Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Chromatography A, 1077 (2005) 1–18

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

Review

Limitations of current proteomics technologies

Spyros Garbis [a], Gert Lubec [b], Michael Fountoulakis [a,c,*]

[a] *Academy of Athens, Foundation for Biomedical Research, Division of Biotechnology, Soranou Ephessius 4, 11527 Athens, Greece*
[b] *University of Vienna, Pediatrics Department, Vienna, Austria*
[c] *F. Hoffmann-La Roche Ltd., Center for Medical Genomics, Basel, Switzerland*

Received 21 March 2005; received in revised form 15 April 2005; accepted 22 April 2005
Available online 10 May 2005

Abstract

Application of proteomics technologies in the investigation of biological systems creates new possibilities in the elucidation of biopathomechanisms and the discovery of novel drug targets and early disease markers. A proteomic analysis involves protein separation and protein identification as well as characterization of the post-translational modifications. Proteomics has been applied in the investigation of various disorders, like neurological diseases, and the application has resulted in the detection of a large number of differences in the levels and the modifications of proteins between healthy and diseased states. However, the current proteomics technologies are still under development and show certain limitations. In this article, we discuss the major drawbacks and pitfalls of proteomics we have observed in our laboratory and in particular during the application of proteomics technologies in the investigation of the brain.
© 2005 Elsevier B.V. All rights reserved.

Keywords: Proteomics; Mass spectrometry; Neuroproteomics; Proteomics limitations; Two-dimensional electrophoresis

Contents

1. Proteomic analysis 2
2. Proteomics methods 3
   2.1. Two-dimensional electrophoresis/matrix-assisted laser desorption ionization time-of-flight mass spectrometry 3
   2.2. Liquid chromatography–mass spectrometry 3
3. General limitations of the proteomic analysis 3
   3.1. Protein samples 3
   3.2. Protein abundance 4
      3.2.1. Low-abundance proteins 4
      3.2.2. High-abundance proteins 5
4. Drawbacks of two-dimensional electrophoresis 6
   4.1. Two-dimensional gel reproducibility 6
   4.2. Protein quantification 6
   4.3. Frequency of detection 7
   4.4. Protein hydrophobicity 7
   4.5. Protein charge 8
   4.6. Protein size 9
   4.7. Protein heterogeneity 9

* Corresponding author. Tel.: +30 210 6597069; fax: +30 210 6597545.
E-mail address: mfountoulakis@bioacademy.gr (M. Fountoulakis).

0021-9673/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2005.04.059

5. Drawbacks of mass spectrometry . . . 9
5.1. MALDI-TOF-MS . . . 10
5.2. Electrospray ionization . . . 12
5.3. Liquid chromatography–mass spectrometry . . . 13
5.4. Tandem mass spectrometry . . . 15
5.5. Instrument-related limitations . . . 16
6. Selection of the proteomic method . . . 16
7. Concluding remarks . . . 17
References . . . 17

1. Proteomic analysis

Proteomic analysis of biological systems has as goal the study of mechanisms and the discovery of novel drug targets and early disease markers. A proteomic analysis has usually two steps: (i) protein separation and (ii) protein identification which also includes the characterization of the post-translational modifications. There exist two major proteomics approaches: (i) two-dimensional (2D) gel electrophoresis for protein separation, followed by matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS) for protein identification and (ii) one- or two-dimensional liquid chromatography for protein and peptide separation combined with electrospray ionization (ESI) or tandem mass spectrometry (MS) for protein identification. A combination of the separation and identification technologies is also frequently applied. For the detection of low-abundance gene products, protein enrichment and separation techniques have to be employed prior to the proteomic analysis [1]. Fig. 1 gives the general workflow in proteomics.

In the proteomic analysis of the central nervous system (neuroproteomics), the first approach 2D gels/MALDI-TOF-MS has been mainly used, whereas liquid chromatography interfaced with mass spectrometry (LC–MS) has been applied less often in this area of research. Neuroproteomics mainly involves the analysis of the brain and of cerebrospinal fluid (CSF). The brain analysis is directed to post-mortem samples from patients with neurodegeneration diseases, like Alzheimer's disease (AD), down syndrome (DS), Creutzfeldt-Jakob disease, Pick's disease, and animals serving as models of human diseases, like AD, DS, anxiety, hypoxia, stroke, pain and control animals to study post-mortem changes, development-related changes, effect of neurotoxins and others. Usually, total proteins or proteins of subcellular fractions, cytosol, mitochondria, membranes, from human and animal brain regions have been analyzed [2,3]. Analysis of cerebrospinal fluid is almost exclusively limited to human material. Proteomics has been applied in the study of neurological cell lines as well [4]. At the beginning of brain proteomics, analysis of total extracts by 2D gels and MALDI-TOF-MS resulted in the identification of up to about 500 different gene products [5,6]. The use of pre-fractionation methods prior to the 2D gel analysis enabled the identification of more than 1000 gene products

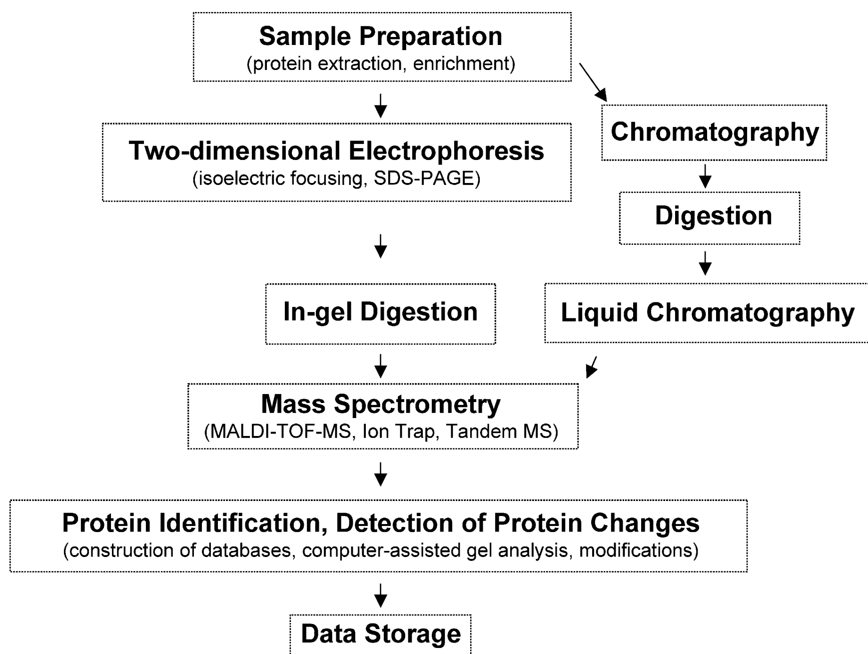

Fig. 1. Workflow schemes in proteomics. SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis.

(ref. [7] and unpublished results). The application of LC–MS/MS technologies resulted in the identification of 1685 proteins from hippocampus tissue [8].

2. Proteomics methods

2.1. Two-dimensional electrophoresis/matrix-assisted laser desorption ionization time-of-flight mass spectrometry

Proteomics in its classical definition involves protein separation by two-dimensional electrophoresis and identification by MALDI-TOF-MS. Two-dimensional electrophoresis has the advantage that it enables the simultaneous visualization of thousands of protein spots, the quantification of their levels and the detection of post-translational modifications. It is a well-established, robust and the most widely used approach in neuroproteomics and other research fields. In the first-dimensional separation, usually immobilized pH-gradient (IPG) strips [9,10] are used. The increased application of two-dimensional gel electrophoresis today is to a large extent due to the introduction of the IPG strips. The major advantage of using IPG strips is that they enable the application of larger sample quantities (up to several milligrams), a requirement for the efficient identification of low-abundance proteins by mass spectrometry, and the ability to maintain high reproducibility. The increase in the amount of sample applied and the improvement in reproducibility contributed to the realization of a high-throughput analysis of neuroproteomes and of any proteome in general. Although the application of several mg of protein amount has been reported, to our experience, use of more than 1.0–1.5 mg of protein may result in artefacts and irreproducible results.

In the second dimensional separation, multi-gel systems are usually employed. The use of multi-gel systems is essential for high-throughput proteomics because of the economical gel preparation and the simultaneous analysis of a large number of samples. Gels of large format (about 20 cm × 20 cm) allow an efficient spot resolution and are preferred when the goal is the detection of post-translational modifications. In such gels, a relatively high protein amount can be applied (0.3–1.5 mg of total protein or more) for the efficient detection of a large number of spots and most spots are correspondingly large. If the protein amount applied in a large format gel is relatively low (about 0.1–0.2 mg), the spot visualization may not be efficient and in addition protein losses may occur. In gels of small format (about 10 cm × 10 cm), the spot resolution is often compromised, however, a low protein amount (about 0.1 mg or less) is required and the spots are of small-size, which however include sufficient protein quantity for identification by mass spectrometry. This is essential when a low amount of total protein is available, for example in a sample of mitochondrial proteins from mouse brain hippocampus.

2.2. Liquid chromatography–mass spectrometry

The combination of liquid chromatographic techniques with subsequent mass spectrometric analysis constitutes a complementary approach to 2D gel based methods in that hydrophobic, low-molecular mass and basic proteins, the identification of which from 2D gels is inefficient, can be detected. Typically, protein mixtures are digested with various proteases, most common of which is trypsin, generating a composite mixture of peptides. The resulting peptide mixtures are then analyzed on-line by nanobore LC–MS using a C18 reverse phase capillary column with mobile phase flow rates ranging from 100 to 300 nl/min and a nano-electrospray ionization source. This general method, also referred to as one-dimensional (1D) LC–MS, is simple and amenable to automation and provides enhanced sensitivity primarily due to the small inner HPLC column diameter and the high ion sampling efficiency afforded by the nano-ESI source design. The proteomic analysis becomes more efficient with the use of multiple chromatographic steps prior to mass spectrometry. The term multidimensional chromatography is widely used but up to now only two-dimensional chromatography has been reported. Currently, two-dimensional liquid chromatography mass spectrometry has the disadvantage of inefficient detection of post-translationally modified proteins and protein quantification. Moreover, multidimensional LC is still under development, whereas 2D gels represent a well established, robust separation system today, but the method will further develop and it will find a wider application in the future.

3. General limitations of the proteomic analysis

The current proteomics technologies show certain limitations, which must be overcome in order for the science to develop its full power and capabilities. The limitations observed in a general proteomic analysis are met in neuroproteomics as well. An additional limitation in neuroproteomics is the quality of the samples themselves. The limitations in a proteomic analysis are of two kinds: (i) those related to the composition of the proteome to be analyzed, mainly concerning protein expression levels (which appear in all proteomics analyses), and (ii) limitations of the analytical methods. Proteomics is, therefore, simultaneously trying to increase the quantity of the low-abundance proteins to allow an efficient detection, and to apply the proper analytical methods to visualize all proteins of a mixture.

