the case following gross microbial contamination of columns of both anion and cation exchangers.

A key element of process validation is the chemical stability of the ion exchanger to the cleaning/sanitizing regent systems. We have developed protocols for monitoring hydrolysis of functional groups, referred to as ligand leakage, and also matrix dissolution, in order to address these concerns.

These process validation studies are typically conducted in scale-down mode, and confirmatory checks made subsequently at process scale.

Process Scale Ion Exchange Chromatography of Proteins

Having defined the feedstock and mobile phases, selected the ion exchanger and selected a batch or column mode of operation, the chromatographer should find himself or herself in a position to scale-up the process.

It is difficult to predict cost information on process scale ion exchange separations since much depends on the upstream and downstream process requirements and reusability of the ion exchanger, etc. We have reported a cost estimate for single usage of 25 kg of DE52 in batch and column operation (see Ganetsos and Barker), but this is exemplary only. Unfortunately information of this type, while in existence, is proprietary and therefore withheld.

Industry has carried out large scale ion exchange chromatography of proteins for the last few decades and with the established principles described above, there is no reason to assume that things will change significantly over the short to medium term. In the longer term, developments in fluidized/expanded beds and membrane adsorbers may offer new opportunities in this area of chromatography.

See also: **II/Affinity Separation:** Rational Design, Synthesis and Evaluation: Affinity Ligands. **III/Proteins:** Centrifugation; Electrophoresis; High-Speed Countercurrent Chromatography.

Further Reading

Ganetsos G and Barker PE (eds) (1993) *Preparative and Production Scale Chromatography*. New York: Marcel Dekker.

Levison PR, Mumford C, Streater M, Brandt-Nielsen A, Pathirana ND and Badger SE (1997) Performance comparison of low-pressure ion-exchange chromatography media for protein separation. *Journal of Chromatography A* 760: 151.

Shillenn JK (ed.) (1996) *Validation Practices for Biotechnology Products*. West Conshohocken: ASTM.

Sofer GK and Nystrom L-E (1991) *Process Chromatography. A Guide to Validation*. London: Academic Press.

Stryer L (1981) *Biochemistry*, 2nd edn. San Francisco: Freeman.

Subramanian G (ed.) (1991) *Process-scale Liquid Chromatography*. Weinheim: VCH.

Verrall MS (ed.) (1996) *Downstream Processing of Natural Products*. Chichester: John Wiley.

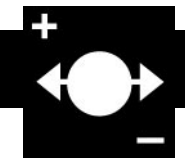
Metalloprotiens: Chromatography

See **III / METALLOPROTEINS: CHROMATOGRAPHY**

Thin-Layer (Planar) Chromatography

See **III / PEPTIDES AND PROTEINS: Thin-Layer (Planar) Chromatography**

PROTEOMICS: ELECTROPHORESIS



M. J. Dunn, National Heart and Lung Institute, Imperial College of Science, Technology and Medicine, Harefield Hospital, Middlesex, UK

Introduction

The first complete genome sequence, that of *Haemophilus influenzae*, was published in 1995. Intense effort over the last three years has resulted in the completion of the genomes for a further 12 micro-organisms

ranging in complexity from *Mycoplasma genitalium*, with a genome size of only 0.58 Mb encoding less than 500 proteins, to *Escherichia coli*, with a genome size of 4.6 Mb encoding more than 4000 proteins (**Table 1**). The complexity of the eukaryotic genomes has resulted in slower progress, with only one organism, the yeast *Saccharomyces cerevisiae*, having been completed (Table 1). However, significant progress is being made for a variety of other species, with the estimated date for the completion of the human genome currently being 2001 (Table 1).

The vast amount of information being generated by the various genome sequencing programmes has the potential to contribute significantly to our understanding of how an organism functions and its evolutionary relationships with other life forms. However, it has already become clear that genomics alone will not provide all of the answers. For those organisms whose genomes have been completed, typically around 30% of the genes can be assigned definite functions with up to a further 30% being attributed functions on the basis of homology with other genes of known function. The remaining 40% of the structural genes can often not even be attributed putative functions. A further limitation of the genomic approach is that it does not provide any insights into the ways an organism may modify its pattern of gene expression in response to different conditions.

Table 1 Some organisms whose genomes have been completely sequenced and others which are the subject of active genome sequencing programmes

Organism	*Size (Mb)*	*ORFs*	*Year completed*
Microorganisms			
Mycoplasma genitalium	0.58	470	1995
Ureaplasma urealyticum	0.75	640	
Mycoplasma pneumoniae	0.81	679	1996
Treponema pallidum	1.14	1000	
Borrelia burgdorferi	1.44	843	1997
Aquifex aeolicus	1.50	1512	1998
Helicobacter pylori	1.66	1590	1997
Methanococcus jannaschii	1.66	1738	1996
M. thermoautotrophicum	1.75	1855	1997
Haemophilus influenzae	1.83	1743	1995
Streptococcus pyogenes	1.98	1900	
Archaeoglobus fulgidis	2.18	2436	1997
Nisseria gonorrhoreae	2.2	2100	
Pyrobaculum aerophilum	2.22	1900	
Synechocystis PCC6803	3.57	3168	1996
Bacillus subtilis	4.20	4100	1997
Mycobacterium tuberculosis	4.41	3924	1998
Escherichia coli	4.60	4288	1997
Eukaryotes			
Saccharomyces cerevisiae	13.0	5885	1996
Dictyostelium doscoideum	70	12500	
Arabidopsis thalania	70	14000	
Caenorhabditis elegans	80	17800	
Drosophila melanogaster	170	30000	
Homo sapiens	2900	50000	

ORFs = open reading frames.

Analysis of Gene Expression

These problems can only be solved by direct investigation of gene expression, which can be achieved at either the level of messenger RNA (mRNA) or protein. A variety of techniques such as cDNA microarrays and serial analysis of gene expression (SAGE) make it possible to undertake mass screening of mRNA expression and establish which particular mRNAs are expressed in an organism under any given condition. However, recent studies have highlighted an important limitation of this approach in that *there is not always a direct correlation between RNA abundance and the amount of the corresponding functional protein within the cell.*

A further major limitation of studies at the level of mRNA is that they are unable to provide any information of processes of co- and post-translational modification of proteins. The modification of proteins by processes such as phosphorylation, glycosylation, sulfation, hydroxylation, N-methylation, acylation, prenylation and N-myristoylation, can result in significant modulation of their functional properties. Knowledge of these processes is therefore important for a complete understanding of gene expression.

