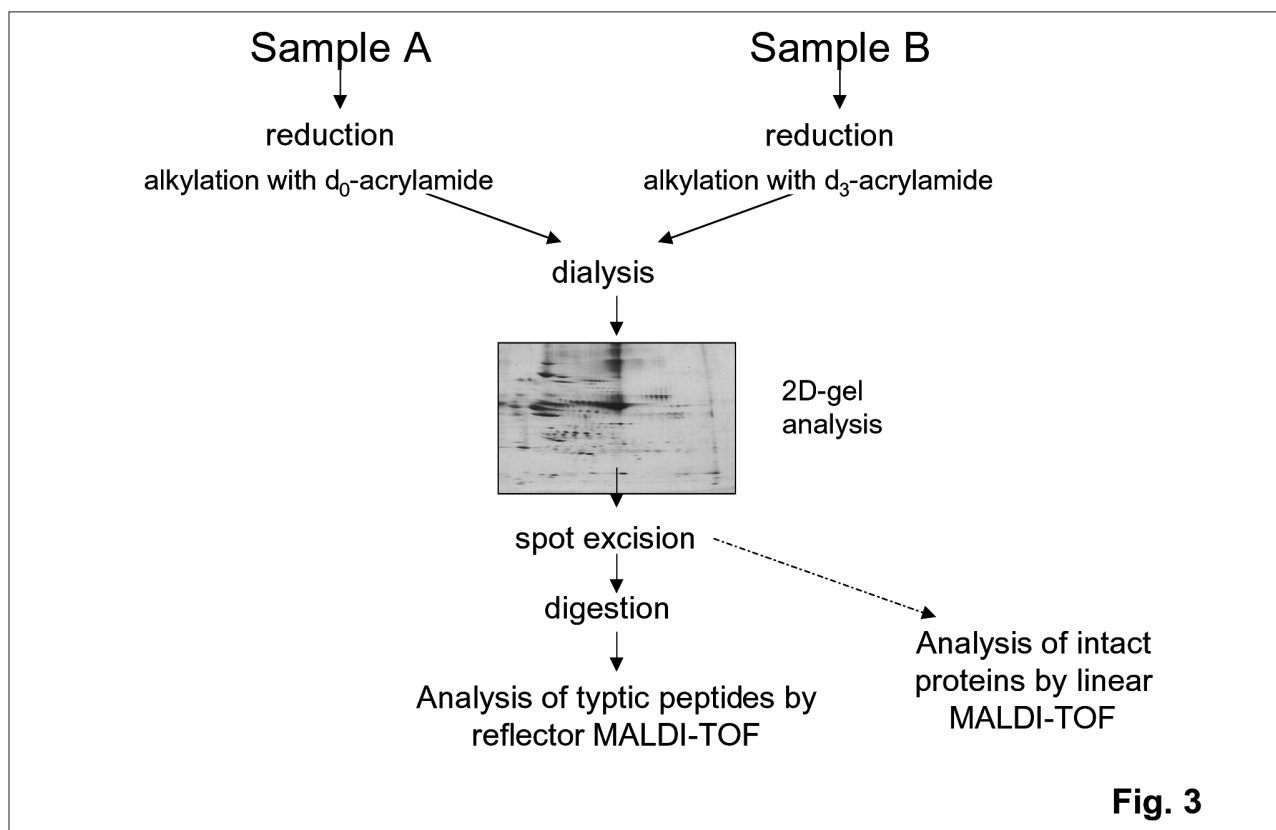
Therefore, the position of the bulk amount of unlabelled protein could be shifted about one spot diameter down (lower Mr values), but this could lead to excision of contaminants, different from the protein of interest. Should one carry the labelling of Lys to higher extents, the situation would be even 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.

### 2.3 Isotope-coded two-dimensional maps

In a third variant for quantitative proteome analysis by 2-D PAGE, one could exploit the technique of stable isotope tagging. This approach would utilize the same strategy depicted in Figure 2, but adopt labelling strategies involving light/heavy forms of the same tagging molecule. An example of such an approach could be the use of  $d_0/d_3$  acrylamide for blocking Cys residues in intact protein molecules. The use of light/heavy acrylamide to alkylate proteins prior to their 2-D electrophoretic separation was recently reported by Sechi (15) and by Gehanne *et al.* (16). 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 ap-

proach are depicted in Figure 3. Basically, relative quantification of individual proteins in two different samples is achieved by alkylating one sample with  $d_0$ -acrylamide, and the second with its  $d_3$ - counterpart; the two samples are then combined with predetermined ratios, dialyzed, and subjected to 2-D gel-electrophoresis. Following visualisation of the separated proteins, each spot can be excised, 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 2-D map of rat serum shown in Figure 4 is illustrated below. The map in Figure 3, 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  $d_0$ -acrylamide, and the second (70%) reacted with  $d_3$ -acrylamide. A representative example of a reflector MALDI spectrum that pertains to apo-transferrin is given in Figure 4(a-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.