3.1. Protein samples

Sample preparation for a proteomic analysis is delicate work and the intrinsic difficulty lies in the fact that the vast diversity of protein molecular sizes, charge state, hydrophobicity and hydrophilicity indices, protein conformational states, post-translational modifications, complexation

with other macromolecular biological entities and enzymatic co-factors and cellular distribution, makes it unfeasible to use a single sample preparation protocol that sufficiently captures the entire proteome for a given biological system. An ideal sample preparation protocol would reproducibly and non-intrusively isolate the full complement of proteins in a given biological sample, concurrently eliminating post-extraction modifications and non-specific contaminants, like auto-digest peptides, keratins, fatty acids, plastic polymers, phthalates, nucleic acids, and artifacts, like MALDI matrix components, salt clusters and adduct species.

The brain and CSF analyses are complex, because additional factors may be involved, such as differences among individuals, differences in age and sex, possibly other diseases, treatment with medicines, as well as technical, disease-unrelated factors, like post-mortem time (for the brain samples), improper treatment of the samples etc., all of which can affect a clear discrimination between healthy and diseased states of interest. Significant variations for certain proteins have been observed from individual to individual even in control groups. Fig. 2 shows examples of differences in the spots representing glial fibrillary acidic protein in control brain. Proteomic analysis of the brain of animals serving as models of human neurological diseases may be advantageous as most of the sample limitations mentioned above are not present. However, we have observed differences in the brains of control animals as well which can be of allelic nature or can be the result of artifacts of the technology [11].

Human brain samples are taken in an average time period of 30 h after death, so that within this period certain alterations may occur. For example, additional spots representing dihydropyrimidinase-related protein 2 were detected within 6 h after death [12]. Furthermore, it is possible that the tissue excision from the various brain regions may not have been accurately performed and that the samples analyzed were not excised from corresponding brain regions in the control and the disease groups. Use of laser-capture microdissection may be useful to excise pure populations of the corresponding regions [13]. However, this approach has the limitation that the tissue pieces dissected contain a low protein quantity so that excision of a large number of pieces is required to obtain a sufficient protein amount for a proteomic analysis.

In addition, the human brain samples used in a proteomic analysis are usually frozen as timely coordinated biopsy and proteomic analysis are difficult to achieve. Preparation of subcellular fractions from frozen samples is not optimal and a cross-contamination of subfractions and organelles may occur, so that in many cases we can speak of subfractions enriched in a protein class [6]. Inefficient separation of organelles may have as a consequence that certain low-abundance gene products are not visualized and not detected. Moreover, differences observed in 2D gels may be partially due to gel artifacts, for example it can often happen that not the whole protein amount enters the IPG strip, in particular for high-abundance components. Furthermore, there is often a variation from gel to gel. The effects of artifacts can be diminished with the analysis of a large sample set at least in duplicate and preferably in triplicate (see Section 4.1 also).

3.2. Protein abundance

Proteins are not expressed in equal amounts and there are large differences in the protein levels in all proteomes. A proteomic analysis has to employ the proper technologies for the detection of all proteins. This can not be achieved by one single approach. There exist a plethora of approaches which can be used for the enrichment and visualization of the low-abundance proteins and also for the depletion of the high-abundance proteins which may hinder the detection of the less abundant counterparts.

3.2.1. Low-abundance proteins

In the small sample volume that is usually applied in a proteomic analysis (about 10–300 μl in a 2D gel), a large percentage of the expressed proteins, like transcription factors, occur in low abundance levels and can not be readily detected during the analysis of total proteins. The study of the low-copy-number gene products is interesting because such proteins are most likely the potential drug targets and disease markers. With today's status of the analytical methods, identification is usually successful if the protein is present

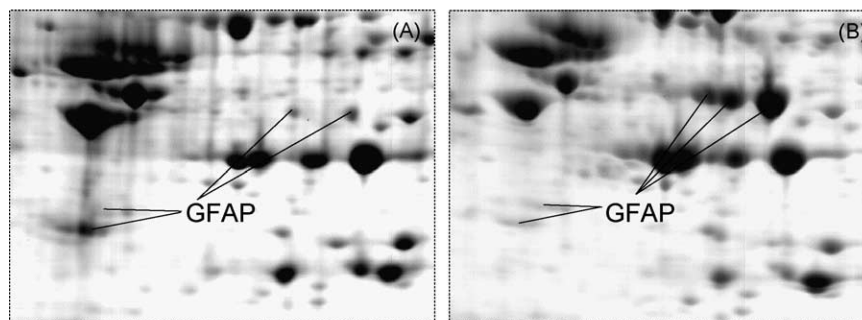

Fig. 2. Partial gel images showing human brain from the parietal cortex of control individuals (A, B). The proteins were separated on 3–10 NL IPG strips and the gels were stained with colloidal Coomassie blue. The spots representing glial fibrillary acidic protein (GFAP) are indicated. Different levels of larger and shorter forms of GFAP in the two samples were found.

at such an amount that the corresponding spot is visible in a polyacrylamide gel after stain with Coomassie blue. The practical protein amount for a MALDI-TOF-MS analysis lies in the amol range, however, because of losses during peptide extraction from gels and sample dilution requirements, in order to have sufficient volume for the automatic operations, the starting material should be in the low pmol or high fmol range i.e., the corresponding spot visible after stain with Coomassie blue (about 20 ng).

In order to be present in sufficient quantities and be detected, the low-abundance proteins should be enriched from crude extracts. Enrichment is performed by biochemical protein-enriching approaches. The original protein mixture is separated into less complex fractions, and each contains a lower number of total proteins in comparison with the starting material. This fractionation increases the likelihood of detecting low-abundance proteins. Two approaches are usually employed: (i) separation of the mixture into subcellular fractions and organelles, and (ii) enrichment of proteins from larger volumes by selective fractionation, immunoprecipitation, chromatographic, or electrophoretic methods [14–16].

Chromatographic methods reduce the complexity of the protein mixtures because of different binding principles, and every approach adds a unique resolving power. The proteins are separated by affinity, charge, hydrophobicity, or size [17]. The choice of the chromatographic method best-suited to fulfill the experimental requirements is essential for the success of the experiment [18]. Sequential chromatographic steps are often required. Enrichment of low-abundance brain proteins has been performed by various chromatographic steps like heparin chromatography and ion-exchange chromatography prior to proteomic analysis [19,20]. Detection of low-abundance gene products can be achieved with the use of equalizer beads. The method consists of a library of combinatorial ligands coupled to small beads. When the beads are impregnated with complex proteomes, they are able to significantly reduce concentration differences and to enhance detection of low-abundance proteins [16].

The electrophoretic methods comprise the separation of protein mixtures by preparative polyacrylamide gel electrophoresis on the basis of protein size, or by preparative isoelectrofocusing on the basis of protein charge. Preparative electrophoresis is performed in the presence of ionic detergents. With that approach, mainly low-molecular mass, neuron-specific, calcium-binding and 14-3-3 proteins brain proteins can be enriched [21,22]. Preparative isoelectrofocusing is performed either in the presence of ampholines (Rotofor system, Bio-Rad) or with the use of multi-compartment electrolyzers with isoelectric IPG membranes [15,16,23], in an IPG-based approach (Proteome Systems).

The chromatographic approaches for protein enrichment have certain limitations: (i) The chromatography steps used to enrich low-abundance proteins result in the enrichment of both low- and high-abundance proteins. This enrichment is to be expected, because a number of high- and low-abundance proteins bind to the column matrix on the basis of the same principles. Certain enriched proteins may represent up to 50% of the protein content in a particular fraction, and thereby suppress the signals of the low-abundance proteins and hinder their detection in 2D gels. To reduce the presence of the high-abundance proteins, for example heat shock proteins and house-keeping enzymes or albumin, they can be specifically removed by affinity chromatography with the use of specific antibodies, prior to the application of a general enriching method, for example ion exchange chromatography. However, removal of the high-abundance proteins may also simultaneously remove several low-abundance proteins which are associated with the high-abundance counterparts. (ii) Proteins eluted from columns usually need further treatment, for example to adjust the pH, to remove salt, chaotropes and other agents, which have been used for their elution and which are incompatible with IEF, or need to be concentrated because they are eluted in large volumes. Such operations may affect the composition of the protein mixture and may result in protein loss. (iii) After having applied two or more fractionation methods, the use of additional chromatographic separations may not result in the detection of significant numbers of new proteins that were not detected or enriched by the previous methods.

3.2.2. *High-abundance proteins*

Detection of low-abundant proteins is laborious and protein enrichment steps are required for their visualization. On the other hand, the presence of high-abundance proteins in certain samples makes the detection of the other, less abundant proteins a difficult undertaking. Such samples are plasma and cerebrospinal fluid. The latter is widely used in neuroproteomics studies. Body fluids, like plasma, are very interesting from the medical point of view in that many cells communicate with them and a large number of cells release at least part of their content into them. Although they can be easily obtained and provide important biomedical information, their proteomic analysis, however, is difficult [24]. The reason is that plasma contains about ten high-abundance proteins which together represent about 98% of the total protein content. Probably, less than 1% of all proteins are prime targets for the identification of novel markers [25]. The most abundant proteins are albumin and the immunoglobulin heavy and light chains, together representing about 80% of the plasma protein content. CSF samples mainly contain plasma proteins and therefore the detection of low-abundance proteins is extremely difficult in this sample as well. Therefore, the successful search for disease markers in plasma and CSF is associated with the depletion of at least these two high-abundance proteins. This is performed with the use of chromatography and precipitation steps [26–29]. Although several hundred gene products have been identified in body fluids, the classical proteomic analysis (2D gels) may not be the right tool in analyzing samples which include dominating proteins like albumin. Complete removal of such proteins, even if it can be achieved, may not be desirable because they will trap many of the low-abundance proteins

along with, their fragments and peptides, which will thus be lost and not detected. In contrast to plasma, which can be obtained in relatively large volumes (about 10 ml) and has a high protein concentration (about 65 mg/ml), the CSF analysis is more difficult as this sample is obtained in small volumes (2 ml) and the protein concentration is relatively low (about 0.5–1 mg/ml). In parallel to the classical method, alternative proteomic approaches may be useful, such as 2D chromatography for protein and peptide separation and enrichment and subsequent identification by tandem MS or MALDI-TOF-MS. The presence of the high-abundance proteins will still be an issue for these methods in that their removal will entail the drawbacks discussed above.