Proteome Analysis

The realization that these problems can only be addressed through studies at the level of protein expression has resulted in increasing interest in the area which has become known as 'proteome analysis'. The term 'proteome' was first coined by a collaborative team of scientists at Macquarie and Sydney Universities in 1995 and is defined as the **prote**in complement of the gen**ome** of an organism. Increasing genomic complexity together with the potential for co- and post-translational modifications make proteome analysis a difficult task for higher organisms. As a consequence, active proteome programmes are currently restricted to some of the simpler organisms such as mollicutes (*M. genitalium*, *Sprioplasma melliferum*), prokaryotes (*E. coli, Chlamydia trachomatis*) and yeast (*S. cerevisiae*).

The complexity of eukaryotic proteomes has resulted in the term 'proteomics' or 'proteome analysis' being used in a narrower context in which it is used to characterize patterns of protein (and thereby gene) expression in particular cell type and tissues. This

approach can provide new insights into a variety of biological processes such as development, apoptosis and the cell cycle and add to our knowledge of the mechanisms that control gene expression. There is also considerable interest in applying proteomics to the study of diseases, where it promises further understanding of these processes at the molecular level and may lead to the discovery of new diagnostic markers and novel therapeutic targets. The pharmaceutical industry is also expressing considerable interest in the potential of proteomics in the process of drug discovery, as well as for analysis of the pharmaceutical and toxicological effects of existing therapeutics.

Need for Protein Separation

The five main steps of proteome analysis are shown in **Figure 1**. The primary requirement is that we must be able to separate the very complex protein mixtures present in lysates of cells, tissues and organisms. It is generally accepted that the best method currently available is two-dimensional polyacrylamide gel electrophoresis (2D electrophoresis). While there are several possibilities for combination of electrophoretic procedures, the most effective approach is a combination of a first-dimension separation by isoelectric focusing (IEF) under denaturing conditions with a second-dimension separation by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). This results in the sample proteins being separated according to their charge properties (i.e. isoelectric point, p*I*) in the first dimension followed by a size-based (molecular weight) separation in the second dimension. As the charge and size properties of proteins are essentially independent parameters, this results in the sample proteins being distributed across the whole area of the 2D separation (**Figure 2**).

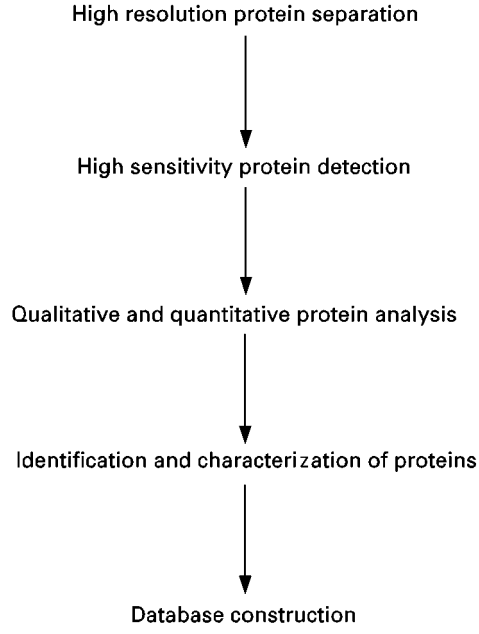


Figure 1 The five main steps in proteome analysis.

O'Farrell Method of 2D Electrophoresis

The method of 2D electrophoresis (2DE) described by O'Farrell in 1975 has formed the basis of almost all subsequent developments in 2-DE. This method used a combination of IEF in cylindrical gels (cast in capillary tubes) containing 8 M urea and 2% of the nonionic detergent, Nonidet P-40 (NP-40), with the SDS-PAGE system of Laemmli. However, for effective use in proteome analysis, 2DE must be capable of consistently reproducible high resolution protein separations. This proved to be problematic largely due to the nature of the synthetic carrier ampholytes (SCA) used to generate the pH gradients for IEF. SCA are small molecules which are freely mobile within the IEF gel, and the electroendosmotic flow of water which occurs during IEF results in their migration towards the cathode. This process is known as cathodic drift and results in pH gradient instability, with loss of the more basic proteins from the final 2D gel pattern.

2DE using IPG IEF

This problem has been largely overcome with the development of the Immobiline reagents (Amersham Pharmacia Biotech) for the generation of immobilized pH gradients (IPG) for IEF. The Immobiline reagents are acrylamide derivatives which give a series of buffers with different pK_a values distributed across the range pH 3–10. The appropriate Immobiline reagents, calculated from the extensive series of published recipes, are added to the mixture used for gel polymerization, resulting in the buffering groups which will form the pH gradient being covalently attached via vinyl bonds to the polyacrylamide backbone. This immobilization of the pH gradient renders it immune to the effects of electroendosmosis, resulting in highly stable IEF separations.

Despite early problems which were encountered in the application of IPG IEF to 2DE separations, this method has now become the method of choice for the first dimension of 2DE. The procedure which is now used is largely based on the work of Görg and her colleagues (see Further Reading). Briefly, IPG IEF is performed in individual gel strips, 3–5 mm wide, cast on plastic supports. After IEF, the gel strips are equilibrated to allow the separated proteins to interact with SDS, and then applied either to the

