

# Proteomics

## Making Sense of Genomic Information for Drug Discovery

Julian P. Whitelegge<sup>1</sup> and Johannes le Coutre<sup>2</sup>

1 Department of Chemistry and Biochemistry and The Pasarow Mass Spectrometry Laboratory, University of California  
Los Angeles, Los Angeles, California, USA

2 Nestlé Research Center, Lausanne, Switzerland

### Contents

Abstract	29
1. Goals	30
2. Technology	30
2.1 Two-Dimensional Gel-Electrophoresis	30
2.2 Mass Spectrometry (MS)	30
2.2.1 Peptide Mass Tags	30
2.2.2 Sequence Tags	31
2.2.3 Intact Mass Tags	31
2.2.4 Membrane Proteins	31
2.2.5 Quantitative Proteomics Using MS	32
2.3 Determination of Protein Structures	32
2.4 Data Management and Evaluation	33
2.5 Expression Patterns, Microarray Data, DNA Chips and Protein Chips	33
3. Proteomics, Drugs and Therapy	33
4. Conclusion	34

### Abstract

As an increasing number of available genomes triggers a gold rush in modern biology, the scientific challenge shifts towards understanding the total of the encoded information, most notably the proteins, their structures, functions and interactions. Currently this work is in its early stages but the near future will bring a merger of biology, engineering and informatics with a far broader impact on society than pure genomics has had so far. The challenge of characterizing the structures and functions of all proteins in a given cell demands technological advances beyond the classical methodologies of protein biochemistry. Mass spectrometry techniques for high-throughput protein identification, including peptide mass fingerprinting, sequence tagging and mass spectrometry on full-length proteins are providing the driving force behind proteomics endeavors. New technologies are needed to move high-resolution protein structure determination to an industrial scale. Nonetheless, improvements in techniques for the separation of intrinsic membrane proteins are enabling proteomics efforts towards identifying drug targets within this important class of biomolecules. Beyond the acquisition of data on sequences, structures and interactions, however, the major work in drug discovery remains: the screening of large candidate compound libraries combined with clever medicinal chemistry that guarantees selective action and defined delivery of the drug.

To separate proteomics from other '-omics', the term should be defined as: characterising relative abundance, all structures and functions and the interactions of the complement of proteins in a given entity. This global approach is the difference between proteomics and conventional protein biochemistry. In contrast, genomics is defined as obtaining and understanding the complement of genetic information available to a given organism.

Therefore, proteomics is dependent on genomic information and utilizes it to understand the 3-dimensional world of functional proteins. This is an important step, not only conceptually, but also practically, which leads from raw and unprocessed linear information to the mechanical structure of the cell and its temporal and spatial organization. Moreover, proteomics seeks to understand interactions between proteins to explain the more complex

phenomena of physiology such as regulation, signal transduction or development and differentiation. For the human genome with 30 000 to 40 000 genes<sup>[1,2]</sup> an equally high number of proteins can be expected, and in fact many more when splicing and post-translational modifications are considered.<sup>[3]</sup>

## 1. Goals

To benefit from genomic information generated in recent years, a full understanding of a gene's function is necessary, an effort for which the term functional genomics has been coined. Even though parts of a genome such as regulatory regions, stem/loop structures or tRNAs reside on the level of nucleic acids, most genomic information is translated into proteins with distinct structures, functions and interactions. The ultimate goal is to understand the working of an entire living cell in a specific tissue, a massive web of processes that are predominantly based on proteins. Thus, proteomics goes beyond characterization of individual proteins, its strength and goal lies in the characterization of all proteins in a given cell. Understanding all protein structures as well as their functions and dynamics in a cell is a daunting task which still requires substantial new technical developments, especially on the side of protein structure determination.

