PART I

GENERAL PROTEOMICS OF MICROORGANISMS/MODEL ORGANISMS

CHAPTER 1

Holistic Biology of Microorganisms: Genomics, Transcriptomics, and Proteomics

VALERIE WASINGER

University of New South Wales, Sydney, Australia

1.1 INTRODUCTION

An organism's phenotype is determined by the environment and its genetic content, and as such understanding the relationship between genome, transcriptome, and proteome is one of the fundamental goals of biology. Over the last decade quantitative biological analysis has been performed at the genomic, transcribed, and translated level. This has occurred through large-scale DNA sequencing, genomewide genetic analysis, DNA and protein chips, and two-dimensional gel electrophoresis (2DE) or multidimensional liquid chromatography integrated with mass spectrometry (LC-MS) for the fast and highly sensitive analysis of proteins. The rapid speed of progress in genomics has been impressive and unrivaled in proteomics, with more than 160 organisms having been sequenced, including the human genome. Despite this achievement, the information provided by any given genome has given little insight into the workings of a cell. A better understanding of the cellular machinery, biology, and disease processes will ultimately result from mining these genomes for their *cognate* proteins—proteomics. This chapter aims to highlight the contribution proteomics has made to our collective knowledge of whole-organism microbiology.

The expansion of the field of proteomics from the display of large numbers of proteins using 2DE to the incorporation of new technologies and a global perspective of protein expression, requisite to genomic sciences, has refined proteomics into a robust scientific field. A focus on quantitative measurement of proteins, patterns of changes in protein expression, and protein interactions in the context of a whole cell has given the field of proteomics a new level of maturity previously not seen.

1.2 PROTEOME IN PERSPECTIVE

Unlike genome sequencing, there has been no completion of a proteome, while a proteomic endpoint remains ill defined. Despite this, the study of the total protein

---

Microbial Proteomics: Functional Biology of Whole Organisms, Edited by Ian Humphery-Smith and Michael Hecker.

complement of a genome is feasible [1–3]. What remains the biggest challenge to the identification and analysis of complex biological samples is the collective variability of all protein's physicochemical properties and a dependence on in vivo and ex vivo parameters. These parameters contribute to proteome complexity and are phenotypically manifested as a variation of relative protein abundance; modifications or truncations, for example enzymatic cleavage; altered molecular or protein interactions; complex formation or breakdown; and presence or absence of proteins. Proteome complexity can be explained from three diverse perspectives: evolutionary complexity, internal complexity, and sample complexity (addressed later).

1.2.1 Evolutionary Complexity

Complexity is often attributed to the number of base pairs in a sequence that give rise to functional genes [4]. We readily admit that as we move up the evolutionary ladder, the most complex organism of all is our own kind. Nonetheless, our genetic likeness to plants (*A. thaliana*) and worms (*C. elegans*) is striking [5]. However, annotation of a gene within a DNA sequence is not an indication that a gene is expressed or able to serve a useful metabolic or structural purpose. The "one gene–one protein" tenet has long been rejected [6, 7] because alternate gene-splicing and posttranslational modifications can result in multiple active forms of a protein. Gene numbers from theoretical proteomes of fully sequenced organisms are compared in Figure 1.1 and are also available from http://www.ebi.ac.uk/. This figure demonstrates evolutionary complexity is not governed by gene content, nor is sequence information sufficient to describe the individuality of an organism [8]. Complexity is more significantly a reflection of protein regulation [9, 10]. As we move up the evolutionary hierarchy there is a distinct increase in gene product modification: for *Mycoplasma genitalium* this has been quantified at a rate of 1.2 times more proteins than

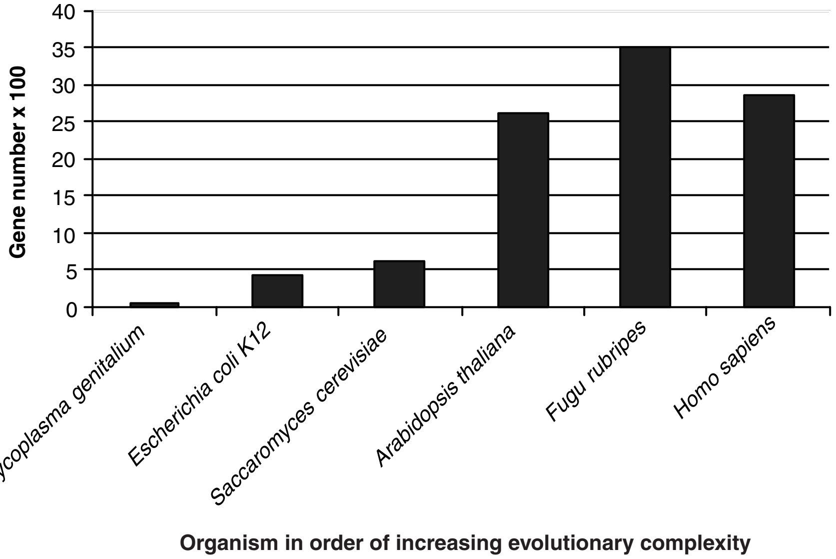

Figure 1.1 Predicted proteome of fully sequenced organisms collated from nonredundant proteome sets from SwissProt and TREMBL entries (http://www.ebi.ac.uk/).

genes [1]; for *Escherichia coli* it is 1.3 proteins per gene; for yeast, 3 proteins per gene; and for human, as many as 10 proteins per gene [8]. This has been estimated from both two-dimensional (2D) sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and tandem MS (MS/MS) work.