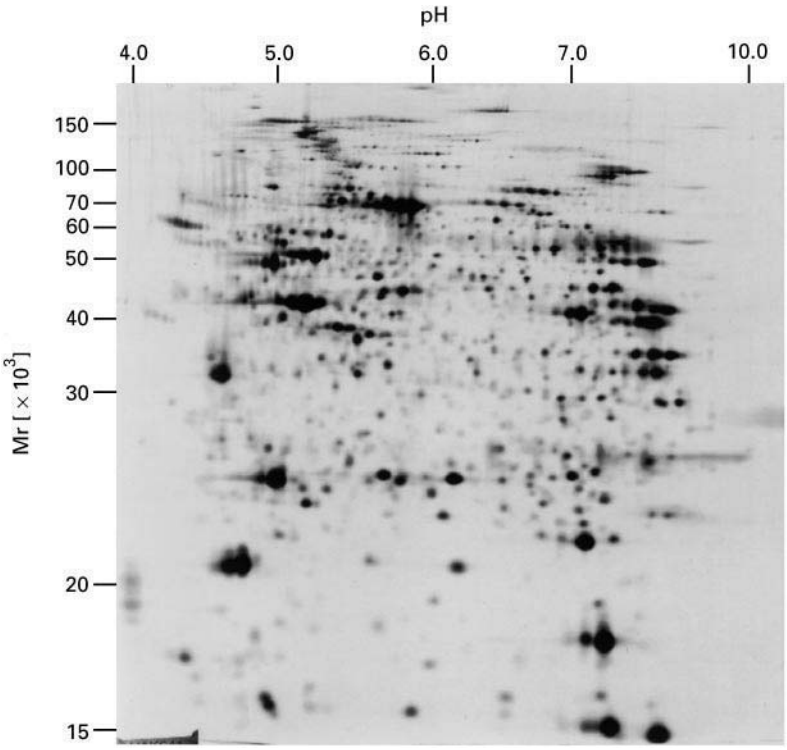


Figure 2 A 2DE separation of 100 μg of human heart using a nonlinear pH 3.5–10 IPG IEF gel was used in the first dimension and 12% SDS-PAGE in the second dimension. The separated proteins were visualized by silver staining.

surface of a horizontal SDS-PAGE gel to the top of a vertical. Interlaboratory studies of heart, barley and yeast proteins have unequivocally demonstrated the excellent reproducibility of both protein spot position and quantity that can be achieved with this method.

Separation Capacity of 2DE

The ability of 2DE to resolve complex mixtures of proteins is dependent on the gel area used for the separation. Thus, the 'standard' combination of 18 cm IPG strips with 20 cm SDS-PAGE gels is capable of routinely separating 2000 proteins from lysates of whole cells and tissues. It has been shown that very large gel formats (up to 30 cm in each dimension) are capable of separating up to 10 000 proteins from such samples, but this is achieved at the expense of the ease of gel handling and processing. In contrast, only a few hundred proteins can be separated using mini-format 2DE, but this approach has the advantage of rapid separations for screening purposes.

IPG IEF also provides great flexibility in the choice of pH gradient used for the separation, providing an additional approach to maximize the efficiency of separation of the particular protein mixture under investigation. Thus, IPG IEF gels spanning the range pH 3.5–10 are ideal for examining the diversity of protein in a sample (Figure 2). Optimal resolution of proteins in a particular pH range can be achieved using narrower pH gradients (**Figure 3**).

A further advantage of IPG IEF is that it has a very high capacity for micropreparative 2DE protein separations, particularly using a recently described method in which IPG strips are reswollen directly in a solution containing up to several mg of the protein sample to be analysed.

Protein Detection

The next requirement for effective proteome analysis is detection of the separated proteins at high sensitivity. The Coomassie brilliant blue dyes have been the most commonly used reagents for detecting protein zones separated by gel electrophoresis, but their limited sensitivity (around 0.5 μg per protein spot) necessitates the use of relatively high sample loadings (> 500 μg). Much higher sensitivity of detection (0.1 ng per protein spot) can be achieved by silver staining (Figures 2 and 3), but there can be problems in using this method as a quantitative procedure as it is known to be non-stoichiometric. Silver staining intensity is linear over the range from 0.04 to 2 ng protein/mm^2, but above this limit the stain becomes nonlinear, resulting in saturation and even negative staining effects.

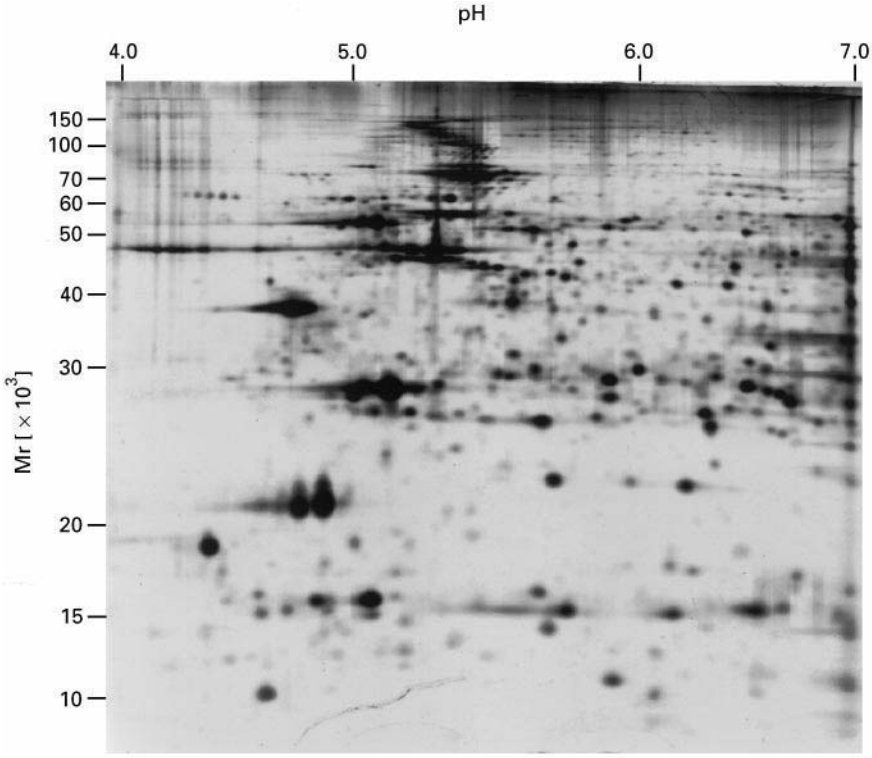


Figure 3 A 2DE separation of 100 μg of human heart using a linear pH 4–7 IPG IEF gel was used in the first dimension and 12% SDS-PAGE in the second dimension. The separated proteins were visualized by silver staining.

