Proteomics in drug discovery

Jack H. Wang and Rodney M. Hewick

The promise of genomics has dramatically altered the way drug discovery is now viewed. Overshadowed by the exuberance for genomics are the observations that most disease processes and treatments are manifest at the protein level and that there may not be a good correlation between gene expression and protein expression. An alternative and complementary approach to genomics is protein expression profiling – proteomics. The authors describe the technology, its advantages and some applications.

The advent of the Human Genome Project has spawned a new industry – genomics – with numerous companies hoping to capitalize on the promise of this available information for drug discovery. With less of a fanfare, an alternative and complementary approach to gene expression analysis – proteomics – has emerged in the past few years. Based on the recently coined term, proteome[1], which was defined as the entire protein complement expressed by a genome or by a cell or tissue type, the aim of proteomics is to profile the proteins from a cellular or tissue source. As with genomics, when this is performed in a comparative manner between diseased and healthy samples, the hope is that new drugs or drug targets will be identified.

Value of proteomics

This new field of proteomics has seen a rapid increase in interest and investment by companies ranging from large pharmaceutical companies to specialized start-up companies. The basis for this interest stems from (1) the belief that gene-based expression analysis alone will not be comprehensive and, in certain cases, will be totally inadequate for drug discovery, and (2) the compelling observation that most disease processes and disease treatments are manifest at the protein level. To be convinced of this latter point, one only needs to look at the list of top-selling pharmaceuticals and confirm that their mechanisms of action are mediated through proteins.

Genomics was enthusiastically embraced because of the sensitivity and high-throughput nature of the technology and the availability of multiple methods for expression analysis, such as serial analysis of gene expression (SAGE)[2], differential display[3], oligonucleotide array technology[4] and cDNA microarrays[5]. Also critical to genomics is the widespread belief that protein expression can be inferred through the presence of its mRNA. However, a recent study by Anderson and Seilhamer[6] offers evidence that there is not necessarily a tight correlation between transcript and protein levels. Therefore, the expression of a gene has no definitive relationship to the ultimate expression or abundance of its protein product, emphasizing the need to also profile protein expression in order to understand disease at the molecular level. This lack of correlation can be attributed to:

- Temporal differences between gene expression and actual protein expression
- Spatial differences between the site of gene expression and where protein products act
- Differential stability and turnover of mRNA versus protein product
- Post-transcriptional splicing of the mRNA to yield various protein products

Of equal importance is the total inability of nucleic acid sequence and expression to be used for the determination of post-translational modifications (PTM), such as post-translational processing, phosphorylation and glycosylation. As an example, signal transduction involves complex protein cascades with phosphorylation a crucial

Jack H. Wang* and **Rodney M. Hewick**, Genetics Institute Inc., 87 Cambridgepark Drive, Cambridge, MA 02140, USA. *tel: +1 617 498 8110, fax: +1 617 498 8878, e-mail: jwang@genetics.com

 1359-6446/99/$ – see front matter PII: S1359-6446(99)01306-9

component of the process. As cells use this mechanism for immediate responses to external stimuli, no involvement of gene expression is necessary. These PTMs play important roles in the structure, activity and compartmentalization of proteins, and have been shown to be involved in disease states. Only a protein-based approach will be able to detect these changes.

The ability to study protein relationships directly is another unique advantage of proteomics. This ability will carry more significance in the future, as only 2% of diseases are believed to be monogenic[7]. The challenge now is to understand how the networks of proteins are related and lead to the other 98% of diseases. Proteomics is ideally suited for this purpose because it looks directly at the protein level and also detects PTMs, which undoubtedly play a role in protein–protein interactions.