4. Drawbacks of two-dimensional electrophoresis

For the detection of a protein in a 2D gel, three prerequisites must be fulfilled: (i) the protein should be available in a sufficient amount in the protein mixture prior to the proteomic analysis, (ii) the protein should be brought into solution with mild detergents and chaotropes, compatible with isoelectric focusing (IEF), and kept in solution during the whole 2D separation, (iii) the protein should belong to the category of proteins that can be visualized by 2D electrophoresis; i.e., it should have average p*I* and molecular mass values and should not include strong hydrophobic stretches. Proteins which fulfill these conditions can be visualized in 2D gels. The gel technology and the protein detection show certain limitations which are discussed below.

4.1. Two-dimensional gel reproducibility

Although the use of IPG strips resulted in increased gel reproducibility, still the issue remains a major limitation of the method. The reproducibility of separation in 2D gels is usually satisfactory when relatively low protein amounts are applied on the gels (less than 0.5 mg). When larger protein amounts are applied (about 1 mg or more), which is often the case because it is advantageous for the detection of low-abundance proteins and for the subsequent MS analysis of the spots, significant variations in the spot number can be observed because not all proteins enter the IPG strips. Fig. 3 provides an example of the reproducibility of separation of brain proteins, following IEF on pH 3–10 non-linear IPG strips. Although several differences can be detected, the reproducibility concerning both the position and the intensity of the protein spots can be considered as satisfactory. Reproducibility of results between laboratories is more difficult because of variations in the 2D electrophoresis approaches and protocols and artifacts of the technology. Therefore, identifications solely derived from 2D gel comparisons should be considered with caution. In general, 2D gel preparation is the most delicate part of a proteomic analysis scheme and requires skilled personnel.

4.2. Protein quantification

Protein quantification is usually performed with commercially available software for spot detection and spot volume measurement. Spots are first outlined automatically, manually corrected, and then quantified using specific software. The percentage of the volume of the spot(s) that represent a certain protein is determined in comparison with the total proteins present in the 2D gel after background subtraction. Several gels (at least three, usually five) that carry the same sample are evaluated to determine average values and reduce the effect of gel artifacts. Only changes in the protein levels which are statistically significant are considered ($P < 0.05$). This approach has the advantage that it is relatively simple and relies exclusively on 2D gels. The major drawback are the 2D gel artifacts (the whole amount of a protein does not

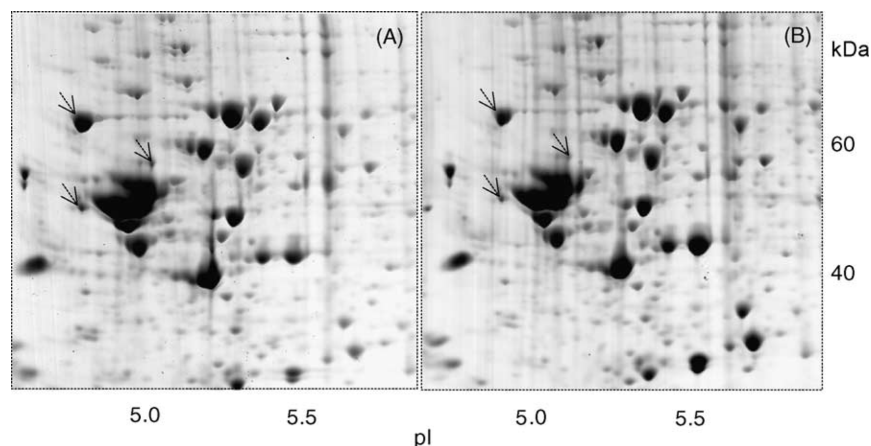

Fig. 3. Partial 2D gel images showing the reproducibility of separation of rat brain proteins (A, B). The proteins were separated on 3–10 NL IPG strips and the gels were stained with colloidal Coomassie blue. (A, B) separation of rat brain proteins, 1.0 and 1.5 mg, respectively. All spots are present in both gels and the spots in image B are larger in comparison with the corresponding spots of image A, where a lower protein amount was applied. However, there are differences between the two images, which should not exist (indicated by the arrowheads).

always enter the IPG strip, in particular, for large components of a protein mixture) and the inaccuracy of level determination because of the large volume difference between weak and strong spots. Moreover, the method is laborious as it requires careful spot definition and detailed control for possible mismatches. There exist several software for protein quantification from 2D gels which are evaluated in [30].

There are also other, more sophisticated, quantification approaches, which are based on chemical modification of the sample, like the 2D fluorescence difference gel electrophoresis approach (DIGE, Amersham Biosciences) [31] and methods that involve sample labeling, chromatography steps and MS analysis, like the isotope-coded affinity tags (ICAT) method, the visible isotope-coded affinity tag (VICAT) and the isotope tags for relative and absolute quantitation (iTRAQ) [32–34]. Such methods are more accurate, however, they are more complex to perform compared to the gel comparison, and rely on the presence of certain residues in the peptides (cysteine, lysine) which react with the dyes or specific groups carrying heavy and light isotopes and the availability of chromatography and MS equipment and of experienced personnel.

4.3. Frequency of detection

In most cases, the proteomic analysis of a certain protein sample, like brain, results in the detection of the same gene products, whereas other proteins are never found. Thus, most of the proteins identified in a brain sample have also been detected in other human protein samples [5]. In 2D gels from total brain extract, which has not been subjected to any fractionation or protein enrichment procedure, the most frequently detected proteins are heat shock proteins, heat shock cognate, heat shock 70 kDa protein, and 78 kDa glucose-regulated protein, which have been found in more than 300 human samples analyzed by mass spectrometry in our laboratory. Other frequently detected proteins are structural proteins, like tubulin chains and house-keeping enzymes, like α-enolase and ATP synthase β-chain, which have been found in more than 200 samples. A relatively low number of gene products have been detected only once in one of the gels which carry a total protein extract from a brain sample. These are mainly low-abundance proteins. The majority of the proteins have been found in 5–50 samples [2].

The frequency of detection delivers certain messages concerning the biological significance of the results, and the applied MS approach: (i) When total proteins or main subcellular fractions are analyzed, there are certain proteins, such as heat shock proteins and house-keeping enzymes, which are consistently detected. These proteins are abundant, hydrophilic, and are easily solubilized, and consequently they are present in most gels. Moreover, they are easily digested and deliver a sufficient number of peptides, so that an identity can be almost always assigned. (ii) During the analysis of a new sample, it is possible that not all frequently detected components (the proteins found in 200 or more samples), will be found. The failure of their detection indicates that there is a variation in the identification process concerning mainly technical aspects, like spot excision, digestion, peptide recovery, mass spectra acquisition, and identity assignment. (iii) The most frequently detected proteins in a sample like brain are the heat shock cognate and glucose-regulated protein. Those proteins have also been most frequently detected in other proteomes; for example, mouse and rat liver [35]. Therefore, they can be considered as internal markers (positive control) of a successful identification process during a protein batch analysis. (iv) The information that derives from the detection of a gene product in the proteome, which is currently analyzed, is probably of limited value if the gene product has been already detected in most previously analyzed protein samples. The less-frequently detected gene products are possibly more interesting in proteomic studies, because such proteins are most likely the potential drug targets and diagnostic markers, and their changed levels or modifications may carry more significant biological information than those of their frequently detected counterparts.

4.4. Protein hydrophobicity

Detection of hydrophobic and membrane proteins in 2D gels is related to certain limitations, and a relatively low number of real hydrophobic and membrane proteins have been detected. Using current 2D electrophoresis products (i.e., non-ionic or zwitterionic detergents like CHAPS, and non-ionic, mild chaotropes like urea or thiourea), mainly the abundant, hydrophilic components of a protein mixture are resolved and visualized [36–38]. Analysis of membrane proteins is of particular interest because they exert important functions in signal-transduction pathways, ion transport, cell–cell interactions, and other processes. Many of those proteins are drug targets, and are consequently of medical interest.

The unsuccessful detection of hydrophobic proteins is usually because of two reasons, the low solubility of certain proteins and their hydrophobicity. A protein can only be visualized and analyzed if it can be brought and kept in solution during the whole enrichment process and 2D electrophoretic analysis. In general, the solubility problem arises at two time points along the performance of the 2D electrophoresis: (i) during the initial extraction step of membrane proteins (the cytosolic proteins are already in solution), using agents that are compatible with isoelectric focusing, such as urea and CHAPS, and (ii) during the performance of the first-dimensional separation, when hydrophobic proteins could precipitate at their application positions. Poorly soluble proteins can be brought in solution with the use of strong detergents and chaotropes, like SDS or LDS [38]. The ionic detergents must be exchanged against IEF-compatible detergents prior to the analysis. During the exchange of the solubilizing agents serious protein losses may occur. Hydrophobic proteins with multiple transmembrane regions do not usually enter the immobilized pH gradient strips during the first

dimensional separation, and consequently a low number of transmembrane proteins have been detected by the 2D gel approach.

For the visualization of all proteins of a proteome, they should all be solubilized and kept in solution during the proteomic analysis. However, no single solubilizing agent is sufficient for the solubilization, resolution, and visualization of all proteins of a proteome. Various detergents have been tried for the efficient protein extraction and their use resulted in improved identification rates [39,40]. A further limitation is that proteins solubilized in a detergent and chaotrope system might not be detected in 2D gels on account of their hydrophobic domains as discussed above. In spite of the limitations mentioned, 2D gel analysis of membrane fractions is useful and may result in the visualization of a relatively large number of protein spots [38]. However, one should be aware that the detected proteins are mainly hydrophilic when judged with their grand average hydrophobicity (GRAVY) values, their transmembrane domains may not be strongly hydrophobic, or they may be contaminants from other fractions.

There exist alternative approaches to detect hydrophobic proteins: (i) Separation in a discontinuous, two-detergent system, where the proteins are separated first in the presence of a cationic and then of an anionic detergent [41]. This system does not include an IEF step, and it is efficient in the visualization of hydrophobic proteins and proteins that are difficult to solubilize, but it is compromised by the fact that a relatively low number of spots are detected and overlapping of spots representing high- and low-abundance components might occur. (ii) Separation of the original protein sample into detailed subfractions and organelles and analysis of the proteins of each fraction by one-dimensional SDS gels and identification by MALDI-TOF-MS or LC–MS. This method has the advantage of the distribution of the proteins into smaller fractions and the solubilization of the membrane proteins with a strong detergent. It has the disadvantage of the poor resolution in one dimension only but this drawback can be overcome by the use of LC–MS. Following two-dimensional chromatography, including hydrophobic interaction steps, LC–MS may result in the detection of hydrophobic proteins, for which the classical 2D gel proteomic analysis is less beneficial.

The degree of hydrophobicity of a protein can be calculated by the GRAVY values [42]. GRAVY scores provide a picture of the hydrophobicity of the whole protein, and usually vary in a range of ±2. Positive scores indicate hydrophobic, and negative scores hydrophilic proteins. GRAVY values do not appear to represent a reliable criterion on whether a protein will enter the IPG strip. It seems that the amino acid sequence of the hydrophobic stretches is decisive whether a protein will enter the IPG strip and not the hydrophobicity of the entire protein. As example, we mention cytochrome P450 2D6, which is a hydrophilic protein, judged by its GRAVY value, and it was not detected in 2D gels, but at the sample application position, on the border of the gels. The protein comprises one transmembrane region and it probably did not enter the IPG strips because of the strong hydrophobic domain [38].