Many of these problems can be overcome using detection methods based on the use of fluorescent compounds. Such methods are highly sensitive and generally exhibit excellent linearity and a high dynamic range, making it possible to achieve excellent quantitative analysis if a suitable imaging device is used. A variety of fluorescent compounds are available for labelling proteins prior to electrophoresis. However, such pre-electrophoretic staining often results in charge modification, resulting in alterations in spot positions during 2DE. Recently, compounds which react with cysteine or lysine residues have been used successfully for 2DE. The cysteine-reactive reagent, monobromobimane was found to have a sensitivity of 1 pg protein per spot when imaged using a cooled CCD camera. The amine-reactive cyanine dyes, propyl Cy3 and methyl Cy5, have been used to label *E. coli* proteins. These dyes are claimed to have an inherent positive charge, thereby preserving the overall charge of the proteins after coupling. Due to their spectral properties, the two dyes can be distinguished when present together, allowing two different samples each labelled with one of the dyes to be mixed together and separated on the same 2DE gel. This method, which has been termed difference gel electrophoresis (DIGE), has great potential for improving the efficiency of detection of differences in 2DE protein patterns between different samples.

One approach to overcoming the problems associated with charge modification during the IEF dimension is to label the proteins while present in the first dimension gel after IEF, prior to the second dimension separation by SDS-PAGE. The fluorescent compound 2-methoxy-2,4-diphenyl-3(2H)-furanone (MDPF) and a fluorescent maleimide derivative have been used in this way. The alternative approach is to label the proteins after the 2DE separation has been completed. Recently, two post-electrophoretic fluorescent staining reagents, SYPRO orange and red have been described. These stains have a very high sensitivity (2 ng protein per spot) and excellent linearity with a high dynamic range.

Metabolic labelling of proteins with a radiolabelled amino acid prior to their separation by 2DE provides a very sensitive method for protein detection. This approach is most commonly used with *in vitro* cell culture systems, but it is also possible to radiolabel synthetically the proteins in small pieces of fresh tissue. While proteins can be readily radiolabelled postsynthetically by methods such as radioiodination with [^{125}I] or reductive methylation with [^{3}H]-sodium borohydride, these result in significant charge modifications precluding their use in proteome analysis.

Qualitative and Quantitative Analysis

The first step in the analysis of 2DE protein profiles is to produce a digitized image. Stained gels can be digitized using a flat-bed scanning laser densitometer or a suitably modified document scanner. Autoradiographic film images of 2DE separations of radiolabelled proteins can also be imaged in this way, but more accurate quantitation can be achieved using a phosphorimaging scanner. Fluorescently labelled protein separation patterns can be visualized using either a dedicated scanning laser densitometer (fluorimager) or a camera system fitted with cooled CCD array. Several commercial software systems for the analysis of 2DE gels are now available running on desktop workstations (Unix, PC, Mac). These systems make it possible to extract quantitative and qualitative information from individual 2DE gels, to match protein patterns from large collections of 2DE gels, and thereby establish comprehensive databases which can be used to investigate quantitative protein expression in cells, tissues and organisms.

Protein Identification and Characterization

It is clear from the foregoing that 2DE provides information on the p*I*, molecular weight and relative abundance of the separated proteins. However, it provides no direct clues to their identities or functions. The p*I* and molecular weight information can be used to search sequence databases for proteins with similar properties, for example using the TagIdent tool (http://www.expasy.ch/www/guess-prot.html), but the uncertainty of molecular weight estimation by SDS-PAGE (typically around $\pm$ 10%) makes this process unreliable. Recently, mass spectrometry (MS) has been used to measure directly the mass of proteins separated by 2DE. In this approach the proteins are transferred by Western blotting onto the surface of a nitrocellulose or PVDF membrane which is then treated with a matrix required for MS. The protein spot of interest is excised, mounted directly into a matrix-associated laser desorption ionization mass spectrometer (MALDI-MS) and the mass of the intact protein measured (**Figure 4**). We have found that this method is very accurate, usually within 1% of the predicted mass, but requires a MALDI-MS fitted with an infrared laser. While such mass data can be invaluable in characterizing post-translational modifications of proteins, it is unlikely on its own to result in unequivocal protein identification. Fortunately, over the last few years, several methods have been developed which make it possible to identify and characterize proteins separated by 2DE (**Figure 5**).

Western Blotting

The first major advance in the characterization of proteins separated by gel electrophoresis was the development of Western blotting. In this technique, the separated proteins are transferred ('blotted') from the restrictive gel matrix, usually by application of an electric field perpendicular to the plane of the gel ('electroblotting'), onto the surface of a suitable membrane support such as nitrocellulose or PVDF. The proteins can then be probed with a variety of ligands,

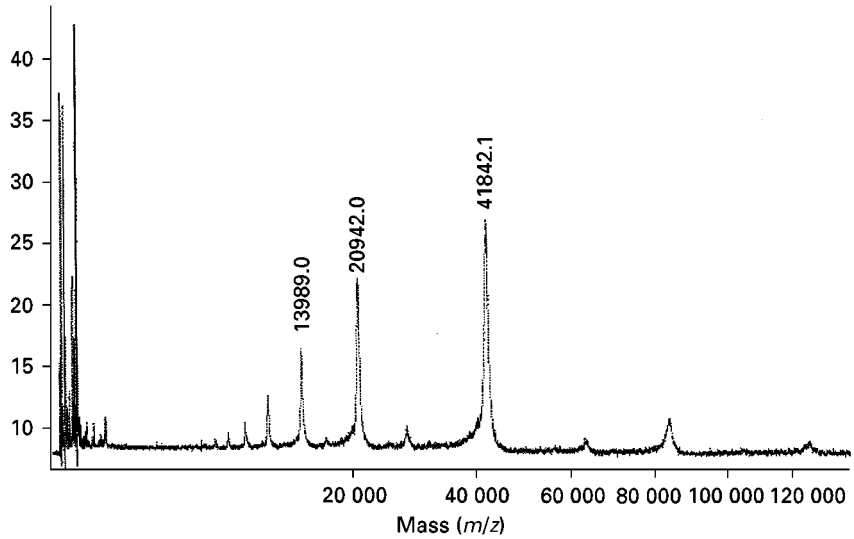


Figure 4 IR-MALDI mass spectrum of a protein spot from a 2D gel electroblotted onto a PVDF membrane. Mass peaks indicated are multiply charged or dimers of the molecular ion. The protein spot is known to be cardiac α-actin. The measured mass of the molecular ion (41842.1) is very close to the theoretical value determined from its amino acid sequence (41784.6).