Nevertheless, and especially in the case of soluble proteins, a growing number of high-resolution structures is available allowing the study of protein function, protein/protein and also protein/ligand interactions toward rational drug design for treatment/prevention not only of infectious diseases but also of genetic disorders and inherited diseases. Unfortunately, a huge discrepancy still exists between available information and the determination of metabolic significance in the case of membrane proteins. On one hand these proteins are preferred targets for drugs because the exposed location on the cell surface makes them indispensable participants for drug entry into the cell. On the other hand, membrane proteins are notoriously difficult to study and still only about two dozen high-resolution structures of membrane proteins are known.

## 2. Technology

The technical demands to accomplish successful proteomic investigations are tremendous at every stage of the process and so far not a single cellular proteome has been characterized to completion. Large numbers of genes ( $10^4$  to  $10^6$  per organism) have to be analysed in order to describe large numbers of proteins that in turn must be characterized structurally and functionally down to the atomic level. The information obtained will be stored in appropriate database systems that allow sufficient annotations, cross referencing and rapid access. Therefore, every aspect of the

data acquisition side of a proteomics project is labeled as 'high throughput', typically on the 'micro(nano) scale', whereas activities on the data storage and analysis side are often called 'data-mining'. Consequently, throughout the experimental stages of a large proteomics project, use is being made of robotic systems all the way from sample storage through sample preparation to data acquisition. To obtain proteomic information a set of key technologies is being used.

### 2.1 Two-Dimensional Gel-Electrophoresis

The classical, biochemical method for analysis of large and complex protein mixtures, high-resolution 2-dimensional gel-electrophoresis (2D gels), remains the key technology for proteomic research. Separation of mixtures is accomplished by isoelectric focusing of proteins according to their charge followed by SDS polyacrylamide gel electrophoresis (PAGE) that separates according to size.<sup>[4]</sup> Over a thousand proteins can be resolved this way and samples from healthy versus diseased tissues, for example, can be compared to provide a global picture of expression changes. With the recent and dramatic progress that has been made in the development of computer technology and mass spectrometry, as well as the availability of genome data, a leap has taken place in proteomic research. Proteins on 2D gels can be quantified by advanced imaging software and the spots identified rapidly using mass spectrometry. Recent advances in 2D PAGE technology in Proteomics have been extensively reviewed.<sup>[5]</sup>

### 2.2 Mass Spectrometry (MS)

Mass spectrometry is driving proteomics as modern equipment allows high-throughput identification of proteins via comparison with genomic data. Different approaches have distinct advantages and disadvantages.

#### 2.2.1 Peptide Mass Tags

In the technique of peptide mass tagging (peptide mass fingerprinting) proteins are treated with a cleavage agent, typically the protease trypsin that cleaves specifically after arginine and lysine residues, and the resulting fragment mass is measured as accurately as possible. This measurement is usually carried out by matrix-assisted laser desorption ionization, time of flight mass spectrometry (MALDI-TOF MS).<sup>[6]</sup> Experimentally derived masses are then matched to databases (reviewed in Fenyó<sup>[7]</sup>). The database software translates genes or expressed sequence tags (ESTs) in all possible frames, digests the 'cyberprotein' *in silico* and calculates then the theoretical masses of the fragments. Positive identification is achieved when the observed masses closely match theoretical masses; it is improved by more closely

matching measured with calculated masses and increasing the number of peptides matched. Accuracy of mass measurement is critical<sup>[8]</sup> but modern MALDI-TOF instruments achieving 10 ppm ( $\pm 0.1$  Da at 10 kDa) or better, combined with sample preparation robots, provide massive throughput (10 to 100 samples/hour) making this the most widely used methodology currently used in proteomics.)