1.2.2 Internal Complexity

Internal complexity is a property of the network of interacting protein, transcript, and genes within a cell. As life is not just an assembly of these individual components, the study of gene function has been explored in terms of the relatively static localized gene and protein interaction networks and the massively parallel study of global networks embracing the higher order and collective behavior of genes and proteins as a proteome-scale network [11]. The study of the sum of all interacting parts in a biological system is called systems biology and addresses the network of interactions of biochemical/signaling pathways and their modulation, the influence of spatial and temporal differences in these networks, and how this knowledge can be applied to providing therapeutic targets in disease processes [12]. The recent interest in large-scale identification of functionally linked proteins can be affiliated to the development of high-throughput experimentation and computational procedures and in extensive database curation of this information. Computational approaches for finding gene and protein interactions complement and extend experimental approaches such as synthetic lethal and suppressor screens, yeast two-hybrid experiments, and high-throughput MS interaction assays [13]. Computational methods based on sequence do not assume knowledge about protein function and can therefore be used to assign function to uncharacterized proteins linked to network pathways of known function. Computational methods for studying protein associations include (i) phylogenetic profiling, based on the co-occurrence of proteins in different genomes [14]; (ii) domain fusion or “Rosetta stone” sequences [15, 16] where fused domains of a protein in one organism are used to predict interaction of these domains separated through evolution in another organism; and (iii) gene clustering in one organism where the genes have been separated by evolution in another organism [17].

The very nature of proteomics embodies all of these complexities. As a result, diverse approaches have and are being used to overcome some of the issues associated with these complexities (e.g., large protein numbers), thereby creating unique data sets. Furthermore, comparing multiple technologies then has the advantage of highlighting differences in measurements, thereby extending the proteomic coverage, but only if these approaches can be unified meaningfully.

1.3 AMALGAMATION OF TOOLS FOR A PROTEOMIC “TOOL BOX”

The field of proteomics has always been heavily reliant on protein characterization technologies. A powerful repertoire of tools has been implemented for the separation and identification of thousands of proteins simultaneously. In 1995, this consisted of Edman sequencing amino acid analysis, peptide mass fingerprinting, ladder sequencing, and expression of cloned inserts, to name a few, and these were approached in an hierarchical manner for protein identification. Success using these tools was contingent on the sensitivity of instruments for accurate mass determination [1] and available protein databases for protein identification. There was little scope for characterization of