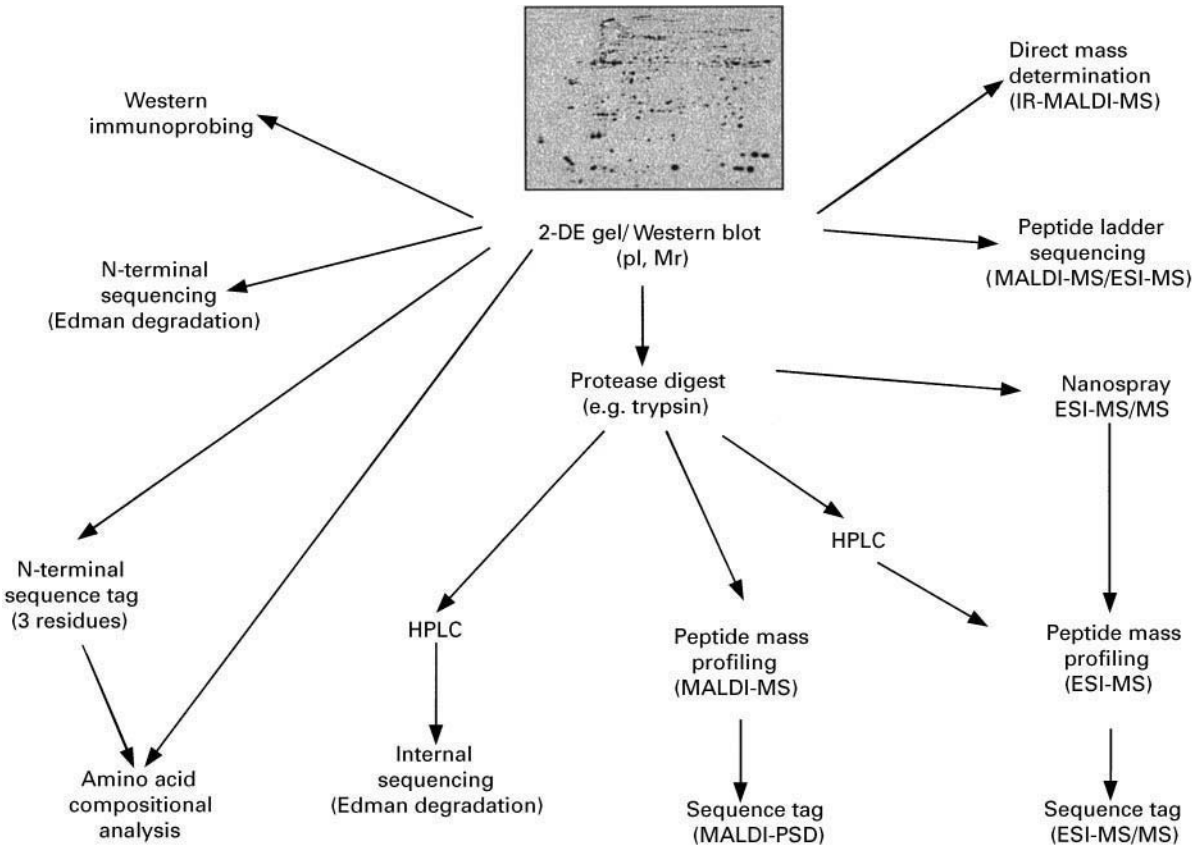


Figure 5 Methods currently used for the identification and characterization of proteins separated by 2DE.

particularly poly- and monoclonal antibodies. This approach has been used quite extensively for the identification of known proteins separated by 2DE, but is a very time-consuming process that is dependent on the availability of a suitable panel of specific antibodies reactive with the denatured proteins in 2DE gels.

Amino Acid Sequence Determination by Edman Degradation

Amino acid sequence, even if this is only a few residues in length, is the most specific method of protein identification. The chemical sequencing of proteins has been possible for half a century since the development of the method known as Edman degradation in 1949. This remained a laborious manual procedure until the development of the first automated protein sequenators, with the first commercial 'spinning cup' instrument becoming available in 1971. This instrument was relatively insensitive, requiring at least 10 nmol of sample (equivalent to 500 μg for a 50 kDa protein). However, progress in sequenator technology has resulted in the current generation of gas–liquid sequenators which are capable of generating N-terminal sequence information from low picomole quantities of protein (1 pmol is equivalent to 0.05 μg for a 50 kDa protein). This level of sensitivity is compatible with the amount represented by many of the spots present on micropreparative 2DE gels and this method remains the method of choice if extended runs of N-terminal protein sequence are required. This is a particularly important consideration for the analysis of apparently 'novel' proteins, i.e. sequences not present in protein and nucleotide databases. Although chemical protein sequencing is a sensitive and informative method of protein identification, throughput is very low, typically one or two samples per day. Thus, there is a need for alternative approaches which allow rapid and sensitive screening of gel proteins separated by 2DE, so that only those which cannot be identified unequivocally or appear to be novel require further detailed characterization.

Problem of N-Terminal Blockage

A major problem with protein sequencing by Edman degradation is that many proteins lack a free α-amino group, due to co- or post-translational modification.

Such N-terminal blockage occurs typically in up to 50% of eukaryotic proteins and results in no sequence being obtained. This problem can be overcome by subjecting the separated protein, either *in situ* within the gel matrix or after Western blot transfer onto a nitrocellulose or PVDF membrane, to chemical (e.g. CNBr) or enzymatic (e.g. trypsin) cleavage to generate shorter peptides which can be isolated and sequenced. The cleavage products are then usually separated by RP-HPLC, and selected peptide fractions directly applied to the protein sequenator. This procedure is highly efficient, but the determination of multiple stretches of sequence usually requires two to three times more protein than does N-terminal protein sequence analysis.