Common protocols involve selection of important targets from 2D-gel experiments for subsequent identification but more ambitious protocols seek to systematically scan entire 2D gels in an automated fashion.<sup>[9]</sup> Software improvements now permit consideration of covalent modifications that alter experimental masses in a predictable fashion.<sup>[10]</sup> Very recent success using this approach has been achieved at Large Scale Biology Corp. (see URL:<http://www.lsb.com>) by characterising HPI v1.0, a subset of the human proteome covering 115 693 proteins derived from some 28 000 genes.<sup>[11]</sup>

### 2.2.2 Sequence Tags

Generation of sequence tags involves tandem mass spectrometry (MS-MS) usually after electrospray-ionization (ESI)<sup>[12]</sup>. First, a mass selective filter is used to select an ion derived from a single protein peptide just as for mass tags (section 2.2.1). The selected ion is then fragmented in the mass spectrometer using collision activated dissociation (CAD), which typically breaks at the peptide bond, and the masses of the pieces are measured. Thus a sequence tag is the mass of a peptide plus the masses of CAD fragments derived from it. Identification against the database is similar to the technique described for mass tags except that the theoretical masses of CAD fragments have to be calculated based upon the sequence. Since this technique incorporates true sequence data, confidence in protein identification is raised, though computation is more cumbersome. Furthermore, the ESI mass spectrometers used to generate sequence tags can only perform one MS-MS experiment at a time, as opposed to the MALDI-TOF instruments that can record the masses of many different peptides in a single experiment. As a result of this constraint, reverse-phase liquid chromatography (LC) is typically coupled to ESI (LC-MS) in order to separate peptides from the digest mixture, and the throughput drops at least an order of magnitude compared to MALDI.

### 2.2.3 Intact Mass Tags

If a full-length protein can be mass measured accurately enough and its sequence matches that predicted by its gene, it is possible to identify proteins by their intact mass.<sup>[13]</sup> The widespread occurrence of post-transcriptional and post-translational modification in eukaryotes and especially humans generally precludes use of intact mass tags for identification, though future

advances in software may bring increased focus to this arena. However, the mass spectrum of an intact protein defines the native covalent state of the gene product and its heterogeneity.<sup>[14]</sup> Thus, in order to monitor any covalent modification it is useful to record the mass spectrum of the intact protein. This is of special importance for monitoring subtle alterations that might not alter protein mobility on 2D-gels. One example is methionine oxidation. Even single methionine oxidation events can significantly modulate protein function<sup>[15]</sup> without altering charge or migration in SDS-PAGE. Toward this end protocols for extraction of intact proteins from gels are being developed.<sup>[16]</sup>

Alternative strategies for separation of proteins from complex mixtures are also attractive because of the potential for covalent modification during 2D-gel analysis.<sup>[17]</sup> Furthermore, recent studies indicate that proteomics with whole cell extracts separated on 2D-gels is biased against less abundant proteins<sup>[18]</sup> and this has been hailed as a serious deficiency. At present there is no 'quick fix' for this problem and enrichment and purification of sub-fractions may be the only way to bypass this dilemma.<sup>[19]</sup> Howell, Yates and co-workers recently reported improved detection of minor abundance proteins of Golgi membranes by employing sub-fractionation techniques.<sup>[20]</sup> In studies of intact proteins by electrospray-ionization MS an added benefit arises in the ability to detect minor covalent variants of much more abundant proteins.<sup>[14]</sup> With current levels of investment in proteomics it seems likely that the near future will bring machines that automate the enrichment/purification process. The least abundant proteins may only be characterized after their capture by antibodies and a substantial effort is underway to provide proteome-wide antibody collections (reviewed in Holt et al.<sup>[21]</sup>). Through use of Fourier-Transform Ion Cyclotron Resonance (FTICR) mass spectrometry,<sup>[22]</sup> which provides the ultimate in accuracy and resolution, McLafferty and co-workers have proposed 'top-down' proteomics. This involves the recording of a mass spectrum of an intact protein prior to fragmentation by CAD or other means to generate fragment ions that can be used to identify the protein and characterize its modifications.<sup>[23,24]</sup>