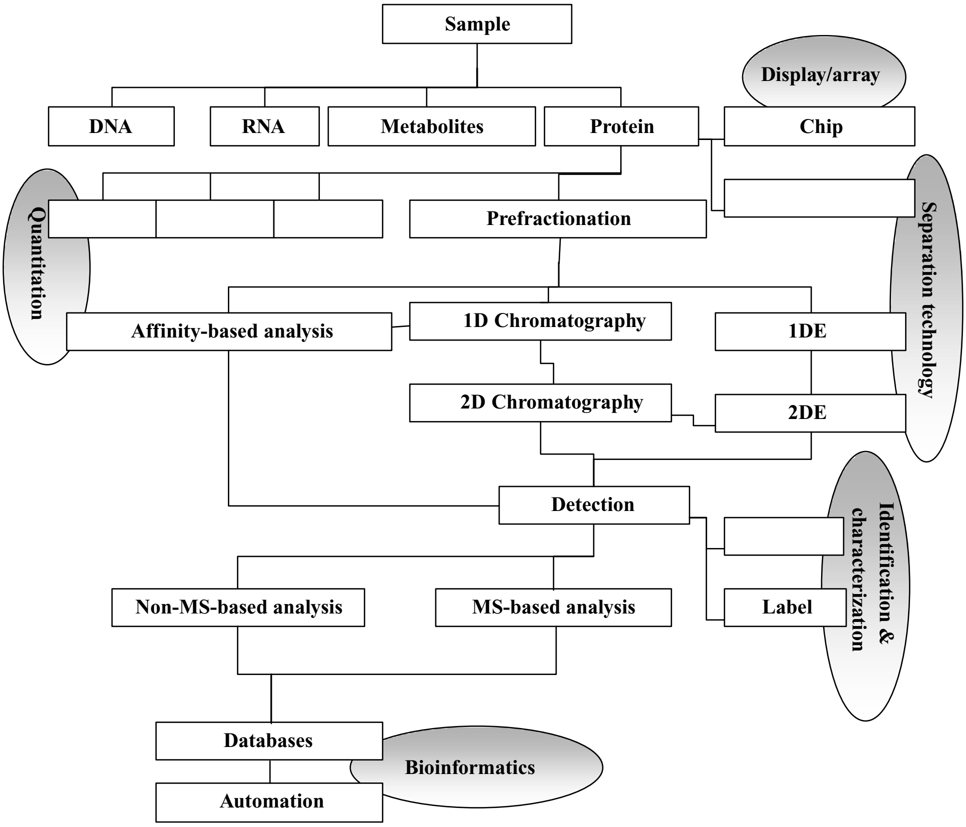

Figure 1.2 Schematic of integration of technologies for proteomics. The unification of DNA, mRNA, and metabolite information is essential for a complete understanding of biological function [58].

“unknown” proteins on a large scale. Today, the proteomic tool box has assimilated genomic, metabolomic, transcriptomic, and proteomic tools, with MS technologies still providing the backbone for much of the proteomic analysis. A representation of an integrated proteomic tool box is given in Figure 1.2.

One of the major technological challenges facing protein analysis is the dynamic range of expression of proteins within massively complex samples [18]. This difference can be as great as 12 orders of magnitude, and this is currently beyond the range of detection capable for our existing technologies. Additionally, a few highly abundant proteins making up a large percentage of total proteins can mask the detection of lower abundance and biologically significant proteins [19]. A third technological challenge, alluded to earlier, is sample complexity. The expression of a myriad of different protein species with similar physicochemical properties is compounded by the numerous modifications that can take place posttranscriptionally. Over 300 different modifications have been described for proteins [20, 21].

These on-going challenges have been addressed in a number of ways and have benefited agriculture, industry, and biomedicine. Two-dimensional gel electrophoresis for proteomic applications is still widely accepted as the technology capable of delivering the greatest separation power for highly complex samples. The expansion of pH separation using a series of overlapping narrow-range gels to assemble a contiguous collage of protein expression and allow for a more meticulous analysis of proteomes was established in 1997 for *O. anthropi* [22] and *M. genitalium*. Close to two-thirds of *M. genitalium* open reading frames (ORFs) were shown to be expressed using this approach [2]. This approach is now commonly used and bacterial examples include *Spiroplasma melliferum* [23], *Mycobacterium tuberculosis* [24], *Corynebacterium glutamicum* [25], and *Streptococcus mutans* [26].

Proteomics is also providing practical applications through the identification of immunogenic proteins as potential vaccine targets. Many hypotheses generated in silico by genomics have been validated through functional studies of the transcriptome and proteome and have led to the identification of essential genes. Using 2DE [27] and DNA microarrays [28] for differential expression analysis has confirmed that the resistance of *M. tuberculosis* to the anti-tuberculosis cell envelope drug isoniazid was the result of overexpression of components of the fatty acid synthase system, in particular AcpM and a carrier protein synthase KasA. A compilation of *M. tuberculosis* [29] and *Mycobacterium bovis* [30] induced immunoresponses in splenocytes using 2D liquid-phase electrophoresis has also identified multiple antigens associated with posttranslationally modified proteins. In 1999, Jungblut et al. [31] compared the proteomes of two nonvirulent vaccine strains of *M. bovis* BCG with two virulent strains of *M. tuberculosis* to identify protein candidates for potential vaccine development as well as for diagnostic and therapeutic purposes. Over 30 differences were identified among the strains, including three cell envelope proteins, some antigenic proteins, and novel unannotated proteins. This work, as well as work from others, has contributed to a better understanding of the pathogenic and physiological mechanism available to this organism and also contributed to the development of new vaccine candidates. *Mycobacterium bovis* BCG (live attenuated) has been the only widely used vaccine available against tuberculosis until recently, with new tuberculosis vaccine trials begun in 2004.

Two-dimensional gel electrophoresis is a powerful protein separation technique in combination with MS, yet it does not always deliver the appropriate sensitivity for the discovery of low-level proteins [32], proteins with extreme isoelectric point (pI) or mass [11], or membrane-associated proteins because of solubility issues [33]. For these reasons, the development of alternative approaches to 2DE, such as multidimensional chromatography coupled to MS, has been vital to the success of *Bifidobacterium* proteomics. *Bifidobacterium* is a gram-positive prokaryote that naturally colonizes the human gut exerting health-promoting effects. Clinical studies have claimed that bifidobacterial probiotics promote gastrointestinal tract homeostasis and health because of antidiarrheal, immunomodulating, and possibly anticarcinogenic properties. Proteomics has contributed to the comprehensive understanding of the physiological mechanisms underlying these properties [34]. A predominant portion of *Bifidobacterium infantis* proteome consists of enzymes of the glycolytic and pentose-phosphate pathways, enzymes of anaerobic metabolism, transcriptional factors, shock proteins, ribosomal proteins, and proteases. The high level of these proteins during the exponential phase of growth underlines their central role in cell survival, replication, and energy metabolism, giving an indication of the basal functions, which are essential for the vitality of the bifidobacteria biological system.