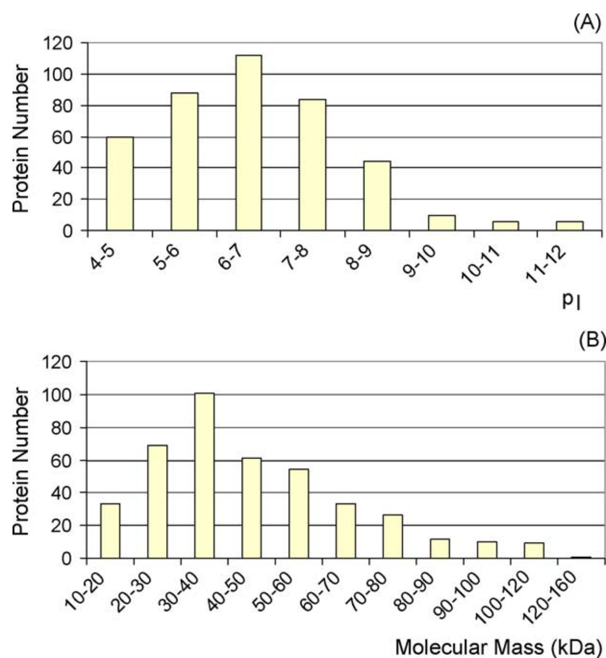

Fig. 4. Distribution of the human brain proteins identified in our laboratory in relation to their theoretical p*I* (A) and molecular mass (B) values. The bars indicate the number of proteins found in the indicated p*I* and the molecular mass intervals.

4.5. Protein charge

Proteins with average p*I* values are those that are frequently detected in 2D gels, whereas proteins with extreme values are less often detected. For example, Fig. 4A shows the distribution of the human brain proteins identified from broad pH range 2D gels according to their p*I* values. About 70% of the proteins have theoretical p*I* values between 5 and 8, 15% between 4 and 5 and 15% higher than 8. Acidic proteins with p*I* values lower than 4 have seldom been detected in 2D gels probably because of detection limitations (the lower p*I* limit is about 3.5). Certain proteins have theoretical p*I* values higher than 10, even if the analysis is performed in pH 3–10 IPG strips. Proteins with p*I* values higher than 10 should theoretically not be detected in 2D gels. Those proteins are most likely represented by multiple spots, and the most acidic forms are detected within the nominal pH range of the strips. Approaches to detect acidic and basic proteins are (i) use of narrow pH range, acidic or basic, IPG strips instead of the broad pH range strips. The detection of very acidic (p*I* 2.5) and very basic (p*I* 12) proteins in 2D gels have been reported [43–46], (ii) separation of the protein mixture into simpler fractions, for example by chromatography, and analysis of each fraction by one-dimensional SDS–polyacrylamide gels, which are mostly independent of the protein charge and (iii) analysis of the protein mixture by LC–MS.

4.6. Protein size

The majority of the proteins visualized in 2D gels have average molecular mass values and proteins with low- and high-molecular masses are underrepresented in 2D gels, because of limitations of the technology. Fig. 4B shows the distribution of the human brain proteins detected from 2D gels sorted according to their masses. Approximately 70% of the identified proteins have masses between 20 and 70 kDa, and eight percent of the proteins have masses between 10 and 20 kDa. Proteins smaller than 10 kDa and larger than 120 kDa are not frequently detected. Detection of low-molecular mass proteins is inefficient because (i) the lower mass limit of the gels is about 8 kDa and smaller proteins are practically not detected. In general, a relatively small number of protein spots that migrate below 10 kDa are visible in any type of 2D gels. (ii) Low-molecular-mass proteins are not efficiently stained with Coomassie blue. (iii) Small proteins are usually lost during the 2D electrophoretic process (this is independent of the stain with Coomassie blue). We tested the detection of small proteins after staining with Coomassie blue, silver and radiolabeling and no significant differences were observed, because the small proteins were simply lost [47]. (iii) Moreover, small proteins are difficult to identify by MALDI-TOF-MS with the typical matrix species used for proteomic studies (this is a limitation of the MS method).

There are some ways to partially overcome the size limitations: (i) Small proteins can be visualized in Tricine gels [48]. With this system, polypeptides of about 6–25 kDa can be efficiently separated. (ii) Enrichment of low-molecular-mass proteins can be achieved by preparative electrophoresis [21]. This method efficiently enriches small proteins whereas enrichment of large proteins is inefficient. (iii) Use of LC–MS may help in the detection of a larger number of low-molecular mass proteins. Except of the low-molecular mass, the high-molecular mass proteins are not efficiently detected in 2D gels either. Large proteins enter the IPG strips with low efficiency, and therefore, a large percentage of them are not detected in the gels, although they yield a high number of matching peptides and could be easily identified by MALDI-TOF-MS. Overall, a larger number of high-molecular-mass proteins are usually detected in 2D gels in comparison with the small proteins. High-molecular-mass proteins can be efficiently detected with the use of "soft" IPG gels [49].

4.7. Protein heterogeneity

Detection of multiple spots representing one gene product does not really represent a limitation of proteomics on the contrary, it constitutes a power of the 2D gels for the detection of post-translational modifications. We include the heterogeneity issue in the limitations just to emphasize the complexity of the brain proteome (and most proteomes, in general) and the necessity of developing more powerful technologies for the detection and understanding of heterogeneities.

A large percentage of the brain proteins show heterogeneity and are represented by more than one spot. Only about 100 proteins were represented by one or two spots out of the 450 different gene products identified in fetal brain. These are mainly enzyme subunits and structural proteins. For about 200 proteins, up to five spots were detected. All other gene products were represented by a larger number of spots-five of them (mainly tubulin chains or abundant enzymes) by more than 100. Based on our data, we estimate that, on average, one brain gene product is represented by three to five spots [2,5]. The heterogeneity observed in 2D gels is most likely higher because not all possible spots that represent the various proteins are excised for MS analysis or the identification process is successful.

The high-abundance proteins show the highest heterogeneity. The multiple spots may be partly the consequence of different splicing, processing, and post-translational modification, which result in an alteration of the p*I* of the polypeptides and consequently of the focusing position. Post-translational modifications include phosphorylation, glycosylation, deamidation, and other alterations. Heterogeneity may also result from artifacts of the technology, such as carbamylation of the proteins upon prolonged contact of the sample with urea, although carbamylation seems to be rather a seldom event and does not occur during the IEF process [50]. For most of the observed heterogeneities, we do not know their origin or their biological significance. For the efficient detection of post-translational modifications, a significant improvement in the sensitivity and throughput of the analytical techniques is required. The high-throughput protein search tools provide an indication on whether certain post-translational modifications, such as methionine oxidation or phosphorylation, are present. Confirmation of the actual existence of biologically significant modifications requires the use of more sophisticated tandem mass spectrometry based analysis. Glycosylation analysis is more complex and has not yet been established as a routine technology.

5. Drawbacks of mass spectrometry

Protein identification is usually accomplished by mass spectrometry. Success in the identification may vary with the sensitivity of the mass spectrometer, the completeness of the database, the presence of isobaric masses, post-translational modifications and other factors. During the past 20 years, advances in mass spectrometry instrumentation and associated techniques played a pivotal role in the fuller understanding of protein chemistry [51–55]. A breakthrough of the application of mass spectrometry to study peptides and proteins is the generation of stable gas-phase ions. Of the ionization techniques available today, only electrospray ionization and matrix assisted laser desorption ionization have the ability to exhibit high sensitivity and to generate ions without causing significant chemical decomposition. Mass spectrometry, on the other hand, faces certain limitations as well, mainly re-

ferring to sample preparation, instrument performance, and measurement of certain protein classes.

5.1. MALDI-TOF-MS

MALDI-TOF-MS is the most efficient approach in high-throughput proteomics today [56]. With the automation and software available, more than 1000 spots can be prepared and measured by one person per day. To date no satisfactory explanation on the mechanism of the MALDI process has been reported. However, functional features of this technique can be provided. The MALDI process has as its energy source the laser pulse, as opposed to the electrostatic potential in ESI, to ionize peptides. The MALDI process relies in the use of small UV- or IR-energy absorbing molecules, referred to as the matrix molecules, mixed with internal standard peptides, such as bradykinin and ACTH, which co-crystallize with the peptide sample, containing a proton donor such as trifluoroacetic acid, on a sample target often functionalized to attract hydrophilic peptides. The matrix molecule of choice for peptides is α-cyano-4-hydroxycinaminic acid. The MALDI process generates predominately singly charged ions [57].

Preparation of reliable sample spots onto the MALDI target demands a homogeneously applied peptide sample at the region upon which the laser shots will take place. Several factors affect the efficacious preparation of a MALDI target. These factors include the accurate application of matrix (pipette tip must not excessively touch the steel target but yet apply the full liquid volume), sample and re-crystallization solutions onto the steel plate spots, and the use of freshly prepared matrix and re-crystallization solutions. The laboratory temperature should be consistent at around 20 °C and the relative humidity less than 50%. Failure to comply with any of these conditions will result in inhomogeneous target preparation (i.e. sample and matrix molecules will tend to accumulate at the periphery of the raster spot), which commonly leads to loss of spectral signal response. Because the peptide and matrix solutions usually contain the volatile solvent acetonitrile, the solutions may not be left in open containers for a prolonged time as evaporation of acetonitrile may affect the matrix crystallization, and consequently the identification rate. (This factor mainly concerns automated procedures.) Contact of the organic solvents with certain plastic surfaces should be avoided as much as possible to reduce extraction of plasticizers and consequently the probability of polymer formation. In particular, presence of plasticizers in the sample may severely affect the performance of targets with hydrophobic surface for sample concentration. Consequently, efficient MALDI target preparation requires much practice under tightly controlled protocols.

A limitation of MALDI-TOF is the identification of low molecular mass proteins, which deliver few peptides, and the identification is often based on a low number of matches (for example, four matches) [47]. This mainly happens when α-cyano-4-hydroxycinnamic acid matrix species are used. In that case, it should be confirmed that the matching peptides are simultaneously the major peptides, and that theoretical and observed M_r and pI values are in a good agreement. In general, identification of small proteins by MALDI-TOF-MS is not efficient, and the application of MS/MS technologies like TOF-TOF, Qq-TOF, or LC–MS may be more advantageous.