### 2.2.4 Membrane Proteins

As the benefits of mass spectrometry of intact proteins become more widely appreciated, increased attention will be devoted to the separation techniques coupled to MS. Membrane proteins make up a significant proportion of all genomes analysed to date and are an important class of drug targets. Therefore, we have devoted considerable effort to develop separation techniques for intrinsic membrane proteins that are fully compatible with the electrospray-ionization process. ESI is preferred because of the superior accuracy and resolution attained over MALDI-TOF

analysis, though the latter is very useful in cases where excessive heterogeneity precludes interpretation of ESI spectra.<sup>[25]</sup> By performing reverse-phase liquid chromatography (LC)-MS on macroporous polymeric stationary phases in the presence of high concentrations of formic acid<sup>[26]</sup> it is possible to separate various intrinsic membrane proteins for ESI analysis. Rhodopsin was the first G-protein coupled receptor to be analyzed in this way; other membrane proteins thus far analyzed include the bacteriorhodopsin holoprotein and the thylakoid D1 herbicide receptor.<sup>[14]</sup> Larger membrane proteins are amenable to ESI-MS after size-exclusion high performance liquid chromatography (HPLC) in chloroform/methanol/aqueous formic acid, as demonstrated in an extensive study on the *Escherichia coli* lactose permease, which has 12 transmembrane helices.<sup>[27]</sup> Further studies demonstrated the broad applicability of the method including the successful analysis of membrane proteins with up to 15 transmembrane helices and molecular weights approaching 100 kDa.<sup>[13,28]</sup> Limited separation of complex protein mixtures has been achieved on extracts containing all *E.coli* membrane proteins, to a degree that at least 1 membrane protein could be identified unambiguously by comparison with the database.<sup>[13]</sup>

### 2.2.5 Quantitative Proteomics Using MS

One reason for the widespread popularity of 2D-gel analysis is the ability to quantify proteins based upon staining intensity. Up or down regulation of a protein's expression under an experimental treatment is visualized, and targets with altered expression identified by mass spectrometry. Mass spectrometry itself is not truly quantitative due to the variable ionization efficiencies associated with different proteins. However, a number of experimental protocols aimed at providing quantitative information from mass spectrometric measurements have been reported. These methods rely upon the distinction of different masses for a given species in the mass spectrometer based upon stable isotope incorporation, thus allowing measurement of relative expression.

In one report, relative expression of *E.coli* proteins was measured as a function of cadmium stress. Proteins from cells grown under experimental conditions were distinguished from controls by growth of one set in rare-isotope depleted media (<sup>15</sup>N, <sup>13</sup>C, <sup>2</sup>H depleted). Protein samples were mixed in equal quantities prior to analysis and relative expression monitored by relative intensities of normal and light protein signals. The experimental protocol was then reversed to illuminate any bias caused by growth in nutrient-depleted media.<sup>[29]</sup> This protocol could also be modified to a pulse-chase experiment to measure protein turnover rates. A similar method was employed to measure relative expression of yeast proteins as well as relative phosphorylation levels in mutants and controls.<sup>[30]</sup> Both of these techniques suffer from the

disadvantage that cells must be grown in isotope-depleted media, a factor that limits the applicability of the method.

An alternative strategy is to modify control and experimental protein samples *after* extraction of proteins. This was accomplished through use of isotope-encoded affinity tag technology (ICAT).<sup>[31]</sup> Chemically identical tags with a thiol-directed reactivity at one end and biotin at the other are synthesized with one batch being deuterium labeled. Parallel protein samples are tagged, pooled in identical quantities and digested with trypsin. Then the biotin tag is used to isolate just the tagged peptides via an affinity procedure, greatly simplifying the complexity of the mixture to be analysed. Once again the mass spectrometer distinguishes the relative expression of a peptide (and thus the protein it was derived from) via light and heavy tags. The technique ignores proteins that lack cysteine and others whose thiols are not accessible for reaction, though alternative specificities can be envisaged. Quantitative proteomics and mass spectrometry have been extensively reviewed.<sup>[32-37]</sup>

### 2.3 Determination of Protein Structures

To obtain high-resolution protein structures of reliable quality on an industrial scale, new technologies are needed to keep up with the pace of DNA sequencing. In principal a high-resolution protein structure can be obtained either by x-ray crystallography or by solution nuclear magnetic resonance (NMR) methods. Both techniques are rather laborious and work reliably only in the case of small to medium size soluble proteins. No reliable technology is available for membrane proteins and so far only about 20 have been crystallized successfully. Therefore other techniques such as Fourier transform infrared (FTIR) spectroscopy, electron paramagnetic resonance (EPR) spectroscopy or intramolecular cross-linking are being used which usually require purification of the protein in milligram quantities.<sup>[38-41]</sup>