A unification of genomics, transcriptomics, and proteomics will provide a comprehensive knowledge base of gene function and a powerful reference of protein properties. There are, however, significant challenges that are being addressed but may remain unresolved until technological advances can overcome them. The goal of deciphering the entire protein complement of an organism and making heuristic relationships is formidable and often limited by cell dynamics. Understanding biological events such as modification, complexity, environmental input, and technological constraints of detection thresholds and bioinformatics will be required to overcome the challenges for a complete proteome study [1, 35].

1.4 RELATIONSHIP BETWEEN GENOME, TRANSCRIPTOME, AND PROTEOME

All life is linked by a common genetic scaffold constrained to 4 nucleotides and 20 amino acids and as such phenotypic differences require far more investigation than was first anticipated. Genomics presents only one level of functional information. Both qualitative and quantitative unique levels of information are also given by the transcriptome and proteome as well as the interactome and metabolome.

1.4.1 Transcriptome and Proteomics

Transcriptome analysis involves messenger ribonucleic acid (mRNA), the relayer of information for protein synthesis. Several methods, including serial analysis of gene expression (SAGE), oligonucleotide and complementary deoxyribonucleic acid (cDNA) microarrays, and large-scale sequencing of expressed tags, are available to measure gene expression at the mRNA level globally and quantitatively. Measurements of protein expression and transcript do not always correlate. This is due to protein abundance also being influenced by protein stability, translation rate, modulation of transcript abundance by other proteins, posttranslational modifications, and half life; therefore, mRNA cannot always be a predictor of protein abundance [36].

The absolute range of transcript abundance is largely unknown in microorganisms; however, a significant amount of work has been done in yeast. A study by Futcher et al. [37] has revealed that for each mRNA there are approximately 4000 molecules of cognate protein produced and a small number of these proteins make up at least 50% of all cellular protein in yeast. For these abundant proteins, several statistical methods have shown a correlation between mRNA and protein abundances [3]. A study involving the use of SAGE [38] for mRNA measurements and 2DE for protein measurements [39] has also shown a close correlation between mRNA and protein levels for high-copy-number proteins but relatively poor correlation for proteins transcribed at 10 or less copies per cell. Seventy-five percent of genes are transcribed at one or fewer copies per cell [38] with some transcripts per cell as low as 0.001 per cell generation [40]. This has also been confirmed recently using immunodetection of high-affinity epitope-tagged ORFs known as tandem affinity purification (TAP) [41] for which ~80% of the *Saccharomyces cerevisiae* proteome was analyzed [3]. Transcript numbers, protein molecules, and the positive correlation between mRNA and protein abundance for some proteins are shown in Figure 1.3, as is a comparison of the range of proteomic coverage for the techniques of TAP, multidimensional protein identification technology (MudPIT), and 2DE [3].

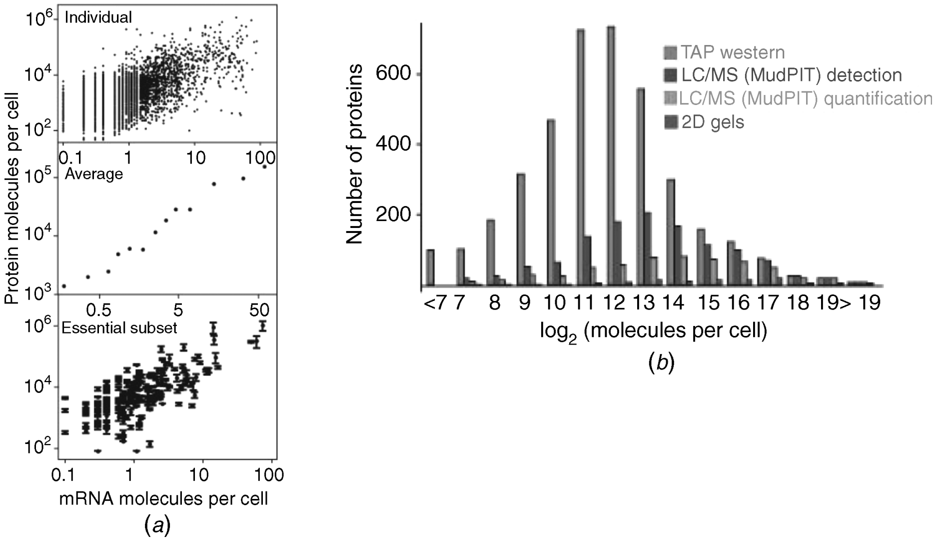

Figure 1.3 Analysis of protein expression in yeast showing (*a*) correlation between protein abundance and transcripts: Top panel shows the relationship between steady-state mRNA and protein levels in yeast as determined by microarray analysis. Middle panel shows ORFs sorted into discrete mRNA levels and means plotted against mean protein abundance. Lower panel shows only a comparison of protein and mRNA levels for some essential soluble proteins. (*b*) Absolute range of expression possible using TAP/Western blot, LC-MS using a multidimensional chromatography approach, and 2DE [3]. (See color insert.)