A specific subset of proteomics, which deals with protein–protein interactions, is protein complex identification. Examples of this include:

- Isolation and identification of heterodimers that cannot be easily inferred from transcript information
- Identification of ligand–receptor interaction or protein–protein binding through the use of Biacore technology
- Identification of components of protein complexes such as the spliceosome and ribosome

Basic technology

Currently, the prevailing operational definition for proteomics is the combined use of mass spectrometry (MS) with two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) analysis to study gene expression at the protein level. At this time, 2D-PAGE is still the most comprehensive, quantitative, high-resolution method for displaying proteins, while MS is the most sensitive method available to identify proteins.

2D-PAGE technology has improved dramatically over the past five years. Since the introduction by Pharmacia Biotech (Uppsala, Sweden) of Immobiline focusing strips, in which the pH gradient is covalently linked to the gel, 2D gels have been shown to be extremely reproducible. 2D gels are capable of resolving thousands of proteins (including isoforms and post-translationally modified proteins) and, when stained with silver or fluorescent dyes, provide a sensitive method for quantitating protein expression. Sample loading advances[8,9] and larger gel formats[9] have increased protein capacity to milligram levels, and improved sample extraction techniques[10,11] are increasing the representation of total cellular protein loaded onto the gel. Recent improvements in sample handling and instrumentation by Amersham Pharmacia Biotech (Uppsala, Sweden) now allow for more simplified sample loading and focusing.

MS has also been advancing at a rapid pace to position itself as the fundamental tool for high-throughput, high-sensitivity protein sequence analysis. Matrix-assisted laser desorption ionization time of flight (MALDI-TOF) mass spectrometry has been adapted for high-throughput peptide mass fingerprinting for protein identification[12–16]. This technique is ideal as an initial screen of 2D gel spots to identify known proteins from the ever-expanding databases. More definitive identifications based on peptide sequence or identification of protein modifications can be achieved with tandem mass spectrometry (MS/MS)[17]. Triple quadrupole or ion trap mass spectrometers have been used for generating fragmentation data of peptides for database searches or sequence analysis[18–20]. The modification of the electrospray ion source to a micro- or nanospray configuration has increased the sensitivity by at least an order of magnitude[21,22]. Consistent, successful identification of faint silver-stained 2D gel spots at low femtomole levels is now attainable.

As is the case with genomics, the availability of an increasingly comprehensive database has been a boon to proteomics. The improvements in accuracy and resolution of mass spectrometers have also helped to fuel the rapid rise of proteomics. A major component in the growth of proteomics not to be overlooked has been the development of database searching software (reviewed in Refs 23,24). These algorithms are capable of using various types of MS data to interrogate databases and yield the critical identifications of 2D gel spots in a high-throughput, automated manner.

Figure 1 provides a generalized overview of the sequence of steps involved in a typical proteomics procedure.

Applications of proteomics technology

Proteomics has unique and significant advantages as an important complement to a genomics approach, and applications for this technology are readily apparent.

Target/marker identification

This application of proteomics provides a protein profile of a cell or tissue that can be used to compare a healthy with a diseased state for protein differences in the search for drugs or drug targets. This is the most popularized of the applications for proteomics; it has the added unique advantage that bodily fluids can be used for profiling.

Identification of diagnostic markers for disease states such as cancer by a differential proteome profiling of diseased