An impressive example of a pioneering and straightforward approach to obtain real proteomic structure information has been established at the University of California Los Angeles, demonstrating the resources needed for this work (see URL: <http://www.doe-mbi.ucla.edu/TB/>). In a consortium approach integrating a few dozen renowned international laboratories with more than a hundred specialized scientists the proteome of *Mycobacterium tuberculosis* will be characterized on the structural, functional and interacting level to 'provide a foundation for a fundamental understanding of biology'. With the goal of high-throughput protein structure determination on the atomic level, the current situation in proteomics is reminiscent of the early attempts to obtain genomic information.

## 2.4 Data Management and Evaluation

To create protein/protein interaction maps and to identify signaling cascades, individual interacting proteins need to be identified as the building blocks for these networks. Both computational and experimental approaches have been harnessed. Using innovative strategies involving phylogenetic profiling, correlated evolution, correlated messenger RNA expression patterns and patterns of domain fusion, the function of many proteins can be predicted (reviewed in Eisenberg et al.<sup>[42]</sup>). Marcotte et al.<sup>[43]</sup> used such methods to assign function to over half of 2557 previously uncharacterized yeast proteins. Computation is also used extensively for protein structural modeling where solved structures are unavailable (reviewed in Sanchez et al.<sup>[44]</sup>), contributing greatly to understanding function and prediction of interactions.

Experimental approaches to assembling large-scale interaction maps have focused on yeast two-hybrid technology in which only interacting proteins activate a genetic switch which is being used as a reporter for interaction. Impressive information has been obtained using two-hybrid technology and, for yeast, complex maps of interactions have been published.<sup>[45-47]</sup> An alternative approach involves biochemical isolation of protein complexes and characterization of proteins by mass spectrometry.<sup>[14,48]</sup> While this may not necessarily be a high-throughput approach it may be the most effective way to completely describe a proteome due to the sensitivity benefits for detection of minor species and low abundance proteins. Both technologies may become less often used as more protein structures are being made available and docking studies will be performed *in silico* as a large-scale approach.

## 2.5 Expression Patterns, Microarray Data, DNA Chips and Protein Chips

Gene expression profiling is achieved using DNA microarrays (DNA chips) that allow high throughput measuring of transcript abundance across a broad range of genes (the transcriptome). However, the frequent lack of agreement between mRNA levels and protein accumulation<sup>[49]</sup> has led many researchers toward the proteomics protocols described above. DNA chips do have a very important role to play in monitoring the functional status of the least abundant elements of the proteome itself, that is, transcription factors.

Though most practical proteomics is being done with 2D gels and mass spectrometry, there are protein chip technologies under development. One of the first is surface-enhanced laser desorption ionization (SELDI) which is showing potential for discovery of novel cancer biomarkers.<sup>[50]</sup> Specially modified surfaces are used to select subsets of proteins from complex mixtures prior to

mass spectrometric analysis of intact proteins. Intact mass tags are then compared across experimental samples to detect changes in abundance. Actual identification of biomarkers then requires further proteomics experiments as described above. Antibodies will likely be incorporated into protein-chip design and an array of 27648 recombinant human-brain proteins were screened in a single experiment, illustrating the potential strength of this technique.<sup>[51]</sup> Many other designs of protein chips are under development and the near future will bring devices to measure protein/biomolecule interactions<sup>[52]</sup> or activities of specific subsets of a proteome, for example, kinases.<sup>[53]</sup>