The MALDI-TOF-based protein identification approach cannot identify multiple components of a mixture. In most cases, in 2D gels, the major component of a protein mixture is identified by MALDI-TOF from one spot. A spot often contains more than one protein and two or more additional proteins can be identified from the same spot as well, but they are usually highly homologous, for example tubulin chains, and the software cannot distinguish between the identification hits. Abundant proteins, which are often represented by multiple, strong spots, can be detected in neighboring spots as well, which represent less abundant components. Table 1 gives examples of multiple identifications from the same spot by MALDI-TOF-MS. If experimentally derived peptide masses for a given mass accuracy tolerance correspond to more than one protein, the resulting search score becomes low. This problem is further aggravated when the number of experimentally derived peptides is small and the intensity of the peptide signal is weak. Additional proteins can be identified from spots by tandem mass spectrometry if a sufficient number of peptides have been analyzed in the MS/MS mode.

In most analyses, only a subset of the predicted masses is actually observed. Such a limitation is the result of multiple contributing factors including loss of peptides during the sample handling process, as described before, ion suppression effects due to the presence of trifluoroacetic acid, surfactant residues and salts, the selective ionization of certain peptide species over others depending on their amino acid sequence, which typically exhibit poor ionization potential and small peptide masses that are beyond the optimum mass range of the MALDI-TOF instrument. In cases where a sufficient peptide coverage exists, the positive protein identification is usually not a problem. Ambiguous protein identification arises when only a very small subset of theoretically predicted masses is actually found under acceptable stringency requirements.

The unassigned peptide masses may be the result of post-translational modifications of peptides such as phosphopeptides, acetylated peptides or glycosilated peptides to name a few. Another reason may be due to chemical derivatization or modification steps taking place during the sample preparation process such as N-terminal biotinilation, N-terminal acetylation, carbamidocetylation, carboxymethylation, or methionine oxidation. These modifications are usually anticipated and are therefore incorporated into the database search parameters. The presence of artifactual peptide mass in the peptide mass fingerprint spectrum occurs in all cases. The certain number of these artifactual peptides are trypsin derived and are often referred to as trypsin auto-

Table 1
Examples of spots from which two or more protein identifications were obtained by MALDI-TOF-MS

Spot	Name	Acc. number	Full name	M_r	Matches	Score
Spot 1	HS72_MOUSE	P17156	Heat shock-related 70 kDa protein 2 (Heat shock protein 70.2)	69982	15	117
	ALBU_MOUSE	P07724	Serum albumin precursor	70700	9	53
Spot 2	GR75_MOUSE	P38647	Stress-70 protein, mitochondrial (75 kDa glucose regulated protein, GRP 75)	73767	11	65
	HS7C_MOUSE	P63017	Heat shock cognate 71 kDa protein	71055	11	64
Spot 3	GFAP_RAT	P47819	Glial fibrillary acidic protein, astrocyte (GFAP)	49969	25	184
	GFAP_MOUSE	P03995	Glial fibrillary acidic protein, astrocyte (GFAP)	49943	17	91
Spot 4	PGK1_CRIGR	P50310	Phosphoglycerate kinase 1 (EC 2.7.2.3)	44802	22	176
	PGK1_RAT	P16617	Phosphoglycerate kinase 1 (EC 2.7.2.3)	44794	20	166
	PGK1_MOUSE	P09411	Phosphoglycerate kinase 1 (EC 2.7.2.3)	44776	19	149

The proteins identified from spot 1 have similar molecular masses and their spots overlapped. The proteins found from spot 2 are highly homologous of the same species (mouse). From spot 3, GFAP was identified which could be from rat or mouse (rat brain was analyzed). Phosphoglycerate mutase found in spot 4 could be theoretically from three species (rat brain was analyzed).

digest peaks. Frequently, however, keratin peaks originating by the analyst, dye related and matrix peaks are also observed in addition. Table 2 lists the MALDI-TOF artifactual peaks observed so far in our laboratory. Artifact peaks occupy crucial mass windows otherwise available for the detection of sample related peptides. As such, potential sample peptide peaks may overlap with the artifact signals making them essentially non-usable. To partially overcome this limitation, the presence of artifactual peaks, may be used as internal calibrants, which assist in improving the mass accuracy of the sample-derived peptides. Currently mathematical algorithms are under development that make use of this approach resulting in mass accuracy values of ≤10 ppm at 1000 Da. Under this scenario, reliable protein identification can take place with smaller numbers of accurately measured peptide masses. Although never completely sufficient, maintaining high mass resolution (≥15,000) will assist in resolving sample related peptide ions from artifactual ions. Fig. 5 shows a typical spectrum generated with a MALDI-TOF-TOF-MS instrument that resulted in the identification of β-dynamin. Several artifactual peptide peaks are indicated.

Table 2
MALDI-TOF-MS artifact peptides derived from contaminant proteins and autodigests

m/z value	Origin
832.3189	Coomasie blue
842.5100	Trypsin
877.0200	Matrix
1045.5642	Trypsin
1126.5655	Trypsin
1165.5853	Keratin I
1179.6010	Keratin I
1277.7105	Keratin I
1300.5302	Keratin I
1365.6399	Keratin I
1475.7494	Keratin II
1493.7348	Keratin I
1707.7727	Keratin I
1716.8517	Keratin I
1940.9354	Trypsin
1993.9767	Keratin I
2211.1046	Trypsin
2225.1202	Trypsin
2239.1359	Trypsin
2283.1807	Trypsin
2299.1756	Trypsin
2383.9524	Cyto-keratin 1
2705.1617	Keratin 9
2825.4056	Keratin I

"Origin", protein from which the peptides were produced.

Un-anticipated enzymolysis products because of either the presence of proteases (i.e. chymotrypsin) or missed cleavage sites also affect the identification process. Miscleavage may be due to poor sample preparation or the inaccessibility of the cleavage site due to the conformation of the target protein. Also it may be due to the juxtaposition of a post-translational modification moiety to the enzymolysis recognition site that inhibits access of trypsin to the cleavage site. Trypsin miscleavage is usually anticipated and incorporated into the search criteria.

Currently, no single protein database is sufficient in characterizing all useful peptide mass fingerprint (PMF) spectra generated by the MALDI-TOF-MS instrument. Typically, an initial search begins with the SwissProt database. The resulting unidentified spectra are then subjected to a different search using a different protein database, such as one of the IPI species-specific databases. The remaining unidentified spectra are further subjected to yet another database search and the process continues until all possible spectra are characterized. In addition to the manual intervention required to set-up the search cycle along with all the search constraints (mass tolerance, fixed and variable modifications, species indication, etc.) the search results in basic differences between all available protein databases. Thus, the number of accession number conventions will be equal to the number of different protein databases used for the iterative search process. In addition, certain hypothetical proteins found with one database may be known, well defined proteins with a dif-

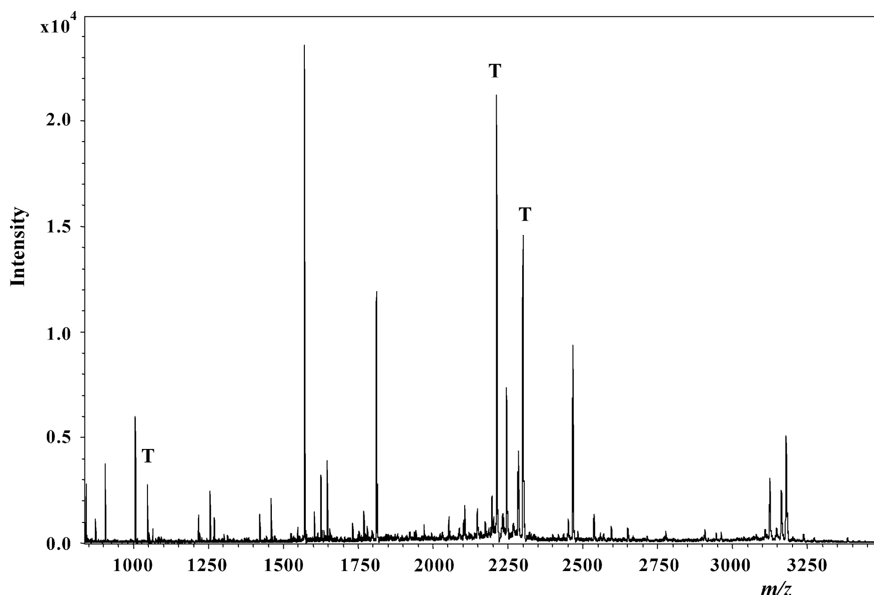

Fig. 5. A typical peptide mass fingerprint (PMF) spectral response. Each peak corresponds to a peptide isotopic cluster. Peaks designated with "T" correspond to trypsin autodigest peptides which are always present in protein samples subjected to trypsin digestion.

ferent database. Also, the actual number of different proteins found may be less than what was originally calculated when protein redundancies and multiple name conventions for a given protein are factored in. The manually intensive process can be greatly simplified with the availability of a single comprehensive database with minimum redundancies and a single accession number convention, protein nomenclature and more accurate and concise protein description.

Typical identification rates, which we obtain in our analyses, at a mass tolerance of 25 ppm are about 55–65% for a well prepared MALDI target when a well balanced mixture of well visible and faint spots is analyzed. The identification rates increase to about 65–75% at a mass tolerance of 50 ppm and higher than 75% at a mass tolerance of 100 ppm. However, mass tolerance threshold values higher than 25 ppm result in unreliable protein identifications with in low or borderline statistical scores. False positive hits may be obtained because large proteins deliver theoretical peptides with masses close to the measured and this factor may result in a random identity assignment. Frequent external calibration of the instrument and mass correction with the use of standard peptides are essential for a good performance and an identification rate of about 70%. Because identification is based on the comparison of mass to charge ratios m/z, a high mass accuracy is essential for a confident identity assignment.

5.2. Electrospray ionization

In the ESI process under acidic conditions, the peptide protonation sites are the primary N-terminal amine moieties along with the basic side groups of lysine, arginine, and histidine resulting in multiply charged species. However, the ionization is never 100% efficient because the solution phase chemistry of the various peptides varies in accordance to their physico-chemical properties, including pK_a value, polarity, the hydrophobic or hydrophilic index, and ionization potential. This variability in solution phase chemistry has a direct effect to the efficiency of ionization in the gas phase. In addition to the variability in ionization potential between the various peptides, the efficiency of ionization is directly impacted by the concentration and type of peptides infused into the API source. The higher the peptide concentration, the lower the capacity to sufficiently ionize all available peptides due to depletion of all available protons and due to upper mass density constraints to the dynamic range of the mass spectrometer. A possible solution in enhancing the ionization potential of the less polar and more hydrophobic peptides is to add ionizable functional groups via chemical derivatization during the sample preparation process. Consideration must be given, however, to ensure that the derivatization will not cause any chemical decomposition of any of the peptides and that it will exhibit complete yields in the functionalized peptides.