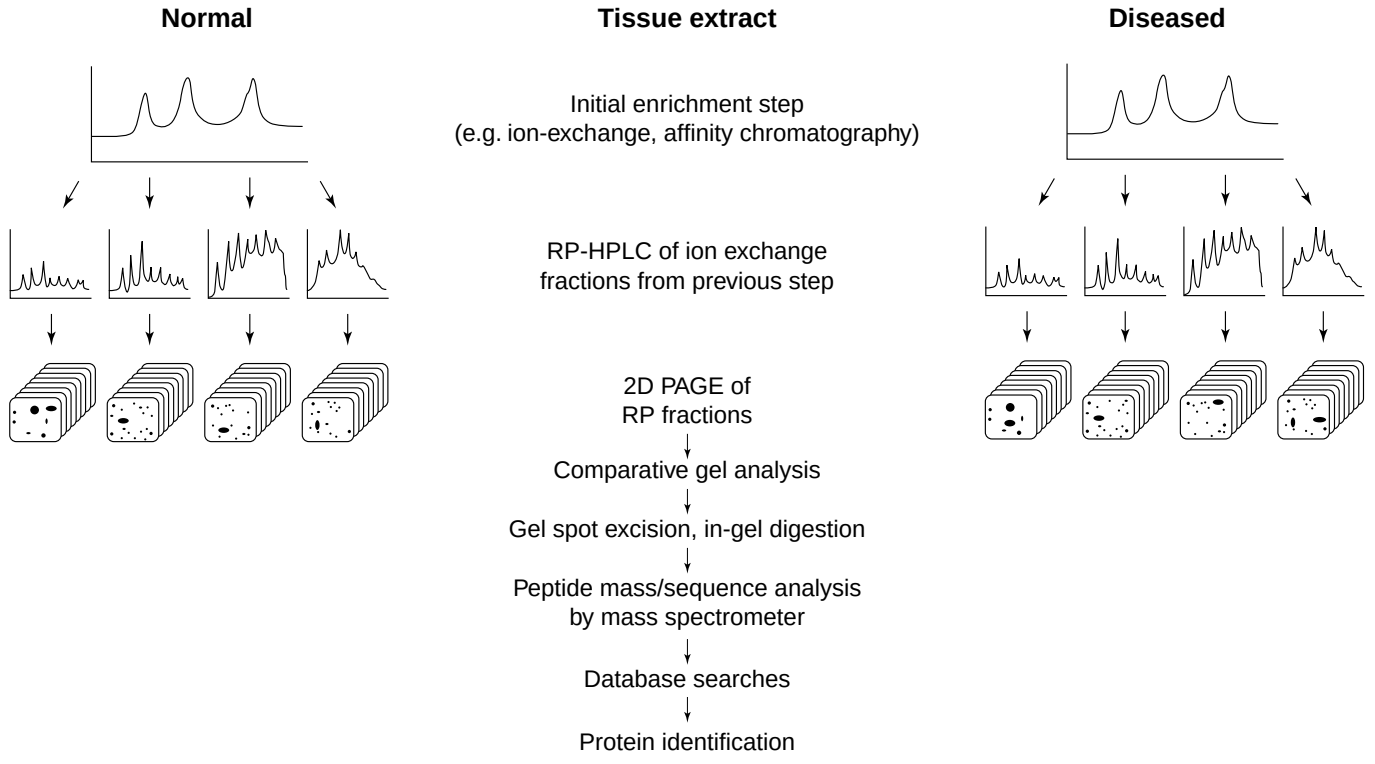

Figure 1. *Generic proteomics scheme. The number of two-dimensional (2D) gels that can be managed will determine the extent of fractionation at each step. The rapid generation of multiple 2D gels, as a result of added enrichment steps and associated fractionation, highlights the need for high-throughput automation of all aspects of gel spot difference analysis (i.e. gel handling, staining, excision, digestion and MS analysis).*

and normal tissue is a subset of target identification. An example of the successful use of this technology for marker identification comes from the research of Julio Celis of the Danish Centre for Human Genome Research (Aarhus, Denmark). His laboratory has identified a biomarker, psoriasin, from the urine of patients with squamous cell carcinoma (SCC) of the bladder[25]. This protein is normally not found in cells of the urinary tract but has now been confirmed by immunohistochemical analysis to be detectable in some cells of bladder SCC.

Target validation/toxicology

Proteomics has an application as an assay for the potential utility of drug candidates. This can be achieved by a comparative analysis of reference protein profiles from normal or disease states with profiles after drug treatment. Proteomics technology could also be integrated with combinatorial chemistry to evaluate comparative structure–activity relationships of drug analogs. These applications could accelerate identification and optimization of lead candidates for clinical development. Large Scale Biology (Rockville, MD, USA) has an active program to study the effects of drugs on protein expression. The 2D gel data have shown a correlation between protein changes and the mechanism of action of the drugs. As an example, lovastatin – a cholesterol lowering agent – was shown to affect proteins involved with cholesterol metabolism[26].