**Figure 3.** Scheme for differential labelling of two samples with  $d_0/d_3$  acrylamide (alkylation of Cys residues). The central map refers to rat sera, labelled separately with either  $d_0$ - or  $d_3$ -acrylamide and mixed in a 30:70% ratio (from Gehanne *et al.*(16), by permission).



**Fig. 3**

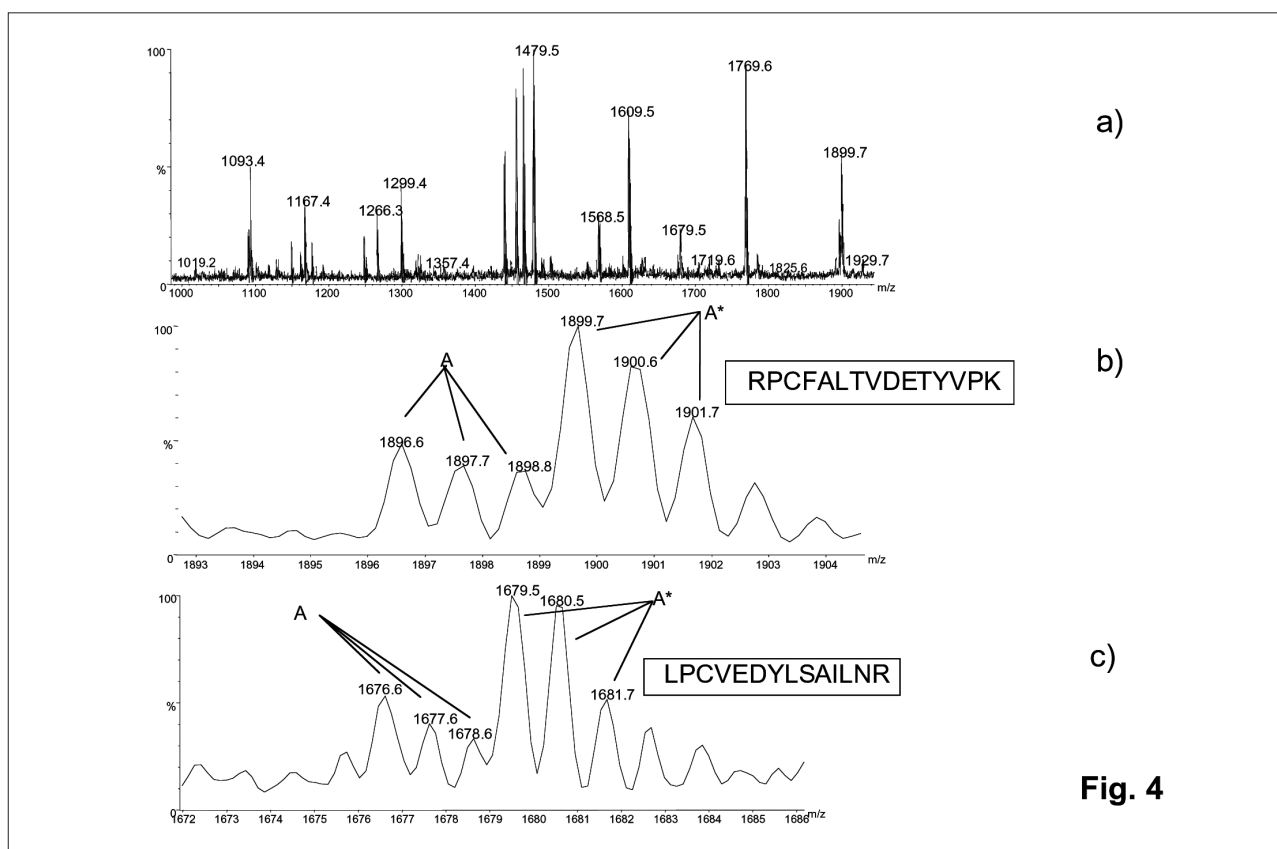
Considering the relative peak heights in both isotopic distributions, a ratio of 34:66 was obtained, which is in good agreement with the labelling ratio 30:70 prior to 2-D separation.