## 3. Proteomics, Drugs and Therapy

Powerful leads for the treatment of human disease are anticipated from large-scale proteomic research that identifies suitable targets. However, as we learn from genomics, the mere discovery of a new disease related gene normally does not contribute substantially to the development of successful drugs, especially since DNA rarely serves as a good target. Moreover, the availability of the human genome does not automatically deliver disease related genetic information.<sup>[1,2]</sup> This can be found systematically only after determining several human genomes, that is, sequencing of different phenotypes. In the future, expression profiling through DNA-arrays or sequencing of genomes from different age groups, races and tissues of patients will help to identify disease-related genes. Proteomics technology can also be used as a diagnostic tool to monitor identity and abundance of individual proteins in body fluids and tissues at a given time. Certainly, proteomics will contribute to the understanding of dynamic interactions between proteins in the living cell, however it will not necessarily provide therapies for diseases. In other words, proteome research will help to ask the right questions but will not automatically give the right answers.

In doing research on the human proteome, technical and ethical problems might arise due to the complexity of the human body and the potential need to identify and purify a sparsely available target protein. Fortunately, one important lesson that has been learned through genomic research is that for almost every gene found in humans a direct homolog can be found in mice. It is reasonable therefore to extrapolate these homologies in DNA sequences to homologies in protein structure, that is, identical folds and similar functions for the human and murine system. Indeed, the driving force behind sequencing the mouse genome as the first mammalian genome is exactly this high degree of homology.<sup>[54,55]</sup>

The main goal of proteomics in drug development is to understand the cellular machinery sufficiently to identify suitable targets against which a selective and efficient lead compound and

finally a successful drug can be synthesized. Amid all the progress in data acquisition for sequences, structures and interactions – once the target has been identified the major work remains: the screening of large candidate compound libraries in conjunction with clever medicinal chemistry that guarantees selective action and defined delivery of the drug.

#### 4. Conclusion

Along with the growing knowledge of genomics, substantial efforts are underway to understand the encoded information – a challenge mainly conceived as proteomics. In the future similar fields will develop which inevitably end in ‘-omics,’ such as metabolomics for the study of small metabolites or lipidomics for lipids and glycomics for sugars. The trend is to perform molecular life science in a general and global array approach that tries to integrate all molecules of a given entity using the power of modern computation. Certainly this is in contrast to traditional biochemical research aimed at in-depth exploration of individual molecules, which is the only way to obtain information about structures, functions and mechanisms. However, to make a simple analogy, if we only look at the individual brickstones of a house we do not obtain any information about the rooms, the windows and the doors. Therefore we might be able to obtain previously inconceivable insights into biology by using proteomics in the framework of careful hypothesis-driven research.

#### Acknowledgements

On the occasion of his 65th birthday we dedicate this article to H. R. Kaback to honor his numerous and invaluable contributions to membrane biochemistry!