1.4.2 Metabolome and Proteomics

There is a nonlinear relationship between the presence of a metabolite and a gene, as many genes may be involved in the synthesis or degradation of one metabolite. An example of metabolomics is provided by a well studied microbe, *C. glutamicum,* important in metabolic engineering as its production of glutamine and lysine is used extensively in the food industry [25]. This organism has been comprehensively studied in terms of genome [42, 43], transcriptome [44, 45], proteome [46–48], and metabolome [45]. However, until recently integration of these studies had not occurred. An in-depth study of glycolysis, pentose-phosphate pathway, tricarboxylic acid (TCA) cycle, and lysine synthesis by Krömer et al. [45] of this organism has revealed a dynamic relationship between transcriptome, metabolome, and fluxome (changes in metabolites). It was found that growth continued despite depletion of essential threonine and methionine from growth media (achieved by scavenging intracellular stores) with an increase in total soluble proteins, indicating protein synthesis was still active. Additionally, genes active in translation such as ribosomal proteins and other protein synthesis machinery were significantly expressed. A maximal flux for most enzymes correlated with maximal gene expression, and this could be affected by down regulation at the transcriptional level of enzymes. A conclusion of this work is that it is essential to measure gene, transcript, protein, and metabolite differences to gain insight into biological systems.

1.5 WHAT HAVE MODEL ORGANISMS CONTRIBUTED TO OUR UNDERSTANDING OF BIOLOGICAL SYSTEMS?

Organisms that are representative of more complex systems are amenable to experimental study and are associated with extensive accumulated information from many sources and can be defined as model organisms [49]. In the last 30 years, fundamental research involving model organisms has profoundly advanced our understanding of biological systems. As a result of considerable effort around the globe, organism-specific databases have been created as knowledge of an organism as well as comparative information of relevance to numerous research directions and perspectives becomes available. Many metabolic and developmental pathways are conserved in nature, irrespective of classification level. This feature invites the use of model organisms to minimize the effort required to understand complex biological systems. Table 1.1 summarizes some of the advances that three model organisms have afforded us and how this information has been applied to other organisms.

Life in its most minimal form is still biologically complex. The smallest bacterial genome capable of independent survival is that of *M. genitalium* (517 genes). Interest in this bacterium has focused on derivation of the minimal gene set or what is essentially required to make a cell alive. By selectively switching off each gene in turn, Hutchison et al. [50] were able to determine an essential requirement of 260–350 *M. genitalium* genes, 100 of which had unknown function. This has since been revised to 250 by comparison to other bacterial genomes [51].

On a similar line of thought, the simplest eukaryotic genome, *S. cerevisiae*, plays an important role as a model organism for understanding more complex genomes such as our own. The best example of the value of yeast as a model involves the study of some human-disease-causing genes and their orthologues in yeast, such as hereditary nonpolyposis colon cancer (*MSH2* and *MLH1*), neurofibromatosis type 1 (*IRA2*), ataxia telangiectasia (*TEL1*), and Werner's syndrome (*SGS1*) [32]. In humans, genetic inheritance of these

TABLE 1.1 Summary of Three Model Organisms and How They Have Influenced Proteomics

Organism	Reason for Model Status	Celebrated Findings	Applications	References
M. genitalium	Smallest self-replicating bacterium; determine core genes that drive life	Minimal gene set of 300, 427 abundantly expressed proteins	Derive a synthetic minimal cell for genetic or biochemical manipulation, uses in nanotechnology, therapeutic vector delivery of DNA	[2, 50, 52]
S. cerevisiae	Simplest eukaryotic model for more complex organisms; best-studied eukaryotic system	Description of protein interaction map, first chromosome ever sequenced	Unknown function prediction in other organisms	[15]; [16]; [53]; [54]; [55]
E. coli	Biochemical, molecular, and metabolically best characterized system	2D SDS-PAGE developed using *E. coli*	Standard protein separation technique in proteomics; virtual cell based on accumulated *E. coli* knowledge and applied across species	[56]; [57]

genes results in disease. Initial insight as to the function of these genes was obtained because of their sequence homology to yeast genes and genes of other organisms.

Escherichia coli is the epitome of bacterial model organisms as it is one of the best-studied microbes in terms of genome sequence, metabolic and regulatory networks, proteome, and mutant phenotype studies. Both the pathogenic (O157:H7) and nonpathogenic (K12) strains have been sequenced and compared and databases exist enabling comparison of metabolic and genomic information (Ecocyc: http://ecocyc.org/; Kegg: http://www.genome.ad.jp/kegg/kegg2.html).