### 3 Quantitative proteomics based on chromatographic approaches

In spite of the contributions of 2-D PAGE to proteomics, there are shortcomings to this technology. High-throughput analysis of proteomes is challenging because each spot from 2-D PAGE must be individually extracted, digested and analyzed, a time-consuming process. In addition, owing to the limited loading capacity of 2-D PAGE gels and the detection limit of staining methods, 2-D PAGE presently has an insufficient dynamic range for complete proteome analysis. These shortcomings have encouraged development of alternative methods, which are today lumped together under the term "multidimensional chromatography" (17), a pompous definition hiding, in reality, just a two-dimensional approach like 2-D PAGE. In one instance, Raida *et al.* (18) coupled two chromatographic columns, first a cation-exchanger followed by RP-HPLC. The effluent of this last column was fed into an ESI MS machine and the peptide masses assessed. Opiteck *et al.* (19,20) described two different chromatographic ap-

proaches for separation of complex protein mixtures. In one approach, an *E. coli* lysate was injected onto a cation-exchange column and the eluates fed stepwise into a RP HPLC column. The eluate after this last step was sprayed into an ESI mass spectrometer. In a second system, size-exclusion chromatography was coupled to a RP HPLC column, again for analysis of *E. coli* cells. In both instances, however, too few proteins could be identified. In yet another approach, Link *et al.* (21) devised a discontinuous 2-D methodology, utilizing strong-anion exchange HPLC for a 1<sup>st</sup> dimension separation. Portions of the eluate were digested with trypsin and analyzed by RP-microcolumn HPLC. The eluted peptides were finally characterized by tandem mass spectrometry. Whereas this is an off-line method, Link *et al.* (22) subsequently devised an on-line approach, by which a peptide mixture from a digested *S. cerevisiae* was fed into a biphasic 2-D capillary column packed with a strong cation exchanger (SCX) juxtaposed to RP beads. Peptides were displaced iteratively by salt from the SCX resin into the RP pearls; this was followed by a classic eluant for RP columns, feeding the eluted peptides into an ESI MS-MS. After re-equilibrating the RP beads, raising the salt concentration displaced more peptides from the SCX resin onto the RP pearls, thus reiterating the process. In a single experiment, this method could resolve and identify 189 unique proteins

**Figure 4.** (a) Reflector MALDI mass spectrum of an *in-situ* digest of apo-transferrin taken from the 2-D map of rat sera displayed in Fig. 4, that were alkylated with  $d_0$ -acrylamide and  $d_3$ -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 Gehanne *et al.*(16), by permission).



**Fig. 4**

from a *S. cerevisiae* whole-cell lysate via SEQUEST interrogation. Most of these chromatographic approaches, unlike conventional 2-D PAGE, do not analyse intact protein species, but their entire digests obtained by trypsin treatment of a total cell lysate. Due to the enormous complexity of such a mixture, and to tackle the problem of properly quantifying the protein/peptide levels in control vs. treated samples, a number of quantitative approaches have been devised, as illustrated below.

### 3.1 Isotope-coded affinity tags

Perhaps one of the earliest protocols described is based on the so-called isotope-coded affinity tags (ICAT) (23,24). In this novel procedure, the 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 (whose general structure is depicted in Figure 5a) 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 "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 (Fig. 5b). 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. The complexity of the label is due to the need of reducing the enormous complexity of the peptide digest (which, for an entire eukariotic cell lysate, could be as high as 300 to 500 thousand components!). This, in turn, generates other problems: the large size of the ICAT probe (ca. 500 Da) complicates database searching for small peptides. Additionally, it will reduce its reactivity at the Cys site, reactivity that seems to be already meager with simpler iodinated molecules such as iodoacetamide (25,26), notwithstanding claims to the contrary (27). It has also been pointed out, by Rabilloud (28), that one protein out of seven does not contain Cys, and therefore it is reasonable to predict that at least 15% of all proteins in a total cell lysate will not be covered by this approach.