#### References

1. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* 2001; 409: 860-921
2. Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. *Science* 2001; 291: 1304-51
3. Black DL. Protein diversity from alternative splicing: a challenge for bioinformatics and post-genome biology. *Cell* 2000; 103: 367-70
4. O'Farrell PH. High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* 1975; 250: 4007-21
5. Banks RE, Dunn MJ, Hochstrasser DF, et al. Proteomics: new perspectives, new biomedical opportunities. *Lancet* 2000; 356: 1749-56
6. Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* 1988; 60: 2299-301
7. Fenyo D. Identifying the proteome: software tools. *Curr Opin Biotechnol* 2000; 11: 391-5
8. Clauser KR, Baker P, Burlingame AL. Role of accurate mass measurement ( $\pm 10$  ppm) in protein identification strategies employing MS or MS/MS and database searching. *Anal Chem* 1999; 71: 2871-82
9. Binz PA, Muller M, Walther D, et al. A molecular scanner to automate proteomic research and to display proteome images. *Anal Chem* 1999; 71: 4981-8
10. Wilkins MR, Gasteiger E, Gooley AA, et al. High-throughput mass spectrometric discovery of protein post-translational modifications. *J Mol Biol* 1999; 289: 645-57
11. Anderson NG, Matheson A, Anderson NL. Back to the future: the human protein index HPI and the agenda for post-proteomic biology. *Proteomics* 2001; 1: 3-12
12. Fenn JB, Mann M, Meng CK, et al. Electrospray ionization for mass spectrometry of large biomolecules. *Science* 1989; 246: 64-71
13. le Coutre J, Whitelegge JP, Gross A, et al. Proteomics on full-length membrane proteins using mass spectrometry. *Biochemistry* 2000; 39: 4237-42
14. Whitelegge JP, Gundersen CB, Faull KF. Electrospray-ionization mass spectrometry of intact intrinsic membrane proteins. *Protein Sci* 1998; 7: 1423-30
15. Whitelegge JP, Penn B, To T, et al. Methionine oxidation within the cerebrosidase-sulfate activator protein (CSAct or Saposin B). *Protein Sci* 2000; 9: 1618-30
16. Loo JA, Brown J, Critchley G, et al. High sensitivity mass spectrometric methods for obtaining intact molecular weights from gel-separated proteins. *Electrophoresis* 1999; 20: 743-8
17. Jensen ON, Houthaeve T, Shevchenko A, et al. Identification of the major membrane and core proteins of vaccinia virus by two-dimensional electrophoresis. *J Virol* 1996; 70: 7485-97
18. Gygi SP, Corthals GL, Zhang Y, et al. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc Natl Acad Sci U S A* 2000; 97: 9390-5
19. Corthals GL, Wasinger VC, Hochstrasser DF, et al. The dynamic range of protein expression: a challenge for proteomic research. *Electrophoresis* 2000; 21: 1104-15
20. Taylor RS, Wu CC, Hays LG, et al. Proteomics of rat liver Golgi complex: minor proteins are identified through sequential fractionation. *Electrophoresis* 2000; 21: 3441-59
21. Holt LJ, Enever C, de Wildt RM, et al. The use of recombinant antibodies in proteomics. *Curr Opin Biotechnol* 2000; 11: 445-9
22. Comisarow MB, Marshall AG. Fourier transform ion cyclotron resonance spectroscopy. *Chem Phys Lett* 1974; 25: 282-3
23. Kelleher NL, Zubarev RA, Bush K, et al. Localization of labile posttranslational modifications by electron capture dissociation: the case of gamma-carboxyglutamic acid. *Anal Chem* 1999; 71: 4250-3
24. Kelleher NL, Lin HY, Valaskovic GA, et al. Top down versus bottom up protein characterization by tandem high-resolution mass spectrometry. *J Am Chem Soc* 2001; 121: 806-12
25. Blanco DR, Whitelegge JP, Miller JN, et al. Demonstration by mass spectrometry that purified native *Treponema pallidum* rare outer membrane protein 1 (Tromp1) has a cleaved signal peptide. *J Bacteriol* 1999; 181: 5094-8
26. Whitelegge JP, Jewess P, Pickering MG, et al. Sequence analysis of photoaffinity-labelled peptides derived by proteolysis of photosystem-2 reaction centres from thylakoid membranes treated with [ $^{14}$ C]azidoatrazine. *Eur J Biochem* 1992; 207: 1077-84
27. Whitelegge JP, le Coutre J, Lee JC, et al. Toward the bilayer proteome, electrospray ionization-mass spectrometry of large, intact transmembrane proteins. *Proc Natl Acad Sci U S A* 1999; 96: 10695-8
28. Turk E, Kim O, le Coutre J, et al. Molecular characterization of *Vibrio parahaemolyticus* vSGLT: a model for sodium-coupled sugar cotransporters. *J Biol Chem* 2000; 275: 25711-6
29. Jensen PK, Pasa-Tolic L, Anderson GA, et al. Probing proteomes using capillary isoelectric focusing-electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Anal Chem* 1999; 71: 2076-84
30. Oda Y, Huang K, Cross FR, et al. Accurate quantitation of protein expression and site-specific phosphorylation. *Proc Natl Acad Sci U S A* 1999; 96: 6591-6
31. Gygi SP, Rist B, Gerber SA, et al. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 1999; 17: 994-9
32. Blackstock WP, Weir MP. Proteomics: quantitative and physical mapping of cellular proteins. *Trends Biotechnol* 1999; 17: 121-7
33. Mann M. Quantitative proteomics? *Nat Biotechnol* 1999; 17: 954-5
34. Chalmers MJ, Gaskell SJ. Advances in mass spectrometry for proteome analysis. *Curr Opin Biotechnol* 2000; 11: 384-90