There are many more microbial model organisms, and some of their contributions to proteomics and science in general are discussed by contributing authors in this book.

1.6 CONCLUSION

There is an amazing lack of cross communication between the fields of genomics, transcriptomics, and proteomics. This dilemma can be attributed to the dynamic and open-ended nature of the proteome. It is compounded by the evolutionary, internal, and sample complexities of studied organisms and the rapid generation of volumes of data from diverse groups. However, it is apparent that the best understood organisms have had information contributed from all fields. This has most easily occurred for the microbes and will also occur with higher organisms as the inertia to propel research forward and unify knowledge is reached. This will result in a holistic understanding that will be beneficial for all biological systems.

ACKNOWLEDGMENTS

I would like to acknowledge support in part from the Australian Government Systemic Infrastructure Initiative grant.

REFERENCES

1. Wasinger, V. C., Cordwell, S. J., Cerpa-Poljak, A., Yan, J. X., Gooley, A. A., Wilkins, M. R., Duncan, M. W., Harris, R., Williams, K. L., and Humphery-Smith, I., Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium, Electrophoresis* 1995, **16**, 1090–1094.
2. Wasinger, V. C., Pollack, J. D., and Humphery-Smith, I., The proteome of *Mycoplasma genitalium*; CHAPS soluble component, *Eur. J. Biochem.* 2000, **267**, 1571–1582.
3. Ghaemmaghami, S., Huh, W-K., Bower, K., Howson, R., Belle, A., Dephoure, N., O'Shea, E., and Weissman, J., Global analysis of protein expression in yeast, *Nat. Biotechnol.* 2003, **425**, 737–741.
4. Adami, C., Ofria, C., and Collier, T. C., Evolution of biological complexity, *PNAS* 2000, **97**(9), 4463–4468.
5. Southan, C., Has the yo-yo stopped? An assessment of human protein-coding gene number, *Proteomics* 2004, **4**(6), 1712–1726.
6. Jaenisch, R., and Bird, A., Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals, *Nat. Genet.* 2003, **33**, s245–254.

7. Strohman, R., Epigenesis: The missing beat in biotechnology? *Biotechnology (NY)* 1994, **12**(2), 156–164.
8. Kellner, R., Proteomics. Concepts and perspectives, *Frenius J. Anal. Chem.* 2000, **366**, 517–524.
9. Harrison, P. M., Kumar, A., Lang, N., Snyder, M., and Gerstein, M., A question of size: The eukaryotic proteome and the problems in defining it, *Nucleic Acids Res.* 2002, **30**(5), 1083–1090.
10. Graveley, B. R., Alternative splicing: Increasing diversity in the proteomic world, *Trends Genet.* 2001, **17**(2), 100–107.
11. Huang, S., Back to the biology in systems biology: What can we learn from biomolecular networks? *Brief Funct. Genomic Proteomic* 2004, **2**(4), 279–297.
12. Zhu, H., Huang, S., and Dhar, P., The next step in systems biology: Simulating the temporospatial dynamics of molecular network, *Bioessays* 2004, **26**(1), 68–72.
13. Date, S. V., and Marcotte, E. M., Discovery of uncharacterized cellular systems by genome-wide analysis of functional linkages, *Nat. Biotechnol.* 2003, **21**(9), 1055–1062.
14. Pellegrini, M., Marcotte, E. M., Thompson, M. J., Eisenberg, D., and Yeates, T. O., Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles, *Proc. Natl. Acad. Sci. USA* 1999, **96**(8), 4285–4288.
15. Marcotte, E. M., Pellegrini, M., Thompson, M. J., Yeates, T. O., and Eisenberg, D., A combined algorithm for genome-wide prediction of protein function, *Nature* 1999, **402**(6757), 83–86.
16. Enright, A. J., Iliopoulos, I., Kyrpides, N. C., and Ouzounis, C. A., Protein interaction maps for complete genomes based on gene fusion events, *Nature* 1999, **402**(6757), 86–90.
17. Dandekar, T., Snel, B., Huynen, M., and Bork, P., Conservation of gene order: A fingerprint of proteins that physically interact, *Trends Biochem. Sci.* 1998, **23**(9), 324–328.
18. Corthals, G. L., Wasinger, V. C., Hochstrasser D. F., and Sanchez, J-C., The dynamic range of protein expression: A challenge for proteomic research, *Electrophoresis* 2000, **21**, 1104–1115.
19. Oda, Y., Nagasu, T., and Chait, B. T., Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome, *Nat. Biotechnol.* 2001, **19**, 379–382.
20. Krishna, R. G., and Wold, F., Post-translational modification of proteins, *Adv. Enzymol. Relat. Areas Mol. Biol*. 1993, **67**, 265–298.
21. James, P., Mass spectrometry and the proteome, in P. James (Ed.), *Proteome Research: Mass Spectrometry*, Springer, 2001, p. 6.
22. Wasinger, V. C., Bjellqvist, B., and Humphery-Smith, I., Proteomic "contigs" of *Ochrobactrum anthropi*, application of extensive pH gradients, *Electrophoresis* 1997, **18**, 1373–1383.
23. Cordwell, S. J., Basseal, D. J., Bjellqvist, B., Shaw, D. C., and Humphery-Smith, I., Characterisation of basic proteins from *Spiroplasma melliferum* using novel immobilised pH gradients, *Electrophoresis* 1997, **18**(8), 1393–1398.
24. Urquhart, B. L., Cordwell, S. J, and Humphery-Smith, I., Comparison of predicted and observed properties of proteins encoded in the genome of *Mycobacterium tuberculosis* H37Rv, *Biochem. Biophys. Res. Commun.* 1998, **253**(1), 70–79.
25. Schaffer, S., Weil, B., Nguyen, V. D., Dongmann, G., Gunther, K., Nickolaus, M., Hermann, T., and Bott, M., A high-resolution reference map for cytoplasmic and membrane-associated proteins of *Corynebacterium glutamicum*, *Electrophoresis* 2001, **22**(20), 4404–4422.
26. Len, A. C., Cordwell, S. J., Harty, D. W., and Jacques, N. A., Cellular and extracellular proteome analysis of *Streptococcus mutans* grown in a chemostat, *Proteomics* 2003, **3**(5), 627–646.
27. Mdluli, K., Slayden, R. A., Zhu, Y., Ramaswamy, S., Pan, X., Mead, D., Crane, D. D., Musser, J. M., and Barry, C. E. 3rd., Inhibition of a *Mycobacterium tuberculosis* beta-ketoacyl ACP synthase by isoniazid, *Science* 1998, **280**(5369), 1607–1610.