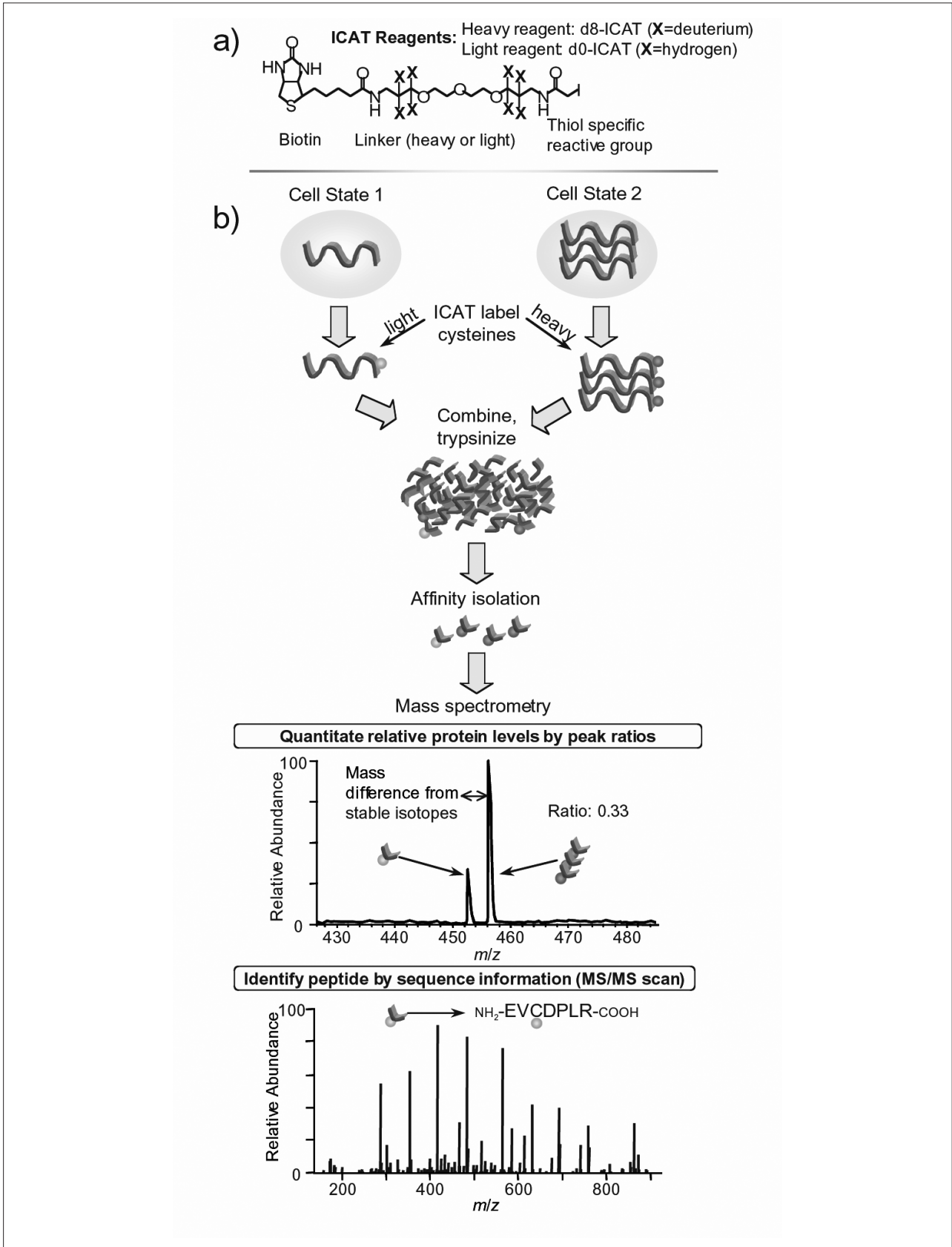
Curiously, however, the ICAT approach has come under attack for some, quite unexpected, reasons. It turns out that, if ICAT-labelled peptides are separated through a reversed-phase column, a disturbing well known isotope effect takes place, by which the deuterated peptide elutes earlier than its non-deuterated counterpart; that separation causes an enormous variation in isotope ratio across the two differently eluting profiles of the isoforms (29). The effect is more pronounced the smaller the tagged peptide. For example, in the case of a simple, Cys-bearing octapeptide, the chromatography resolution between the light/heavy species was as high as  $R_s=0.74$  (it is here recalled that an  $R_s=1.2$  means just base-line resolved peaks). Because the column eluate is continuously analyzed by ESI-MS, obtaining the correct quantitative peak ratio according to the isotopic ratio becomes an extremely difficult task. Zhang and Regnier (29) reported that the resolution of the isoforms, in a  $C_{18}$  column, exceeds 0.5 with 20% of the peptides in the digest. Such a strong chromatographic effect is not, in fact, so unexpected. For instance, even in capillary electrophoresis Chiari *et al.* (30) disclosed ample, base-line separation of benzyl- $H_5$ /benzyl- $D_5$  alcohols and of benzene- $H_6$ /benzene- $D_6$ , brought about by the presence of SDS micelles in the background electrolyte. Clearly, in this micellar electrokinetic process, the residence times of the light/heavy species must be quite different. Conversely, Zhang and Regnier (29) described the complete absence of such an isotope effect in the case of peptides differentially labelled with  $^{13}C$ - and  $^{12}C$ - succinate, so they strongly recommended this type of peptide-coding when attempting separations in  $C_{18}$  columns.