A variant of target validation is the use of proteomics to study the toxicity of drugs. A comparison of the protein profiles from normal tissue or tissue treated with the known toxic agent might give an indication of the drug's toxic activity. An excellent example of the use of proteomics for toxicology was the study of the toxic activity of cyclosporine A (CsA) in kidney by Sandra Steiner (Novartis, Basel, Switzerland)[27]. 2D gel profiles of rat kidney proteins with or without treatment with CsA were compared. One of the downregulated protein spots with CsA treatment was identified as calbindin. This protein is found in the tubules of kidney and is involved in calcium binding and transport. There is now good evidence that the toxic effect of CsA

with respect to intratubular calcification is linked to the decrease of calbindin. This is a clear case of where 2D-gel analysis of kidney tissue gave new insights into the side effects of CsA. As with target validation, a database of the protein profiles after treatment with known toxic agents can be used as a reference for comparative purposes when investigating the side effects of new drugs.

Protein–protein interactions

This use of proteomics was discussed earlier. There is no substitute for dealing directly with proteins when studying their interactions and complexes. The multiple components of spliceosomes and anaphase-promoting complexes have been identified through the use of the latest proteomics technology[28,29]. The combination of high-sensitivity Biacore affinity purification and MS analysis has been shown to be an ideal combination for identification of specific protein–protein interactions at nanogram levels[30,31]. The identification of the heterodimeric subunits of bone morphogenetic proteins (BMP) can be accomplished more directly with a proteomics approach (Ref. 32; J. Wang and coworkers, unpublished).

Alternative proteomics approaches

Even with the recent improvements in 2D gel and MS technology, proteomics is still viewed as a difficult technology relative to genomics because of the complexity of the proteome, the lack of a high-throughput scheme for amplifying and detecting low abundance proteins and the relative differences in protein solubility. Not surprisingly, alternatives to this technology (particularly to replace the use of 2D gels) are being explored in the hope of making the protein approach more competitive with genomics.

One approach that omits 2D gels is the use of multidimensional chromatography to partially resolve mixtures of proteins or their peptide digest products[33–35]. The resolving power of the mass spectrometer can be used to identify the components of the simplified mixture. This technique has been used on fully sequenced and less complex systems such as bacteria and yeast. Potentially, this approach avoids losses associated with 2D-gel technology but loses the quantitative information gained from 2D-gel imaging.

A popular alternative for the screening and identification of proteins has always been the use of antibodies. Cambridge Antibody Technology (Melbourn, UK) has generated an extensive library of antibodies with their phage antibody display system[36]. This library can then be probed with a peptide or protein of interest to isolate the specific antibody to be used for screening purposes. The goal with this technology is to construct a database of protein expression information for the ligands of interest, based on the antibodies from the library. The applications described earlier will be relevant to this approach as well.

The yeast two-hybrid system is another alternative to studying protein expression with emphasis on protein–protein interactions. From the initial observation that an interaction of two proteins, each fused to a separate domain of a transcription factor, allows for the transcription of a reporter gene[37], this technology has expanded with the development of new variations and applications[38–40]. The two-hybrid system is taking a proteomic view towards expression analysis, but uses a genomic approach. The high-throughput screening of genomic libraries to study protein–protein interactions is an advantage, but the lack of direct analysis of protein interactions is evident from the false positives observed with this technology. Because of the high-throughput nature of this screen, the potential for constructing linkage maps of protein networks[41], and the impact this information would have for drug targets and targeting, this technology must be considered a viable alternative for studying protein–protein interactions.