Amino Acid Compositional Analysis

Amino acid compositional analysis (AAA) is the best method for the absolute measurement of protein abundance. Current methods for the analysis of fluorescently derivatized amino acids have sub-pmole sensitivity and so can be applied directly to proteins separated by 2DE. An individual proteins have more or less unique amino acid compositions, AAA can be an excellent method for the rapid identification of proteins separated by 2DE, in which the experimental amino acid composition is compared with amino acid sequences generated *in silico* from protein and nucleotide sequence databases. The major drawback of this approach is that the output is a ranked list of possible protein identities (**Figure 6**) and the 'correct' protein does not necessarily occur as the first ranked entry. This method is better used in conjunction with another rapid method of protein identification such as peptide mass profiling (see later) and this orthogonal approach has been found to be useful for identifying proteins even across the species barrier. Another approach to improving protein identification by AAA is to generate a short N-terminal protein 'sequence tag' by Edman degradation and to use this in combination with the AAA data for protein identification.

Peptide Mass Profiling

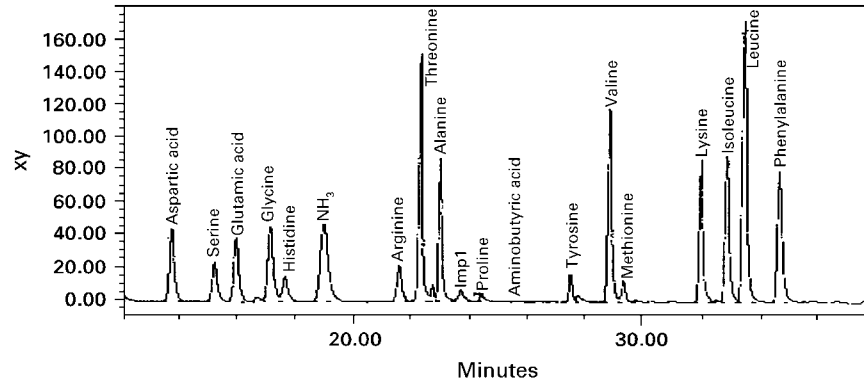
It has long been realized that the peptides generated by chemical (e.g. CNBr) or enzymic (e.g. trypsin) digestion of a protein are characteristic of that protein and such peptide fingerprints or maps analysed by chromatography or electrophoresis have been used for investigating the relatedness of proteins. The advent of MS methods for the analysis of peptides has made this into a much more powerful approach for protein identification. In this method the peptide masses obtained by MS of a protein digest are used to interrogate databases of peptide masses generated *in silico* from protein and nucleotide sequence databases. As in the case of AAA, this technique of peptide mass profiling or fingerprinting produces a list of possible protein identities ranked in order of probability (**Figure 7**). The reliability of this method can be increased by combining peptide mass profiling data from two or more separate digests (e.g. trypsin, Lys-C) or by adopting an orthogonal approach in combination with AAA (see above).

The enzymatic cleavage of the 2DE protein spot can be carried out either *in situ* within the gel matrix or after electroblotting to a suitable membrane (nitrocellulose or PVDF). After recovery, the unfractionated peptide can be readily analysed by MALDI-MS (Figure 7). Alternatively, the peptide mixture can be fractionated by high performance liquid chromatography (HPLC), with the fractions being analysed either offline by MALDI-MS or online by electrospray ionization (ESI)-MS using a quadrupole or ion-trap instrument.

Amino Acid Sequence Determination by Mass Spectrometry

Recently alternative techniques for the determination of the primary sequence of peptides and proteins have been developed based on the use of mass spectrometry (MS). This can be achieved by peptide fragmentation within the spectrometer or by a method termed 'ladder sequencing'. In the latter approach, Edman chemistry or enzymic degradation with aminopeptidase or carboxypeptidase is used under limiting conditions to produce an overlapping series of truncated peptides. These differ in size according to the number of amino acid residues which have been removed from their N- or C-terminus, allowing the sequence to be deduced by measurement of the peptide masses by MALDI-MS. A high mass accuracy is required and it is not possible to distinguish between leucine and isoleucine as these residues have an identical mass.

The alternative approach is to take advantage of the ability of two-stage mass spectrometers, either MALDI-MS with post-source decay (PSD) or ESI-MS/MS triple-quadropole or ion-trap instruments, to induce fragmentation of peptide bonds. It is possible to use this approach to generate extended *de novo* amino acid sequence information, but it requires considerable expertise to interpret the complex spectra that are obtained. However, partial sequence data is an extremely powerful adjunct to the identification of proteins by peptide mass profiling.



(A)

AMINO ACID COMPOSITION

Protein	Unknown protein	Calibration protein	Bias	Weight
Asx	5.70	4.67	1.00	1.00
Glx	5.25	8.01	1.00	1.00
Ser	3.26	3.38	1.00	1.00
His	2.38	3.45	1.00	1.00
Thr	14.08	6.61	1.00	1.00
Pro	0.50	2.48	1.00	1.00
Tyr	1.43	4.85	1.00	1.00
Arg	2.96	4.38	1.00	1.00
Val	12.44	11.12	1.00	1.00
Met	1.45	1.21	1.00	1.00
Ile	10.17	5.48	1.00	1.00
Leu	20.84	22.42	1.00	1.00
Phe	10.92	13.56	1.00	1.00
Lys	8.63	8.40	1.00	1.00

(B)