### 3.2 Other types of isotope coded tags

Due to the shortcomings of just tagging Cys residues, a complementary approach has been reported by Goodlett *et al.* (31) where methyl esterification (using  $d_0$ - or  $d_3$ - methanol) of peptides converted carboxylic acids on the side chains of aspartic and glutamic acids as well as the carboxyl terminus, to their corresponding methyl esters. The separate  $d_0$ - and  $d_3$ - methylated peptide mixtures were combined and subjected to capillary LC/tandem mass spectrometry. Ratios of various proteins in the original mixtures were obtained by normalization of the areas under the curves for the same charge states associated with  $d_0$ - and  $d_3$ - methylated peptides. In another approach, Peters *et al.* (32) reacted tryptic digests of proteins with 2-methoxy-4,5-dihydro-1H-imidazole, thus converting Lys residues to their 4,5-dihydro-1H-imidazol-2-yl derivatives. This accomplishes two goals: first of all, it increases the ionization efficiencies of Lys-terminated fragments; secondly, it permits differential quantitation studies, since  $d_0$  and  $d_4$  forms of this label can be produced. The same Lys residues could be reacted with O-methyl-



**Figure 5.** Scheme of the isotope-coded affinity tag (ICAT) strategy for quantification of protein expression. (A): structure of the ICAT reagent. It consists of 3 segments: an affinity tag (biotin); a linker, which can incorporate either deuterium or hydrogen; and a reactive tail specific for thiol groups. (B): ICAT strategy. Protein from two different cell states is harvested, denatured, reduced and labelled at Cys with the light or heavy ICAT reagents. The samples are then combined and digested with trypsin. After affinity isolation of the ICAT-labelled peptides, they are analyzed in a tandem mass spectrometer. The ratio of the ion intensities for each ICAT-labelled pair quantifies the relative abundance of its parent proteins in the original state. Finally, the MS spectrum enables sequencing and protein identification (from Gygi and Aebersold(24), by permission).



sourea, labelled with  $^{13}\text{C}$  and  $^{15}\text{N}$ -stable isotopes, for differential quantitation experiments (23). This reaction would transform Lys into a more basic homo-arginine residue; however, the mass difference between standard and isotope coded label would be of only 3 Da, with the additional burden of a quite expensive isotope labelling protocol. In yet another variant, a  $\text{H}_4/\text{D}_4$  labelling reagent can be obtained by reacting nicotinic acid with N-hydroxysuccinimide (33). The resulting 1- $[(\text{H}_4/\text{D}_4)]$  Nicotinoyloxy succinimide esters are then reacted with the N-termini of tryptic digests of proteins. This results in the N-terminal nicotinylation of all peptides, via formation of an amido bond between the  $-\text{NH}_2$  and the activated carboxyl of nicotinic acid. This leaves at each terminus a weakly basic, pyridine ring, containing either an  $\text{H}_4$  or a  $\text{D}_4$  moiety, allowing for quantitation of differential protein expression. An analogous method, based on labelling N-termini of peptides, has been recently proposed by Mason and Liebler (34). These authors described the use of  $\text{d}_0$ - and  $\text{d}_5$ -phenyl isocyanate (PIC) as N-terminal reactive tags for essentially all peptides in proteolytic digests. PIC reacts quantitatively with peptide N-terminal amines within minutes at neutral pH and the PIC-labelled peptides undergo informative MS/MS fragmentation. Ratios of  $\text{d}_0$ - and  $\text{d}_5$ -PIC-labelled derivatives of several model peptides were linear across a 10000 fold range of peptide concentration ratios, thus indicating a wide dynamic range for quantitation. Thus, PIC labelling seems to offer a versatile means for quantifying differential protein expression in paired samples.