35. Gygi SP, Rist B, Aebersold R. Measuring gene expression by quantitative proteome analysis. *Curr Opin Biotechnol* 2000; 11: 396-401
36. Pandey A, Mann M. Proteomics to study genes and genomes. *Nature* 2000; 405: 837-46
37. Patterson SD. Proteomics: the industrialization of protein chemistry. *Curr Opin Biotechnol* 2000; 11: 413-8
38. le Coutre J, Kaback HR. Structure function relationships of integral membrane proteins: membrane transporters vs channels. *Biopolymers Peptide Sci* 2001; 55: 297-307
39. Voss J, Hubbell WL, Kaback HR. Distance determination in proteins using designed metal ion binding sites and site-directed spin labeling: application to the lactose permease of *Escherichia coli*. *Proc Natl Acad Sci U S A* 1995; 92: 12300-3
40. Frillingos S, Sahin-Toth M, Wu J, et al. Cys-scanning mutagenesis: a novel approach to structure function relationships in polytopic membrane proteins. *FASEB J* 1998; 12: 1281-99
41. Fleming KG. Riding the wave: structural and energetic principles of helical membrane proteins. *Curr Opin Biotechnol* 2000; 11: 67-71
42. Eisenberg D, Marcotte EM, Xenarios I, et al. Protein function in the post-genomic era. *Nature* 2000; 405: 823-6
43. Marcotte EM, Pellegrini M, Thompson MJ, et al. A combined algorithm for genome-wide prediction of protein function. *Nature* 1999; 402: 83-6
44. Sanchez R, Pieper U, Melo F, et al. Protein structure modeling for structural genomics. *Nat Struct Biol* 2000; 7 Suppl.: 986-90
45. Ito T, Tashiro K, Muta S, et al. Toward a protein-protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. *Proc Natl Acad Sci U S A* 2000; 97: 1143-7
46. Schwikowski B, Uetz P, Fields S. A network of protein-protein interactions in yeast. *Nat Biotechnol* 2000; 18: 1257-61
47. Uetz P, Giot L, Cagney G, et al. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 2000; 403: 623-7
48. Shevchenko A, Zachariae W. A strategy for the characterization of protein interaction networks by mass spectrometry. *Biochem Soc Trans* 1999; 27: 549-54
49. Anderson L, Seilhamer J. A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* 1997; 18: 533-7
50. Vastag B. Detection network gives early cancer tests a push. *J Natl Cancer Inst* 2000; 92: 786-8
51. Holt LJ, Bussow K, Walter G, et al. By-passing selection: direct screening for antibody-antigen interactions using protein arrays. *Nucleic Acids Res* 2000; 28: E72
52. Ge H. UPA, a universal protein array system for quantitative detection of protein-protein, protein-DNA, protein-RNA and protein-ligand interactions. *Nucleic Acids Res* 2000; 28: e3
53. Zhu H, Klemic JF, Chang S, et al. Analysis of yeast protein kinases using protein chips. *Nat Genet* 2000; 26: 283-9
54. O'Brien SJ, Menotti-Raymond M, Murphy WJ, et al. The promise of comparative genomics in mammals. *Science* 1999; 286: 458-81
55. Broder S, Venter JC. Whole genomes: the foundation of new biology and medicine. *Curr Opin Biotechnol* 2000; 11: 581-5

---

Correspondence and offprints: Dr. Johannes le Coutre, Nestlé Research Center, P. O. Box 44, CH-1000 Lausanne 26, Switzerland.  
E-mail: johannes.le-coutre@rdls.nestle.com