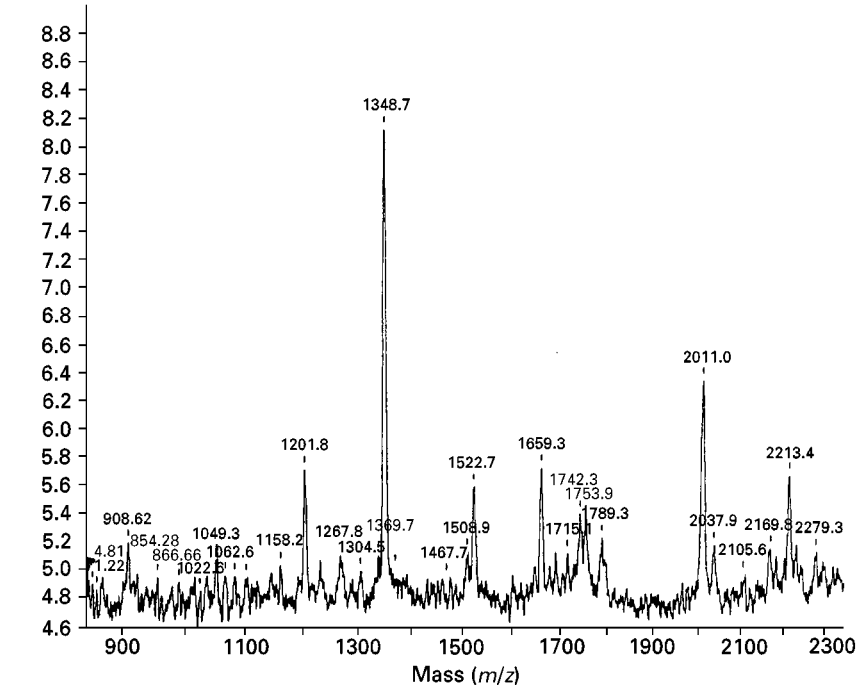
Rank	Score	Protein	(pI)	(Mw)	Description
1	11	FABH_HUMAN	6.34	14727	FATTY ACID-BINDING PROTEIN, HEART (H)
2	17	FABH_MOUSE	5.69	14689	FATTY ACID-BINDING PROTEIN, HEART (H)
3	28	FABH_RAT	5.92	14644	FATTY ACID-BINDING PROTEIN, HEART (H)
4	37	FABL_GINCI	6.79	15079	FATTY ACID-BINDING PROTEIN, LIVER (L)
5	48	FABH_BOVIN	6.91	14648	FATTY ACID-BINDING PROTEIN, HEART (H)
6	56	FABE_RAT	6.73	15059	FATTY ACID-BINDING PROTEIN, EPIDERMAL
7	69	HBB_CHEKU	7.11	16458	HEMOGLOBIN BETA CHAIN
8	70	BTF3_HUMAN	6.85	17699	BTF3B
9	73	FABE_HUMAN	6.60	15164	FATTY ACID-BINDING PROTEIN, EPIDERMAL
10	77	FABE_MOUSE	6.14	15137	FATTY ACID-BINDING PROTEIN, EPIDERMAL
11	84	RET2_MOUSE	6.12	15478	RETINOL-BINDING PROTEIN II, CELLULAR
12	89	RET3_HUMAN	5.31	15434	RETINOIC ACID-BINDING PROTEIN I
13	89	FABI_HUMAN	6.88	15076	FATTY ACID-BINDING PROTEIN, INTESTINAL
14	92	RET3_BOVIN	5.31	15460	RETINOIC ACID-BINDING PROTEIN I

(C)

Figure 6 The identification of a protein spot from a 2DE separation by amino acid compositional analysis. (A) HPLC analysis of pre-column derivatized amino acids resulting from hydrolysis of the protein spot. (B) Amino acid composition determined from the HPLC analysis. (C) Result of the amino acid composition database search indicating that the protein is cardiac fatty acid binding protein.

The method known as 'peptide sequences tagging' is based on interpretation of a portion of the ESI-MS/MS or PSD-MALDI-MS fragmentation data to generate a short partial sequence or 'tag', which is used in combination with the mass of the intact parent peptide ion, and provides a significant amount of additional information for the homology search. A powerful extension of this approach has been the development of a nano-electrospray ion source that allows spraying times of more than 30 min from about 1 μL of sample. The sensitivity of this method is in the low femtomole range and silver stained 2DE protein spots containing as little as 5 ng protein have been successfully sequenced. Moreover, using this

(A)



Valid search masses: 908.6 1049.3 1201.8 1267.8 1304.5 1508.9 1522.7 1659.3 1742.3 1733.9 1789.3 2011.0 2037.9 2169.8 2213.4
Search mass uncertainty = ± 4.0
42655 < Protein mass < 47145
Enzyme: TRYPSIN
Scoring scheme: MOWSE

(B)

Score	No. Hits	Mol. wt.	Accession	Description
667	8	43023.2	KCRM_CANFA	CREATINE KINASE, M CHAIN.
415	6	46899.6	SYH_ECOLI	HISTIDYL-TRNA SYNTHETASE (HISTIDINE--TRN)
304	6	45368.9	gi_1022975	CELRLLF4 cds3
294	5	43200.1	IL3R_HUMAN	INTERLEUKIN-3 RECEPTOR ALPHA CHAIN PRECU
255	5	46571.4	THD1_BACSU	THREONINE DEHYDRATASE BIOSYNTHETIC (THRE
252	5	42982.1	KCRM_RABIT	CREATINE KINASE, M CHAIN.
244	5	44323.3	KAS1_STRCM	PUTATIVE POLYKETIDE BETA-KETOACYL SYNTHA
242	5	45394.0	GUNB_RUMAL	ENDOGLUCANASE B PRECURSOR (ENDO-1, 4-BETA
235	5	46242.2	gi_609282	p46Eg265
228	5	44634.6	BMP8_MOUSE	BONE MORPHOGENETIC PROTEIN 8 PRECURSOR (
224	5	47077.4	TYPH_ECOLI	THYMIDINE PHOSPHORYLASE.
218	5	43929.5	INVO_GALCR	INVOLUCRIN.
214	4	42746.2	EFTU_BURCE	ELONGATION FACTOR TU (EF–TU).
211	4	45847.6	gi_532474	CELTO7E3 cds7
210	4	43463.5	EFTU_BACSU	ELONGATION FACTOR TU (EF-TU)
210	5	42971.2	KCRM_HUMAN	CREATINE KINASE, M CHAIN.
207	4	45285.1	LIPA_HUMAN	LYSOSOMAL ACID LIPASE/CHOLESTERYL ESTER
204	4	45353.3	UDHG_STRPY	UDP-GLUCOSE 6-DEHYDROGENASE.
191	5	46167.1	SAGP_STRPY	STREPTOCOCCAL ACID GLYCOPROTEIN.
189	4	42851.6	ERK1_RAT	EXTRACELLULAR SIGNAL-REGULATED KINASE 1
188	4	43005.8	ERK1_HUMAN	EXTRACELLULAR SIGNAL-REGULATED KINASE 1
188	4	45227.4	GLPB_ECOLI	ANAEROBIC GLYCEROL-3-PHOSPHATE DEHYDROGE
187	4	43917.4	TRPB_PSEPU	TRYPTOPHAN SYNTHASE BETA CHAIN.
186	5	47010.5	CBP1_HUMAN	CARBOXYPEPTIDASE A1 PRECURSOR.
185	4	44679.9	gi_1066945	F34D10.5
183	4	45670.1	ACRO_HUMAN	ACROSIN PRECURSOR.
182	4	43439.4	RHO_BORBU	TRANSCRIPTION TERMINATION FACTOR RHO.
181	4	44666.2	HMPA_AICEU	FLAVOHEMOPROTEIN (HAEMOGLOBIN-LIKE PROTEIN)