Instead of adding isotope-coded tags, one could use extensive, differential tagging by adopting culture media enriched with different types of stable isotopes. For instance, Oda *et al.* (35) grew a yeast culture on a medium that contained the natural abundance of nitrogen, whereas another culture was grown on a medium enriched with  $^{15}\text{N}$ . In another approach, Yao *et al.* (36), during peptide bond hydrolysis, incorporated two  $^{18}\text{O}$  atoms into the carboxyl termini of the tryptic peptides of a first pool, whereas the peptides in a second pool were cleaved analogously with the carboxyl termini containing  $^{16}\text{O}$  atoms. The two peptide mixtures were pooled for fractionation and separation, and were examined by high-resolution MS.

### 3.3 Global internal standard strategy (GIST)

The examples given above can be described as particular cases of what Regnier's group has called "global internal standard strategy". Whereas, for example, in the case of the ICAT label, only Cys-containing peptides are differentially marked, it would be highly desirable to universally label all peptides, independent of their amino acid composition, and select any kind of strategy for the selective capture of any desired population of peptides for the analysis

of complex proteomes. This approach would have the advantage that any of a wide variety of derivatization and separation techniques could be applied for selection, identification, and quantification of peptides. In a first approach, Ji *et al.* (37) and Chakraborty and Regnier (38) introduced acylation of primary amino groups with either N-acetoxysuccinimide or N-acetoxy- $[\text{}^2\text{H}_3]$ succinimide. As usual, one or the other reagents were added to the experimental and control samples, respectively, after the total cell extract had been fully digested with trypsin. As a result, each tryptic peptide would increase its mass by 42 amu if it contained an Arg residue at the carboxy terminus; however, those peptides with a free amino-terminus and a C-terminal Lys would be acetylated twice and thus exhibit a mass increment of 84 amu. In turn, when the peptides from the two populations labelled with the light/heavy isotope, and mixed in a 1:1 ratio, were analysed by MS, the relative peaks will be shifted by 3 mass units, for the singly labelled, or by 6 mass units (for the doubly labelled), having an area ratio that is proportional to their relative abundance in the two sample specimens. In a second approach, the same group adopted acylation of peptides via light and heavy (deuterated) succinic anhydride (39,40). Here too, a few peptides that were differentially labelled and that terminated with an Arg will exhibit mono-isotopic peaks that are spaced apart by 4 amu, whereas those having a Lys at their C-terminus will exhibit a difference of 8 amu (occasional peaks that have a further difference of 12 amu in mass were shown to contain serine; the latter adduct was easily eliminated by treating the derivatized peptides with hydroxylamide at pH *ca.* 12). Because all peptides would carry a light/heavy label, a variety of strategies for their purification could be adopted, in order to reduce the enormous complexity that derives from the total lysate of a proteome of a single organism (just as an example, according to genomic analysis, a total *E. coli* lysate would generate 132,768 peptides; a total yeast lysate would produce as many as 333,247 fragments; and a total human digest would produce the non-negligible number of 2,641,532 peptides). As an example, Wang and Regnier (39) adopted covalent chromatography to capture all Cys-containing peptides from a total *E. coli* lysate. After labelling with light/heavy succinic anhydride, the Cys-carrying peptides were captured via disulfide interchange with sulfidryl groups on a thiopropyl Sepharose gel. After eluting unwanted material, such peptides were released from the gel by washing with an excess of DTT. In another case, the labelled population of peptides was purified via reversed-phase chromatography on a  $\text{C}_{18}$  column (38). In yet another variant, only a narrow fraction of peptides was captured (thus greatly reducing the complexity of the original digest); *i.e.*, those that contain a Cys and a His residue (40). First, all Cys

peptides were collected from a covalent chromatography step that involved thiol disulfide exchange. Upon elution from the thiol column, the same population was adsorbed onto an immobilized metal affinity chromatography column that had been loaded with copper. The emerging peptides were only those that contained both a Cys and a His residue (40). Finally, an isotope coding strategy involving both amine and carboxyl group labelling has been recently reported by Liu and Regnier (41). The method exploits differential derivatization of amine and carboxyl groups generated during proteolysis as a means of coding. Carboxyl groups produced during proteolytic attack incorporate  $^{18}\text{O}$  from  $\text{H}_2^{18}\text{O}$ . Peptides from the C-terminus of proteins were not labelled with  $^{18}\text{O}$  unless they contained a basic C-terminal amino acid. Primary amines from control and experimental samples were differentially acylated after proteolysis with either  $^1\text{H}_3$ - or  $^2\text{H}_3$ -N-acetyloxysuccinamide. This methodology offers several advantages: (a) it distinguishes C-terminal peptides in model proteins; (b) it recognizes N-terminal peptides from proteins in which the amino terminus is acylated; (c) it identifies primary structure variations between proteins from different sources. For more on this topic, one can consult a recent review covering in depth all aspects of quantitative proteome analysis (42).