When performing positive ion ESI, peptides form adducts with Na^+, Li^+, NH_4^+, or K^+. The formation depends on the type and purity of the sample solvent used. Fig. 6 illustrates some of the more commonly observed adduct species in the positive ion mode. The commonality of adduct species formation in ESI results in reduction of the sensitivity of quantitative assays using tandem mass spectrometry, since the signal for each peptide will be distributed among more than one species instead of being concentrated into one precursor ion that is selected for MS–MS. Also, the relative abundance of the protonated molecule (or other cationized species) used for quantitative analysis will vary between assays depending upon the amount of sodium or potassium ions in each sample and in the mobile phase or sample solvent system.

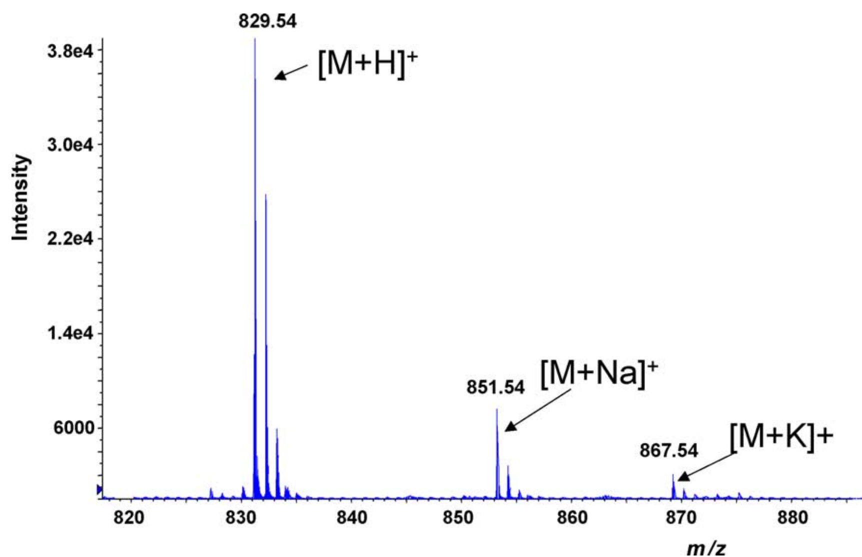

Fig. 6. Adduct species observed for a synthetic peptide used to externally calibrate the ESI-Qq-TOF instrument.

In a ESI process, generation of multiply charged species occurs. The mass of the peptide ion increases so does the number of charges it can accommodate. Even though this ESI feature allows the detection of large biomolecules, it leads to a dilution effect of the monoisotopic mass, upon which the peptide mass assignment is based, and consequently to a reduction in the sensitivity in detecting the higher mass ions. Also, at the higher charge states, higher mass resolution is needed to resolve the individual peaks within the isotopic cluster. The lack of sufficient mass resolution will affect the accurate mass assignment to the monoisotopic peak, which is used for the subsequent product ion in MS–MS experiments. Fig. 7 illustrates the reduction in mass resolution with increasing peptide charge state observed with the Q-STAR XL quadrupole time-of-flight (Qq-TOF) instrument that has been tuned and calibrated to exhibit a mass resolution value of 10,000 at 800 Da.

As in the case of the PMF spectrum generated with MALDI-TOF based systems, artifactual peptide masses also occur in the TOF-MS spectral responses generated with the ESI-Qq-TOF based instruments. These peptidic spectral artifacts are trypsin autodigest peaks originating from the trypsin digestion process or keratins originating from poor sample handling procedures. Table 3 lists the artifactual peaks observed in our lab. As in the MALDI-TOF-MS analysis situation, these artifacts occupy potentially valuable mass windows, which would otherwise be available for the detection of sample related peptide signals. To partially compensate for this pitfall, the artifactual peptide masses can potentially be used as internal calibrants to augment the mass accuracy of the analysis process. However, there remains an instrument specific limit as to how much the mass accuracy can improve. For example, a Qq-TOF-MS based system with mass resolution of 10,000 at 800 Da can provide a mass accuracy of 10 ppm at best.

5.3. Liquid chromatography–mass spectrometry

LC–MS is a complementary technology to 2D gels/MALDI-TOF-MS for the analysis of certain protein classes [54,58,59]. A limitation of the approach is the incomplete chromatographic separation of peptides, which introduces multiple peptides originating from multiple proteins per time unit into the mass spectrometer. The ionization efficiency (ion suppression and selective ionization effects) scanning duty cycle, ion capacity threshold, and ion transmittance characteristics for a given mass analyzer design (i.e. quadrupole ion-trap and quadrupole time-of-flight) add significant constraints in the efficient analysis of all available

Table 3
ESI-Qq-TOF-MS artifact peptides derived from contaminant proteins and autodigests

m/z Value	Charge state	Origin
421.76	+2	Trypsin
453.90	+2	Trypsin
599.70	+2	Keratin I
632.40	+2	Keratin I
707.05	+3	Trypsin
730.90	+3	Keratin I
732.00	+3	Trypsin
737.69	+3	Trypsin
745.00	+3	Trypsin
761.70	+3	Trypsin
767.10	+3	Trypsin
772.07	+3	Trypsin
789.80	+3	Trypsin
795.10	+4	Trypsin
883.90	+2	Keratin I
899.00	+2	Keratin I
1106.07	+2	Trypsin
1142.20	+2	Trypsin
1150.08	+2	Trypsin

"Origin", protein from which the peptides were produced.

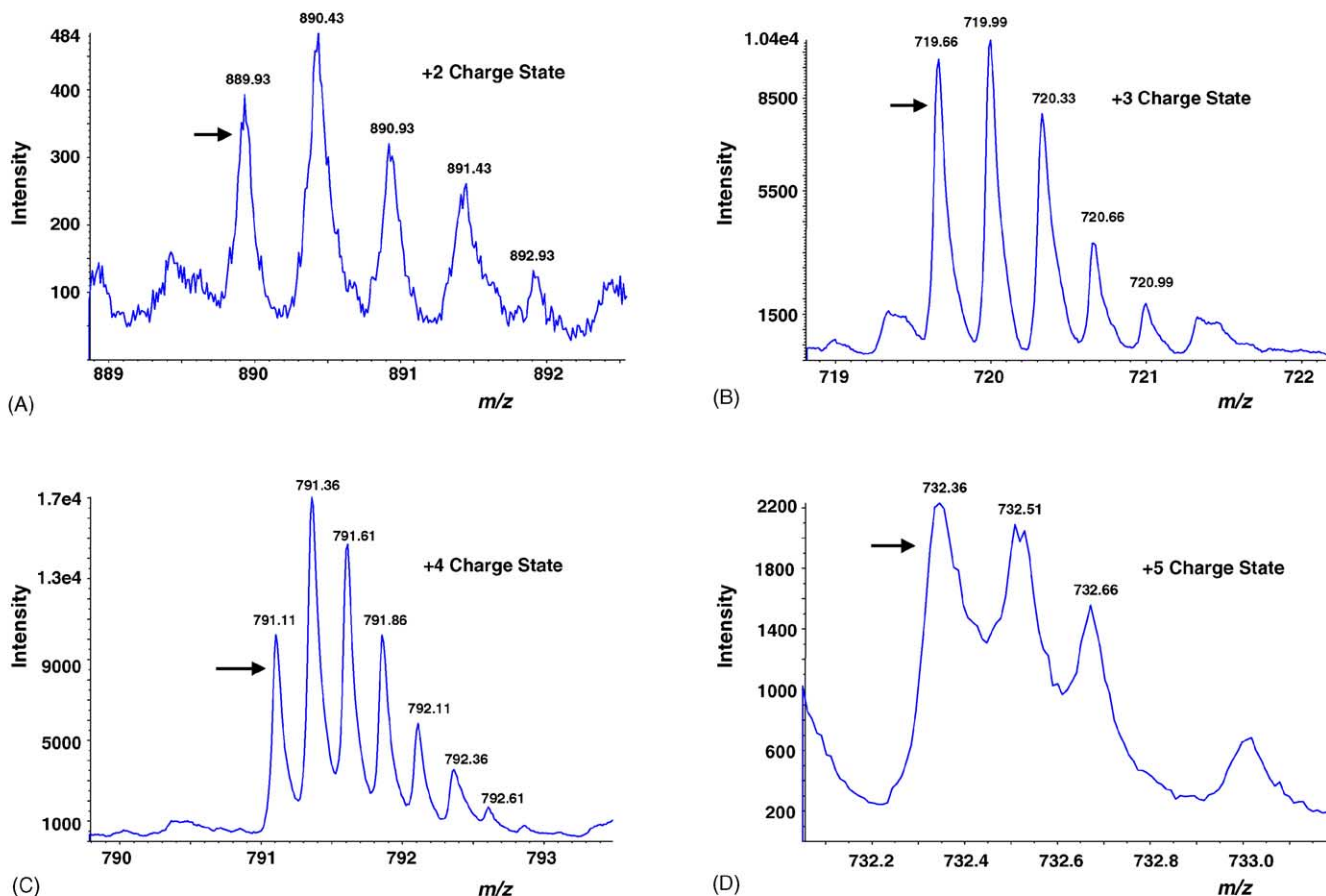

Fig. 7. The monoisotopic mass becomes less resolved with increasing charge state of a given peptide for the ESI-Qq-TOF type instrument exhibiting a mass resolution value of 10,000 at *m/z* 800 type. A, B, C and D, isotope clusters occurring at +2, +3, +4 and +5 charge states, respectively. The monoisotopic mass becomes increasingly mixed with the adjacent peaks. The arrowheads indicate the monoisotopic mass.

peptide ions. Consequently, the identification of the major peptide ions will tend to negate the effective analysis of the minor peptides, which will then reduce the identification rate of the analysis. In addition, the sensitivity, mass resolution and consequently mass accuracy of the MS analysis for the remaining peptide ions is compromised. As a rule of thumb at least 15 fmol for a single peptide is necessary to yield a usable MS–MS spectrum. To illustrate these limitations, a representative analytical problem is described. To conduct an LC–MS experiment, one typically begins with about 0.7–1.5 μg digests originating from a biological sample (e.g. plasma, brain). The typical sample volume injected to the LC–MS system is 1.0–2.0 μl. A linear gradient mobile phase is used to maximize the chromatographic separation over a period of 3–5 h [60]. At the completion of this run more than 5000 tandem MS–MS spectra are generated and the search results in the identification of over 1000 proteins. However, the significant redundancies in identification due to multiple peptides originating from a single protein and ineffective analysis of the minor peptide ions, approximately only 10% of the total protein hits correspond to different proteins. An added difficulty is that these less than ideal results are generated at the expense of extensive computing resources (processing time, data storage and maintenance and data archiving).