28. Wilson, M., DeRisi, J., Kristensen, H. H., Imboden, P., Rane, S., Brown, P. O., and Schoolnik, G. K., Exploring drug-induced alterations in gene expression in *Mycobacterium tuberculosis* by microarray hybridization, *Proc. Natl. Acad. Sci. USA*, 1999, **96**(22), 12833–12838.
29. Covert, B. A., Spencer, J. S., Orme, I. M., and Belisle, J. T., The application of proteomics in defining the T cell antigens of *Mycobacterium tuberculosis, Proteomics* 2001, **1**(4), 574–586.
30. Gulle, H., Fray, L. M., Gormley, E. P., Murray, A., and Moriarty, K. M., Responses of bovine T cells to fractionated lysate and culture filtrate proteins of *Mycobacterium bovis* BCG, *Vet. Immunol. Immunopathol.* 1995, **48**(1–2), 183–190.
31. Jungblut, P. R., Schaible, U. E., Mollenkopf, H. J., Zimny-Arndt, U., Raupach, B., Mattow, J., Halada, P., Lamer, S., Hagens, K., and Kaufmann, S. H., Comparative proteome analysis of *Mycobacterium tuberculosis* and *Mycobacterium bovis* BCG strains: Towards functional genomics of microbial pathogens, *Mol. Microbiol.* 1999, **33**(6), 1103–1117.
32. Botstein, D., Chervitz S. A., and Cherry J. M., Yeast as a model organism, *Science* 1997, **277**(5330), 1259–1260.
33. Anderson, N. L., Polanski, M., Pieper, R., Gatlin, T., Tirumalai, R. S., Conrads, T. P., Veenstra, T. D., Adkins, J. N., Pounds, J. G., Fagan, R., and Lobley, A., The human plasma proteome: A nonredundant list developed by combination of four separate sources, *Mol. Cell Proteomics* 2004, **3**(4), 311–326.
34. Vitali, B., Wasinger, V., Brigidi, P., and Guilhaus, M., A proteomic view of *Bifidobacterium infantis* generated by multi-dimensional chromatography coupled with tandem mass spectrometry, *Proteomics*, in press.
35. Wilkins, M. R., Sanchez, J. C., Gooley, A. A., Appel, R. D., Humphery-Smith, I., Hochstrasser, D. F., and Williams, K. L., Progress with proteome projects: Why all proteins expressed by a genome should be identified and how to do it, *Biotechnol. Genet. Eng. Rev.* 1996, **13**, 19–50.
36. Hatzimanikatis, V., and Lee, K. H., Dynamical analysis of gene networks requires both mRNA and protein expression information, *Metab. Eng.* 1999, **1**(4), 275–281.
37. Futcher, B., Latter, G. I., Monardo, P., McLaughlin, C. S., and Garrels, J. I., A sampling of the yeast proteome, *Mol. Cell. Biol.* 1999, **19**(11), 7357–7368.
38. Velculescu, V. E., Zhang, L, Vogelstein, B., and Kinzler, K. W., Serial analysis of gene expression, *Science* 1995, **270**(5235), 484–487.
39. Gygi, S. P., Corthals, G. L., Zhang, Y., Rochon, Y., and Aebersold, R., Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology, *Proc. Natl. Acad. Sci. USA* 2000, **97**, 9390–9395.
40. Holland, M. J., Transcript abundance in yeast varies over six orders of magnitude, *J. Biol. Chem.* 2002, **277**(17), 14363–14366.
41. Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B., A generic protein purification method for protein complex characterization and proteome exploration, *Nat. Biotechnol.* 1999, **17**(10), 1030–1032.
42. Ikeda, M., and Nakagawa, S., The *Corynebacterium glutamicum* genome: Features and impacts on biotechnological processes, *Appl. Microbiol. Biotechnol.* 2003, **62**(2–3), 99–109.
43. Kalinowski, J., Bathe, B., Bartels, D., Bischoff, N., Bott, M., Burkovski, A., Dusch, N., Eggeling, L., Eikmanns, B. J., Gaigalat, L., Goesmann, A., Hartmann, M., Huthmacher, K., Kramer, R., Linke, B., McHardy, A. C., Meyer, F., Mockel, B., Pfefferle, W., Puhler, A., Rey, D. A., Ruckert, C., Rupp, O., Sahm, H., Wendisch, V. F., Wiegrabe, I., and Tauch, A., The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins, *J. Biotechnol.* 2003, **104**(1–3), 5–25.
44. Hayashi, M., Mizoguchi, H., Shiraishi, N., Obayashi, M., Nakagawa, S., Imai, J., Watanabe, S., Ota, T., and Ikeda, M., Transcriptome analysis of acetate metabolism in *Corynebacterium*