## Conclusions

The aim of the present review was to assemble together the vast array of methodologies today existing for performing quantitative proteomics. It is of interest to notice that, except for the statistical analysis of 2-D PAGE slabs, exploiting software packages, all other methodologies have been reported only in the last few years. The vast majority of them, moreover, were developed specifically for proteome analysis performed by two-dimensional chromatography and cannot be easily adapted for gel-based methodologies, those relying on the classical O'Farrell approach (8). The reason is that most of those differential labels have been developed for tagging peptides, which are easily amenable to any kind of chromatographic procedure. In conventional, 2-D PAGE, only intact polypeptide chains can be mapped, since mixtures of peptides would not focus properly nor would be differentially sieved along the second dimension SDS-gel. Were this not enough, even if some resolution might be achieved in 2-D PAGE analysis, there would still remain the problem of properly fixing and staining all peptides with an  $M_r < 1000$  Da, species recalcitrant to just about any kind of protein fixative (43). Other caveats apply as well: for instance, the differential labelling of Lys residues, while working very well for peptides generated after tryptic digestion, could hardly be applied to intact proteins, since it would mask precisely those resi-

dues susceptible to trypsin attack. Even if one could circumvent this problem by using other proteolytic enzymes, there would still remain the major drawback that, in an intact protein, it is almost impossible to carry to 100% reactions onto amino acid side chains, such as onto Lys, as well as onto Asp and Glu residues. The impossibility of reaching 100% labelling would generate strings of spots, in the focusing dimension, from partially reacted species, exhibiting per force different pI values. This, in turn, would tremendously complicate the final 2-D map and render proteome analysis meaningless. Thus, we have to live with the present paradox in modern proteome analysis: although the method still much in vogue and most popular is the traditional electrophoretic approach based on large-size gel slabs, it is also the method that offers the most limited versatility in differential proteome analysis via the vast array of methods invented for labelling peptides (44,45).

## References

1. Wilkins MR, Williams KL, Appel RD, Hochstrasser D.F. Proteome Research: New Frontiers in Functional Genomics. Berlin: Springer; 1997.
2. Kellner R, Lottspeich F, Meyer H.E. Microcharacterization of Proteins. Weinheim: Wiley-VCH; 1999.
3. Rabilloud T. Proteome Research: Two-Dimensional Gel Electrophoresis and Identification Methods. Berlin: Springer; 2000.
4. James P. Proteome Research: Mass Spectrometry. Berlin: Springer; 2001.
5. Righetti PG, Stoyanov AV, Zhukov MY. The Proteome revisited: Theory and Practice of All Relevant Electrophoretic Steps. Amsterdam; Elsevier; 2001.
6. Figeys D, Ross M. Applied Proteomics. New York: Wiley; 2003 (in press).
7. Grandi G. Genomics, Proteomics and Vaccines, London: Wiley; 2003 (in press).
8. O'Farrell PH. High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* 1975; 250:4007-21.
9. Anderson NG, Anderson L. The Human Protein Index. *Clin Chem* 1982; 28:739-48.
10. Unlu M, Morgan ME, Minden JS. Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* 1997; 18:2071-7.
11. Patton WF. A thousand points of light: the application of fluorescence detection technologies to two-dimensional gel electrophoresis and proteomics. *Electrophoresis* 2000; 21:1123-44.
12. 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; 1:91-8.
13. Zhou G, Li H, DeCamp 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. *Mol Cell Proteomics* 2002; 1:117-24.