Another alternative approach to 2D gels would be a protein version of the DNA chip technology. This lofty goal is far from being realized because of the challenges associated with the obvious differences between nucleic acids and proteins. Ciphergen Biosystems (Palo Alto, CA, USA) has made some early strides in this area with their surface-enhanced laser desorption/ionization (SELDI)[42] and ProteinChip technology. These chips have addressable sites for affinity binding of proteins from crude samples and subsequent detection of bound protein by MS.

The future of proteomics

There is a growing realization of the importance of proteomics in the life sciences industry. This is evident by the increasing investment in this area by pharmaceutical and genomics companies and the proliferation of numerous companies dedicated to proteomics or to hardware and software support of proteomics technologies. This increased attention and investment should fuel the continued rapid advance of proteomics technology. Some of the advances, which can be anticipated or are being realized already, are in the areas of sample handling, automation for high-throughput analysis, miniaturization[43,44], mass spectrometer development and bioinformatics. Improvements in sample solubilization, subcellular fractionation, protein detection and the commercialization of narrower range pH gradients should improve the sensitivity, resolution and representation of proteins using 2D gels[45]. MS technology has been continually improving in sensitivity and ease of use. It is

anticipated that the next generation of this technology will see another order of magnitude increase in sensitivity. Attempts have been made to directly integrate the resolution of polyacrylamide gels with the sensitivity of MS by analyzing proteins or their *in situ* digested peptides directly from gels or blots by MALDI-MS (Refs 46,47). The development of this technique vastly increases the potential for automation and high-throughput protein identification from gels while minimizing gel processing by having a direct interface between gels and mass spectrometer. As this new technique is being developed, several laboratories around the world have built high-throughput 2D-gel analysis platforms that integrate imaging of 2D gel spots with their excision, in-gel digestion and analysis by MS. Continued improvements in automating all aspects of this process, particularly for dealing with multiple gels in parallel, will be necessary if the goal is for proteomics to be considered high-throughput. 2D gel electrophoresis, staining and image analysis are just some of the steps that remain to be fully automated before this process can be truly called high-throughput. Finally, bioinformatics will also need to keep pace with these other improvements. The next generation of software for data collection, filtering, analysis, management and database searching will have to be developed. All the 2D gel images, mass spectra and sequence data will require archiving into a relational database so it can ultimately be integrated with genomic data and data from other sources available over the World Wide Web.

An objective evaluation will lead to the conclusion that proteomics is a necessary component to any complete functional genomics program. The advantages and importance of directly analyzing proteins make a strong argument for the value of proteomics. The shortcomings of throughput and sensitivity (e.g. the lack of a protein equivalent for the polymerase chain reaction) highlight the need for improved automation, enrichment and detection methods. Otherwise, proteomics will continue to be viewed as technically demanding and able to detect only abundant proteins. These challenges will undoubtedly be attacked and overcome with the increased awareness of the importance of proteomics.