glutamicum using a newly developed metabolic array, *Biosci. Biotechnol. Biochem.* 2002, **66**(6), 1337–1344.

45. Krmer, J. O., Sorgenfrei, O., Klopprogge, K., Heinzle, E., and Wittmann, C., In-depth profiling of lysine-producing *Corynebacterium glutamicum* by combined analysis of the transcriptome, metabolome, and fluxome, *J. Bacteriol.* 2004, **186**(6), 1769–1784.
46. Hermann, T., Pfefferle, W., Baumann, C., Busker, E., Schaffer, S., Bott, M., Sahm, H., Dusch, N., Kalinowski, J., Puhler, A., Bendt, A. K., Kramer, R., and Burkovski, A., Proteome analysis of *Corynebacterium glutamicum*, *Electrophoresis*, 2001, **22**(9), 1712–1723.
47. Bendt, A. K., Burkovski, A., Schaffer, S., Bott, M., and Farwick, M., and Hermann, T., Towards a phosphoproteome map of *Corynebacterium glutamicum*, *Proteomics* 2003, **3**(8), 1637–1646.
48. Hermann, T., Finkemeier, M., Pfefferle, W., Wersch, G., Kramer, R., and Burkovski, A., Two-dimensional electrophoretic analysis of *Corynebacterium glutamicum* membrane fraction and surface proteins. *Electrophoresis* 2000, **21**(3), 654–659.
49. Barr, M. M., Super models, *Physiol Genomics* 2003, **13**(1), 15–24.
50. Hutchison, C. A., Peterson, S. N., Gill, S. R., Cline, R. T., White, O., Fraser, C. M., Smith, H. O., and Venter, J. C., Global transposon mutagenesis and a minimal *Mycoplasma* genome. *Science* 1999, **286**(5447), 2089–2090.
51. Koonin, E. V., How many genes can make a cell: the minimal-gene-set concept, *Annu. Rev. Genomics Hum. Genet.* 2000, **1**, 99–116.
52. Zimmer, C., Genomics. Tinker, tailor: Can Venter stitch together a genome from scratch? *Science* 2003, **299**(5609), 1006–1007.
53. Schwikowski, B., Uetz P, and Fields, S., A network of protein-protein interactions in yeast, *Nat. Biotechnol.* 2000, **18**(12), 1257–1261.
54. Mewes, H. W., Hani, J., Pfeiffer, F., and Frishman, D., MIPS: A database for protein sequences and complete genomes, *Nucleic Acids Res.* 1998, **26**(1), 33–37.
55. Goffeau, A., Four years of post-genomic life with 6,000 yeast genes, *FEBS Lett.* 2000, **480**(1), 37–41.
56. O'Farrell, P. H., High resolution two-dimensional electrophoresis of proteins, *J. Biol. Chem.* 1975, **250**(10), 4007–4021.
57. Sundararaj, S., Guo, A., Habibi-Nazhad, B., Rouani, M., Stothard, P., Ellison, M., and Wishart, D. S., The CyberCell Database (CCDB): A comprehensive, self-updating, relational database to coordinate and facilitate in silico modeling of *Escherichia coli*, *Nucleic Acids Res*. 2004, **32**, D293–295.
58. Wasinger, V. C., and Corthals, G. L., Proteomic tools for biomedicine, *J. Chromatogr. B.* 2002, **2002**(771), 33–48.