In order to overcome these limitations of the 1D LC–MS approach, a prior chromatographic fractionation is performed to reduce the number of peptides injected in a single run. Such an approach provides for a more efficient chromatographic separation of peptides and reduces the complications described above in regard to MS analyzer behavior. Two-dimensional liquid chromatography (usually refreed to as multi-dimensional protein identification technology, MudPIT) provides a wide diversity of chromatographic principles to be combined to suit the particular physicochemical properties of the peptides under investigation. For example, membrane bound proteins may be best handled with the combination of ion-exchange (weak anion exchange, strong cation exchange) and hydrophobic interaction (reverse phase such as C18, C8 and C4 or poly-methyl aspartamide) chromatographic chemistries. The only basic restraint in combining these methods is for the chosen chemistries to be orthogonal to each other. The number of aliquots that may be fractionated from the one-dimensional chromatography experiment ranges from 25 to 70. Each aliquot is then subjected to an on-line nanobore LC/MS analysis run. The

principle advantage is that the total redundancy decreases substantially and the number of different proteins identified is substantially larger. For example, of the 1000 proteins that can be found from the MudPIT approach, 600–700 different proteins can reliably be identified after redundancies are factored in. However, the overall experimental run time ranges from 20 to 70 h, and combined with the processing time, the time period increases to weeks. Therefore, the high-throughput analysis that is essential to proteomic studies is not realized yet from the MudPIT approach.

5.4. Tandem mass spectrometry

The two principal means of generating tandem mass (MS–MS) spectra is via low-energy collision induced dissociation (CID) with N_2 as the collision gas taking place in the fragmentation quadrupole chamber (Q2) of a Qq-TOF-based instrument, and post source decay taking (PSD) place in MALDI-TOF instrument designs. The correct execution of tandem mass spectrometry will normally result in protein identification with even one peptide precursor ion. Also, they serve to provide added confirmation to the MS-based protein identifications. Another critical advantage available to tandem mass spectra is that they allow the sequencing of amino acids for the un-assigned peptides or peptides belonging to hypothetical proteins. In such cases, the amino acid sequences are subjected to a BLAST search for protein identification purposes.

As a rule of thumb, in the PSD tandem mass spectra, the precursor ion intensity occurring for the PMF spectrum must be over 1000 otherwise the generation of tandem spectra may be unsuccessful. Fig. 8 illustrates an un-identified PMF spectrum along with poorly executed PSD spectra due to weak precursor ion intensities. This poses a critical constraint to the utility of the MALDI-TOF-MS-MS technique in improving the protein identification process for PMF spectra corresponding to faint 2D gel spots or low concentration protein samples.

On the other hand, CID generally provides better quality tandem mass spectrometric responses in terms of higher signal-to-noise ratios and higher abundances in diagnostic fragment ions. These features of the CID generated tandem spectra allow for the de novo sequencing of peptide ions. However, several limitations apply to the effective amino acid sequencing using CID tandem mass spectral data. These include: (i) The CID spectra of peptides generally result in a limited number of fragment ions. For example, the key sequence-specific b-type ions normally occur at low abundance values and thus may not be usable. (ii) The CID spectra may also result in non-sequence specific ion types such as the immonium ions that cannot be used for sequence determination. (iii) The leucine and isoleucine cannot be differentiated with CID tandem spectra due to their identical *m/z* values. (iv) Glutamine and lysine differ by 0.036 Da and phenylalanine and oxidized methionine differ by only 0.033 Da. The differentiation of these residues can only be accomplished with a well tuned and calibrated Qq-TOF based MS system. Quadrupole ion trap and triple quadrupole MS instruments cannot differentiate these pairs. (v) It is very uncommon to cause fragmentation of all peptide bonds, which is necessary to delineate the complete peptide sequence. Practically, only a partial sequence is provided by the interpretation of the CID mass spectra. (vi) Certain amino acid pairs have the same *m/z* value of a single mass and

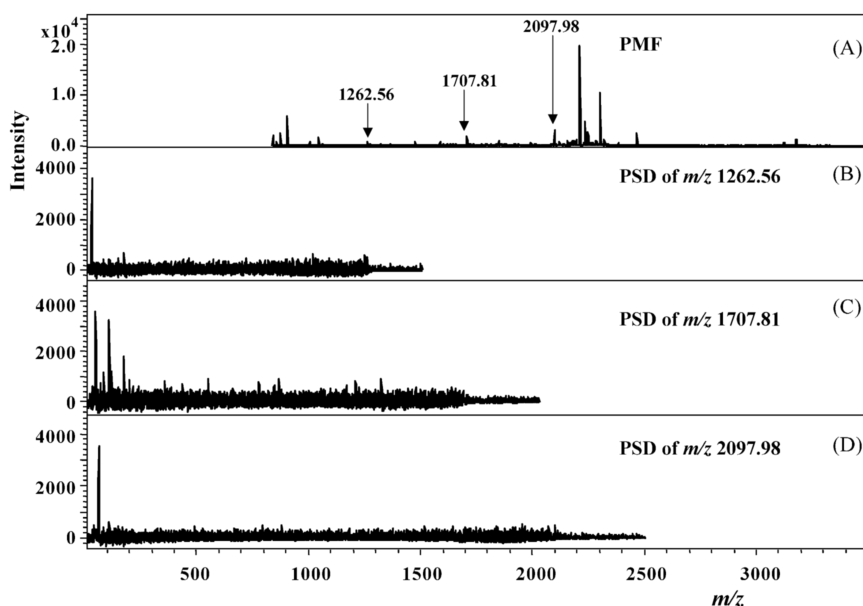

Fig. 8. The post source decay (PSD) experiment to generate tandem mass spectra with MALDI-TOF based instruments requires precursor ions with signal-to-noise values of over 1000:1. Consequently, weak PMF signals result in unsuccessful PSD runs. A, PMF spectrum produced with MALDI-TOF. The arrowheads indicate PMF peaks chosen as the precursor ions for PSD experiments. (B, C, D) PSD spectra from the parent peaks indicated which exhibit poor signal-to-noise values and were unsuccessful in the identification of a protein. At the left part of the figures, low m/z signals of no diagnostic value are observed.

thus further complicate the sequencing process. These cases are as follows (in *m/z* values): Ser + Val = 186.100 versus Trp = 186.079; Gly + Val = 156.090 versus Arg = 156.101; Ala + Gly = 128.059 versus Gln = 128.059 versus Lys = 128.095; Ala + Asp = Glu + Gly = 186.064 versus Trp = 186.079; Gly + Gly = 114.043 versus Asn = 114.043. All these limitations can impose constraints to the data acquisition process in that unit resolution must be applied when acquiring CID tandem mass spectra with the Qq-TOF instrument which in-turn requires sufficient peptide concentration levels in order to ensure usable signal-to-noise values. When using a low-resolution setting during the acquisition process, the amino acid sequencing will be unreliable at best. But even under unit-resolution acquisition conditions, the peptide sequencing results are never 100% accurate, and require an experienced mass spectrometry scientist for the interpretation of the tandem spectra.

5.5. *Instrument-related limitations*

The smooth operation of a mass spectrometry based proteomics facility requires the diligent and consistent oversight of multiple parameters and tasks. These primarily include the following: (i) All instruments need to be well tuned and calibrated on a frequent basis. One ancillary requirement is for laboratory to be maintained under a consistent temperature and humidity environment (typically 20–22 °C and relative humidity below 50%). This is an important factor because instrument power supplies and electronic circuit boards control the behavior of the ion optic, mass analyzer and detection components of mass spectrometers. Fluctuation in temperature may affect optimal tuning and calibration settings thus compromising the analysis results. Elevated temperature environments also cause excessive heating of the compressor motors of the nitrogen generators. Also, high humidity environments may cause premature corrosion to electronic components and may cause premature nitrogen generator problems. An additional critical environmental consideration is a dust free air environment. Excessive dust will cause obstruction to the ventilation ducts and fans of computer and instrument devices causing them to overheat and cause premature operation failures. (ii) MS instrumentation require optimum vacuum conditions that directly depend on properly maintained rotary-vein and turbo-molecular pump. Thus, frequent inspection of oil levels and quality is essential. The current turbo-molecular pumps are air-cooled and thus operate better under lower temperature environments. (iii) The nitrogen source required by N_2-laser based MALDI-TOF-MS instrument is grade 5 (99.9995%) pure. However, this also requires that the nitrogen gas lines be airtight and always well purged. As such, when tanks are exchanged, the air-lines must be purged sufficiently prior to instrument use. (iv) The nano-HPLC pumps are often prone to air bubble occurrence in one or more of their components. When this occurs, it may take several hours to days to expunge. As such, the nano-HPLC pump systems must be purged and primed constantly. (v) The sample handling robotic devices that conduct gel-picking and sample application to MALDI targets also require often operator intervention during the sample handling process. Ideally, sample handling robotic devices should perform un-attended operation in order to allow operators to pursue other tasks as required by a high-throughput laboratory environment. (vi) MS instrumentation also requires frequent cleaning of ionization source ion-optical components depending on the frequency of use. Such intervention is warranted when high background noise is observed in spectral responses. (vii) All instrumentation, including the sample handling robotic devices, require a well maintained and updated log-book to include maintenance performed, troubleshooting operations, standard operating procedures and personnel training. Good documentation allows for the quick and efficient diagnosis and troubleshooting of instrument failures. Good documentation also supports their consistent operation. (viii) The large amounts of data acquired on a daily basis in a high-throughput environment requires frequent and consistent data storage and archiving. Such storage should take place on multiple hard disks and tapes and should be adequately protected. (ix) The protein databases, i.e. Swiss Prot, NCBI, IPI, MSDB, etc., require updating on a weekly basis to ensure that the analysis results are completely updated. Oftentimes, a meta-analysis may be performed whereby the previously acquired spectral data are resubmitted for search using the updated databases. However, these operations are time-prohibitive given the magnitude of the data size. Bioinformatics software programs should be designed to allow the frequent updating of all spectra. As a consequence, a functional MS-based proteomics facility requires dedicated and experienced personnel committed to executing the above operations.

6. Selection of the proteomic method

The selection of the method strongly depends on the sample to be analyzed and the goal of the study. As a rule of thumb, the first choice of a proteomics analysis is the 2D gel/MALDI-TOF-MS approach. Because certain protein classes can not be detected by these technologies, employment of two-dimensional LC–MS or tandem MS should follow. In general, it seems that a larger number of different proteins are identified in a LC–MS run compared to the 2D gel/MALDI-TOF-MS approach. This can be partially due to the limitations of the 2D gel technology and that many identifications from the ion trap approach depend on the sequencing of one peptide and this may not be sufficient for a confident identity assignment. When the ion trap technology is used, at least three peptides should be analyzed in the MS/MS mode for an unambiguous identification.