14. 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.
15. 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.
16. 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/ionization mass spectrometry. *Rapid Commun Mass Spectrom* 2002; 16:1692-8.
17. Washburn MP, Yates JR Jr 3<sup>rd</sup>. in Blackstock W, Mann M. *Proteomics: a Trend Guide*. London: Elsevier; 2000. p. 27-30.
18. Raida M, Schulz-Knappe P, Heine G, Forssmann WG. Liquid chromatography and electrospray mass spectrometric mapping of peptides from human plasma filtrate. *J Am Soc Mass Spectrom* 1999; 10:45-54.
19. Opiteck GJ, Lewis KC, Jorgenson JW, Anderegg RJ. Comprehensive on-line LC/LC/MS of proteins. *Anal Chem* 1997; 69:1518-24.
20. Opiteck GJ, Ramirez SM, Jorgenson JW, Moseley MA 3<sup>rd</sup>. Comprehensive two-dimensional high-performance liquid chromatography for the isolation of overexpressed proteins and proteome mapping. *Anal Biochem* 1998; 258:349-61.
21. Link AJ, Carmack E, J Yates JR Jr 3<sup>rd</sup>. *Int J Mass Spectrom Ion Proc* 1997; 160:303-
22. Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, et al. Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* 1999; 17:676-82.
23. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 1999; 17:994-9.
24. Gygi SP, Aebersold R, in W. Blackstock, M. Mann. *Proteomics: a Trend Guide*. London: Elsevier; 2000.p. 31-36.
25. Galvani M, Rovatti L, Hamdan M, Herbert B, Righetti PG. Protein alkylation in the presence/absence of thiourea in proteome analysis: a matrix assisted laser desorption/ionization-time of flight-mass spectrometry investigation. *Electrophoresis* 2001; 22:2066-74.
26. Galvani M, Hamdan M, Herbert B, Righetti PG. Alkylation kinetics of proteins in preparation for two-dimensional maps: a matrix assisted laser desorption/ionization-mass spectrometry investigation. *Electrophoresis* 2001; 22:2058-65.
27. Smolka MB, Zhou H, Purkayastha S, Aebersold R. Optimization of the isotope-coded affinity tag-labeling procedure for quantitative proteome analysis. *Anal Biochem* 2001; 297:25-31.
28. Rabilloud T. Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains. *Proteomics* 2002; 2:3-10.
29. Zhang R, Regnier FE. Minimizing resolution of isotopically coded peptides in comparative proteomics. *J Proteome Res* 2002;1:139-47.
30. Chiari M, Nesi M, Ottolina G, Rigetti PG. *J. Chromatogr* 1994; A670: 215.
31. Goodlett DR, Keller A, Watts JD, Newitt R, Yi EC, Purvine S, Eng JK, et al. Differential stable isotope labeling of peptides for quantitation and de novo sequence derivation. *Rapid Commun Mass Spectrom* 2001; 15:1214-21.
32. Peters EC, Horn DM, Tully DC, Brock A. A novel multifunctional labeling reagent for enhanced protein characterization with mass spectrometry. *Rapid Commun Mass Spectrom* 2001; 15:2387-92.
33. Munchbach M, Quadroni M, Miotto G, James P. Quantitation and facilitated de novo sequencing of proteins by isotopic N-terminal labeling of peptides with a fragmentation-directing moiety. *Anal Chem* 2000; 72:4047-57.
34. Mason DE, Liebler DC, *J Proteome Res*. In press.
35. Oda Y, Huang K, Cross FR, Cowburn D, Chait BT. Accurate quantitation of protein expression and site-specific phosphorylation. *Proc Natl Acad Sci USA* 1999; 96:6591-6.
36. Yao X, Freas A, Ramirez J, Demirev PA, Fenselau C. Proteolytic 18O labeling for comparative proteomics: model studies with two serotypes of adenovirus. *Anal Chem* 2001; 73:2836-42.
37. Ji J, Chakraborty A, Geng M, Zhang X, Amini A, Bina M, Regnier F. Strategy for qualitative and quantitative analysis in proteomics based on signature peptides. *J Chromatogr B Biomed Sci Appl* 2000; 745:197-210.
38. Chakraborty A, Regnier FE. Global internal standard technology for comparative proteomics. *J Chromatogr A*. 2002 ; 949:173-84.
39. Wang S, Regnier FE. Proteomics based on selecting and quantifying cysteine containing peptides by covalent chromatography. *J Chromatogr A* 2001; 924:345-57.
40. Wang S, Zhang X, Regnier FE. Quantitative proteomics strategy involving the selection of peptides containing both cysteine and histidine from tryptic digests of cell lysates. *J Chromatogr A* 2002; 949:153-62.
41. Liu P, Regnier FE. An isotope coding strategy for proteomics involving both amine and carboxyl group labeling. *J Proteome Res* 2002; 1:443-50.
42. Hamdan M, Righetti PG. Modern strategies for protein quantification in proteome analysis: advantages and limitations. *Mass Spectrom Rev* 2002; 21:287-302.
43. Rigetti PG, Chillemi F. *J. Chromatogr* 1978; 157:243.
44. Beardsley RL, Reilly JP. Quantitation using enhanced signal tags: a technique for comparative proteomics. *J Proteome Res* 2003; 2:15-21.
45. Lilley KS., *Current Protocols in Protein Science*. New York: Wiley; 2000.