Figure 7 The identification of a protein spot from a 2DE separation by peptide mass fingerprinting. (A) MALDI-MS spectrum of the tryptic digest of the protein spot. (B) Result of the peptide mass database search indicating that the protein is the M-chain of creatine kinase.

method it is possible to sequence multiple peptides from a digest without the need for their prior separation by RP-HPLC.

An alternative approach is based on the automated interpretation of ESI-MS/MS fragmentation data which is used to directly search sequence databases. In the first step in this process all those peptides that can be generated from proteins in the sequence databases and whose masses match those of the measured peptide ion are identified. The fragment ions expected for each of the candidate peptides are then calculated and the experimentally determined MS/MS spectrum is then compared with the predicted spectra using cluster analysis algorithms. This

method has been fully automated and sensitivity is at the low femtomole level.

Database Construction

The final requirement for proteome analysis is the construction of databases to store the data generated and to make this readily available within the laboratory and, where possible, accessible to other scientists worldwide. The best approach to this is currently offered by the World Wide Web (WWW) on the Internet. There is currently no international standard for the construction of such databases, but in order that there should be good connectivity between them it has been suggested that they are constructed according to a set of fundamental rules. Databases conforming to these rules are said to be 'Federated 2D Databases' and a list of these can be viewed at WORLD-2DPAGE (http://www.expasy.ch/ch2d/2d-index.html).

Conclusions

From the foregoing discussion it can be seen that proteomics provides an interface with genomics which can provide information on protein expression in biological systems. This information will aid our understanding of complex cellular processes and the way in which they react to varying conditions. Proteomics will also provide insights into processes of disease at the molecular level and should result in the development of novel diagnostics and therapeutics.

Further Reading

Dunn MJ (1987) Two-dimensional polyacrylamide gel electrophoresis. In: Chrambach A, Dunn MJ and Radola BJ (eds) *Advances in Electrophoresis*, vol. I, pp. 1–109. Weinheim: VCH.

Görg A, Boguth G, Obermaier C, Posch A and Weiss W (1995) Two-dimensional polyacrylamide gel electrophoresis with immobilized pH gradients in the first dimension (IPG-Dalt): The state of the art and the controversy of vertical versus horizontal systems. *Electrophoresis* 16: 1079–1086.

Humphery-Smith I, Cordwell SJ and Blackstock WP (1997) Proteome research: Complementarity and limitations with respect to the RNA and DNA worlds. *Electrophoresis* 18: 1217–1242.

Lamond AI and Mann M (1997) Cell biology and the genome projects: A concerted strategy for characterizing multiprotein complexes by using mass spectrometry. *Trends in Cell Biology* 7: 139–142.

Patterson SD (1994) From electrophoretically separated protein to identification: Strategies for sequence and mass analysis. *Analytical Biochemistry* 221: 1–15.

Patterson SD and Aebersold R (1995) Mass spectrometric approaches for the identification of gel-separated proteins. *Electrophoresis* 16: 1791–1814.

Pennington SR, Wilkins MR, Hochstrasser DF and Dunn MJ (1997) Proteome analysis: From protein characterization to biological function. *Trends in Cell Biology* 7: 168–173.

Sutton CW, Wheeler CH, U S, Corbett JM, Cottrell JS and Dunn MJ (1997) The analysis of myocardial proteins by infrared and ultraviolet laser desorption mass spectrometry. *Electrophoresis* 18: 424–431.

Wan JS, Sharp SJ, Poirer GM *et al.* (1996) Cloning differentially expressed mRNAs. *Nature Biotechnology* 14: 1685–1691.

Wasinger VC, Cordwell SJ, Cerpa-Poljak A *et al.* (1995) Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium. *Electrophoresis* 16: 1090–1094.

Wheeler CH, Berry SL, Wilkins MR *et al.* (1996) Characterisation of proteins from two-dimensional electrophoresis gels by matrix-assisted laser desorption mass spectrometry and amino acid compositional analysis. *Electrophoresis* 17: 580–587.

Wilkins MR, Pasquali C, Appel RD *et al.* (1996) From proteins to proteomes: Large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology* 14: 61–65.

Wilkins MR, Williams KL, Appel RD and Hochstrasser DF (1997) *Proteome Research*: *New Frontiers in Functional Genomics*. Berlin: Springer.

Wilm M, Shevchenko A, Houthaeve T *et al.* (1996) Femtomole sequencing of proteins from polyacrylamide gels by nanoelectrospray mass spectrometry. *Nature* 379: 466–469.

Yates JR (1998) Database searching using mass spectrometry data. *Electrophoresis* 19: 893–900.

PURGE-AND-TRAP: GAS CHROMATOGRAPHY

See **III / VOLATILE ORGANIC COMPOUNDS IN WATER: GAS CHROMATOGRAPHY**

PYROLYSIS: GAS CHROMATOGRAPHY

See **II / CHROMATOGRAPHY: GAS / Pyrolysis Gas Chromatography**