REFERENCES

1 Wasinger, V.C. *et al.* (1995) *Electrophoresis* 16, 1090–1094
2 Velculescu, V.E. *et al.* (1995) *Science* 270, 484–487
3 Liang, P. and Pardee, A.B. (1992) *Science* 257, 967–971
4 Lockhart, D.J. *et al.* (1996) *Nat. Biotechnol.* 14, 1675–1680
5 Schena, M. *et al.* (1995) *Science* 270, 467–470
6 Anderson, L. and Seilhamer, J. (1997) *Electrophoresis* 18, 533–537
7 Strohman, R. (1994) *Nat. Biotechnol.* 12, 156–164
8 Rabilloud, T., Valette, C. and Lawrence, J.J. (1994) *Electrophoresis* 15, 1552–1558
9 Sanchez, J-C. *et al.* (1997) *Electrophoresis* 18, 324–327
10 Rabilloud, T. *et al.* (1997) *Electrophoresis* 18, 307–316
11 Molloy, M.P. *et al.* (1998) *Electrophoresis* 19, 837–844
12 Henzel, W.J. *et al.* (1993) *Proc. Natl. Acad. Sci. U. S. A.* 90, 5011–5015
13 James, P. *et al.* (1993) *Biochem. Biophys. Res. Commun.* 195, 58–64
14 Mann, M., Hojrup, P. and Roepstorff, P. (1993) *Biol. Mass Spectrom.* 22, 338–345
15 Pappin, D., Hojrup, P. and Bleasby, A. (1993) *Curr. Biol.* 3, 327–332
16 Yates, J.R., III *et al.* (1993) *Anal. Biochem.* 214, 397–408
17 Hunt, D.F. *et al.* (1986) *Proc. Natl. Acad. Sci. U. S. A.* 83, 6233–6237
18 Hunt, D.F. *et al.* (1992) *Science* 255, 1261–1263
19 Eng, J.K., McCormack, A.L. and Yates, J.R., III (1994) *J. Am. Soc. Mass Spectrom.* 5, 976–989
20 Mann, M. and Wilm, M. (1994) *Anal. Chem.* 66, 4390–4399
21 Wilm, M. and Mann, M. (1996) *Anal. Chem.* 68, 1–8
22 Davis, M.T. *et al.* (1995) *Anal. Chem.* 67, 4549–4556
23 Yates, J.R., III (1998) *Electrophoresis* 19, 893–900
24 James, P. (1997) *Q. Rev. Biophys.* 30, 279–331
25 Celis, J.E. *et al.* (1996) *J. Urol.* 155, 2105–2112
26 Anderson, N.L. *et al.* (1991) *Electrophoresis* 12, 907–930
27 Steiner, S. *et al.* (1996) *Biochem. Pharmacol.* 51, 253–258
28 Gottschalk, A. *et al.* (1998) *RNA* 4, 374–393
29 Zachariae, W. *et al.* (1998) *Science* 279, 1216–1219
30 Fitz, L. (1998) *IBC Conference on Biosensor Technologies: Applications for Pharmaceutical and Biotechnology Discovery*, 14–15 May, Boston, MA, USA
31 Krone, J.R. *et al.* (1997) *Anal. Biochem.* 244, 124–132
32 Israel, D.I. *et al.* (1996) *Growth Factors* 13, 291–300
33 McCormack, A.L. *et al.* (1997) *Anal. Chem.* 69, 767–776
34 Schieltz, D.M. *et al.* (1997) *45th ASMS Conference on Mass Spectrometry and Allied Topics,* 1–5 June, Palm Springs, CA, USA
35 Opiteck, G.J. *et al.* (1998) *Anal. Biochem.* 258, 349–361
36 Winter, G. *et al.* (1994) *Annu. Rev. Immunol.* 12, 433–455
37 Fields, S. and Song, O. (1989) *Nature* 340, 245–246
38 Colas, P. and Brent, R. (1998) *Trends Biotechnol.* 16, 355–363
39 Brachmann, R.K. and Boeke, J.D. (1997) *Curr. Opin. Biotechnol.* 8, 561–568
40 Young, K.H. (1998) *Biol. Reprod.* 58, 302–311
41 Evangelista, C., Lockshon, D. and Fields, S. (1996) *Trends Cell Biol.* 6, 196–199
42 Hutchens, T.W. and Yip, T-T. (1993) *Rapid Commun. Mass Spectrom.* 7, 576–580
43 Figeys, D., Ning, Y. and Aebersold, R. (1997) *Anal. Chem.* 69, 3153–3160
44 Ramsey, R.S. and Ramsey, J.M. (1997) *Anal. Chem.* 69, 1174–1178
45 Herbert, B.R., Sanchez, J-C. and Bini, L. (1997) in *Proteome Research: New Frontiers in Functional Genomics* (Wilkins, M.R., Williams, K.L., Appel, R.D. and Hochstrasser, D.F., eds), pp. 13–33, Springer
46 Ogorzalek Loo, R.R. *et al.* (1997) *Electrophoresis* 18, 382–390
47 Eckerskorn, C. *et al.* (1997) *Anal. Chem.* 69, 2888–2892