Besides the application of the principal proteomics technologies, a preliminary work effort should be performed for the enrichment of the low-abundance or the depletion of high-abundance proteins. It becomes obvious then, given the wide

array of physicochemical techniques, that the experimental design regarding the sample preparation strategy to be used should be guided by various constraints such as the downstream analysis process (i.e. if 2D gel electrophoresis or direct MS analysis is to be used), cost considerations, throughput and available facilities. The number of fractions to be analyzed may escalate the resulting number of analysis samples to an impractical degree, once available instrumentations, timelines, number of personnel and cost estimates are considered. Chemical integrity and analysis reproducibility are directly affected by the complexity of the sample analysis protocol and if the number of sample preparation steps increases so does the probability of having sample loss manifested as non-specific binding to plastic surfaces such as tubes, pipette tips, glass vials, deep well plates and solid-phase enrichment and cleaning materials.

7. Concluding remarks

Proteomics technologies show certain limitations which are mainly related to the capability of the methods to detect difficult protein classes, like low-abundance, hydrophobic and basic proteins. Compared to MS, 2D electrophoresis has not been developed to the same extent. Thus, whereas a protein amount at the amol range is sufficient for identification, and highly sophisticated software can reduce noise, extract spectra, correct masses, identify proteins with a high confidence, and new or updated instruments enter the market in always shorter time periods, the 2D gel technology cannot progress correspondingly. Although certain advancements and automation have been introduced, the detection limitations in 2D gels of the majority of the low-abundance, hydrophobic, acidic, basic, small and large proteins have not been resolved. With the use of LC–MS technologies some of these limitations can be overcome, however, further development for the establishment of a robust proteomics platform is required.

The wide application of proteomics technologies in the investigation of biological systems started a few years ago. In most proteomes, up to now, mainly partial mappings have been performed. Thus, about 1000–1500 different gene products have been found in human and mouse brain. In addition, a large number of changes at the levels and the modifications of mainly abundant brain proteins have been detected by proteomics technologies, and certain of these proteins are potential drug targets and diagnostic markers. However, neuroproteomics has probably not yet resulted in the detection of the majority of the important brain proteins. This task may be difficult, and demands an increase in sensitivity and performance of the analytical techniques and in particular the detection and visualization of certain protein classes, like low-abundance and hydrophobic proteins, which on account of limitations of the current technologies can not be efficiently detected. The proteomics shows severe or less significant limitations, depending on the goal, the most common of which are discussed here as we experienced them in our laboratory. The drawbacks certainly affect our research today but they drive the proteomics science to develop further and faster and to contribute to a detailed understanding of biological systems and the control and diagnosis of human disorders.

References

[1] M. Fountoulakis, Amino Acids 21 (2001) 363.
[2] M. Fountoulakis, Mass Spectrom. Rev. 23 (2004) 231.
[3] G. Lubec, K. Krapfenbauer, M. Fountoulakis, Progress Neurobiol. 69 (2003) 193.
[4] M. Fountoulakis, J. Schlaeger, Electrophoresis 24 (2003) 260.
[5] M. Fountoulakis, J.-F. Juranville, M. Dierssen, G. Lubec, Proteomics 2 (2002) 1547.
[6] K. Krapfenbauer, M. Fountoulakis, G. Lubec, Electrophoresis 24 (2003) 1847.
[7] J.-H. Shin, K. Krapfenbauer, G. Lubec, Electrophoresis, in press.
[8] P.A. Nielsen, J.V. Olsen, A.V. Podtelejnikov, J.R. Andersen, M. Mann, J.R. Wisniewski, Mol. Cell. Proteomics 4 (2005) 404.
[9] B. Bjellqvist, K. Ek, P.G. Righetti, E. Gianazza, A. Görg, R. Westermeier, W. Postel, J. Biochem. Biophys. Methods 6 (1982) 317.
[10] A. Görg, C. Obermaier, G. Boguth, A. Harder, B. Scheibe, R. Wildgruber, W. Weiss, Electrophoresis 21 (2000) 1037.
[11] M. Fountoulakis, E. Schuller, R. Hardmeier, P. Berndt, G. Lubec, Electrophoresis 20 (1999) 3572.
[12] M. Fountoulakis, R. Hardmeier, H. Höger, G. Lubec, Exp. Neurol. 167 (2001) 86.
[13] K.K. Jain, Methods Enzymol. 356 (2002) 157.
[14] M. Fountoulakis, M.-F. Takács, B. Takács, J. Chromatogr. A 833 (1999) 157.
[15] P.G. Righetti, A. Castagna, B. Herbert, F. Reymond, J.S. Rossier, Proteomics 3 (2003) 1397.
[16] P.G. Righetti, A. Castagna, P. Antonioli, E. Boschetti, Electrophoresis 26 (2005) 297.
[17] B.J. Takács, in: R.A. Meyers (Ed.), Encyclopedia of Analytical Chemistry, Wiley, Chichester, 2001, pp. 5955–5970.
[18] M. Fountoulakis, B. Takács, Protein Express. Purif. 14 (1998) 113.
[19] K. Karlsson, N. Cairns, G. Lubec, M. Fountoulakis, Electrophoresis 20 (1999) 2970.
[20] K. Krapfenbauer, M. Berger, G. Lubec, M. Fountoulakis, Eur. J. Biochem. 268 (2001) 3532.
[21] M. Fountoulakis, J.-F. Juranville, Anal. Biochem. 313 (2003) 267.
[22] M. Fountoulakis, J.-F. Juranville, G. Tsangaris, L. Suter, Amino Acids 26 (2004) 27.
[23] P.G. Righetti, A. Castagna, B. Herbert, Anal. Chem. 73 (2001) 320A.
[24] N.L. Anderson, N.G. Anderson, Mol. Cell. Proteomics 1 (2002) 845.
[25] J.W. Zolg, H. Langen, Mol. Cell. Proteomics 3 (2004) 345.
[26] R. Pieper, C.L. Gatlin, A.J. Makusky, P.S. Russo, C.R. Schatz, S.S. Miller, Q. Su, A.M. McGrath, M.A. Estock, P.P. Parmar, M. Zhao, S.T. Huang, J. Zhou, F. Wang, R. Esquer-Blasco, N.L. Anderson, J. Taylor, S. Steiner, Proteomics 3 (2003) 1345.
[27] L. Jiang, K. Lindpaintner, H.-F. Li, N.F. Gu, H. Langen, L. He, M. Fountoulakis, Amino Acids 25 (2003) 49.
[28] M. Fountoulakis, J.-F. Juranville, L. Jiang, D. Avila, D. Röder, P. Jakob, P. Berndt, S. Evers, H. Langen, Amino Acids 27 (2004) 249.
[29] P. Lescuyer, D.F. Hochstrasser, J.C. Sanchez, Electrophoresis 25 (2004) 1125.
[30] E. Marengo, E. Robotti, F. Antonucci, D. Cecconi, N. Campostrini, P.G. Righetti, Proteomics 5 (2005) 654.
[31] M. Unlu, M.E. Morgan, J.S. Minden, Electrophoresis 18 (1997) 2071.
[32] S.P. Gygi, B. Rist, S.A.T.F. Gerber, M.H. Gelb, R. Aebersold, Nat. Biotechnol. 10 (1999) 994.

[33] S.P. Gygi, B. Rist, T.J. Griffin, J. Eng, R. Aebersold, J. Proteome Res. 1 (2002) 47.
[34] Y. Lu, P. Bottari, F. Turecek, R. Aebersold, M.H. Gelb, Anal. Chem. 76 (2004) 4104.
[35] M. Fountoulakis, P. Berndt, U.A. Boelsterli, F. Crameri, M. Winter, S. Albertini, L. Suter, Electrophoresis 21 (2000) 2148.
[36] T. Rabilloud, T. Blisnick, M. Heller, S. Lunche, R. Aebersold, J. Lunardi, C. Braun-Breton, Electrophoresis 20 (1999) 3603.
[37] V. Santoni, M. Molloy, T. Rabilloud, Electrophoresis 21 (2000) 1054.
[38] M. Fountoulakis, R. Gasser, Amino Acids 24 (2003) 19.
[39] S. Konig, O. Schmidt, K. Rose, S. Thanos, M. Besselmann, M. Zeller, Electrophoresis 24 (2003) 751.
[40] L. Carboni, C. Piubelli, P.G. Righetti, B. Jansson, E. Domenici, Electrophoresis 23 (2002) 4132.
[41] J. Hartinger, K. Stenius, D. Hogemann, R. Jahn, Anal. Biochem. 240 (1996) 126.
[42] J. Kyte, R.F. Doolittle, J. Mol. Biol. 157 (1982) 105.
[43] P.K. Sinha, E. Köttgen, M. Stoffler-Meilicke, E. Gianazza, P.G. Righetti, J. Biochem. Biophys. Methods 20 (1990) 345.
[44] P.G. Righetti, A. Bossi, A. Görg, C. Obermaier, G. Boguth, J. Biochem. Biophys. Methods 31 (1996) 81.
[45] A. Görg, W. Weiss, M.J. Dunn, Proteomics 4 (2004) 3665.
[46] M. Fountoulakis, B. Takács, H. Langen, Electrophoresis 19 (1998) 761.
[47] M. Fountoulakis, J.-F. Juranville, D. Roeder, S. Evers, P. Berndt, H. Langen, Electrophoresis 19 (1998) 1819.
[48] H. Schagger, G. von Jagow, Anal. Biochem. 166 (1987) 368.
[49] M. Bruschi, L. Musante, G. Candiano, B. Herbert, F. Antonucci, P.G. Righetti, Proteomics 3 (2003) 821.
[50] J. McCarthy, F. Hopwood, D. Oxley, M. Laver, A. Castagna, P.G. Righetti, K. Williams, B. Herbert, J. Proteome Res. 2 (2003) 239.
[51] R. Aebersold, M. Mann, Nature 422 (2003) 198.
[52] D. Fenyo, J. Qin, B.T. Chait, Electrophoresis 19 (1998) 998.
[53] J. Godovac-Zimmermann, L.R. Brown, Mass Spectrom. Rev. 20 (2001) 1.
[54] K.G. Standing, Curr. Opin. Struct. Biol. 13 (2003) 595.
[55] W.J. Henzel, T.M. Billeci, J.T. Stults, S.C. Wong, C. Grimley, C. Watanabe, Proc. Natl. Acad. Sci. U.S.A. 90 (1993) 5011.
[56] H.-W. Lahm, H. Langen, Electrophoresis 21 (2000) 2105.
[57] M. Karas, M. Glückmann, J. Schäfer, J. Mass Spectrom. 35 (2000) 1.
[58] C.L. Gatlin, J.K. Eng, S.T. Cross, J.C. Detter, J.R. Yates III, Anal. Chem. 72 (2000) 757.
[59] G. Neubauer, M. Mann, Anal. Chem. 71 (1999) 235.
[60] S.D. Garbis, A. Melse-Boonstra, C.E. West, R.B. van Breemen, Anal. Chem. 73 (2001) 5358.