Principles and Practice

Proteome Research

Concepts, Technology and Application

Bearbeitet von
M.R. Wilkins, R.D. Appel, K.L. Williams, D.F. Hochstrasser

Neuausgabe 2007. Buch. XXII, 239 S. Hardcover
ISBN 978 3 540 71240 4
Format (B x L): 15,5 x 23,5 cm
Gewicht: 609 g

1 Ten Years of the Proteome

Marc R. Wilkins and Ron D. Appel

Abstract

The concept of the proteome is now over 10 years old. As with all anniversaries, it is a good time to look back and reflect on what has been achieved in the area that we now call proteomics. What has been done well? What has been done not-so-well? What has been achieved, and what still eludes us? This review will briefly explore some of these questions, with respect to protein separations, mass spectrometry, and proteomic bioinformatics.

1.1 Introduction to the Proteome

The editors of this book have been carrying out research and development in proteomics for more than 20 years. They developed techniques for the analysis of proteins and global protein expression (Williams et al. 1991; Hochstrasser et al. 1988) and software algorithms and tools for the interpretation of the results obtained using such analytical tools (Appel et al. 1988; Wilkins et al. 1995). While the idea of observing the protein expression of genomes in a holistic manner rather than one protein at a time arose with the advent of 2-D gels, the concept of the proteome itself was only introduced by Marc Wilkins in 1994 at a conference in Siena, Italy[1], having coined the term earlier that year in association with his then PhD supervisor Keith Williams. The first papers that began to use the term were published shortly thereafter (Wilkins et al. 1995; Wasinger et al. 1995), and the first book on proteomics was published in 1997 (Wilkins et al. 1997). Ten years has now passed since the publication of that first book, and as with all anniversaries, it is a good time to look back and reflect a little on what has been achieved in the area we now refer to as proteomics. What has been done well? What has been done not-so-well? What has been achieved, and what still eludes us? Here we will suggest answers to these questions. At the same time, we will comment on what we have sought to achieve in this book, and provide a brief précis on its contents.

[1]First Siena conference, 2D electrophoresis: from protein maps to genomes, 5–7 September 1994.

M.R. Wilkins, R.D. Appel, K.L. Williams and D.F. Hochstrasser (Eds.)
Proteome Research: Concepts, Technology and Application

1.1.1 What's in a Word?

The words 'proteome' and 'proteomics' have been widely adopted by the biological community. In the 10 years since their introduction, their use has grown very rapidly (Fig. 1.1). In fact over 4,000 proteomics research and review articles were published in 2005. This has been fuelled by increasing numbers of journals that have arisen to serve the field, including *Proteomics*, *Proteomics-Clinical Application*, *Practical Proteomics*, *Journal of Proteome Research*, *Molecular and Cellular Proteomics*, *Proteome Science*, *Current Proteomics*, *Genomics and Proteomics*, *Briefings in Functional Genomics and Proteomics*, *Genomics Proteomics Bioinformatics* and *Expert Review of Proteomics*. In addition, proteomics research is increasingly published in a variety of other journals, so it has become established as a valuable means to obtain insight into the complexities of biological systems.

If we are simply measuring the progress of a field by its use of language, we might ask if the growth of proteomics is just a reflection of the so-called -omics revolution, or does it show a true growth in the field? The volume of work published in two other newer -omics areas, metabolomics and glycomics, is tiny by comparison, with 433 and 115 manuscripts having been published in 2005, respectively. Proteomics is clearly more widespread and established.

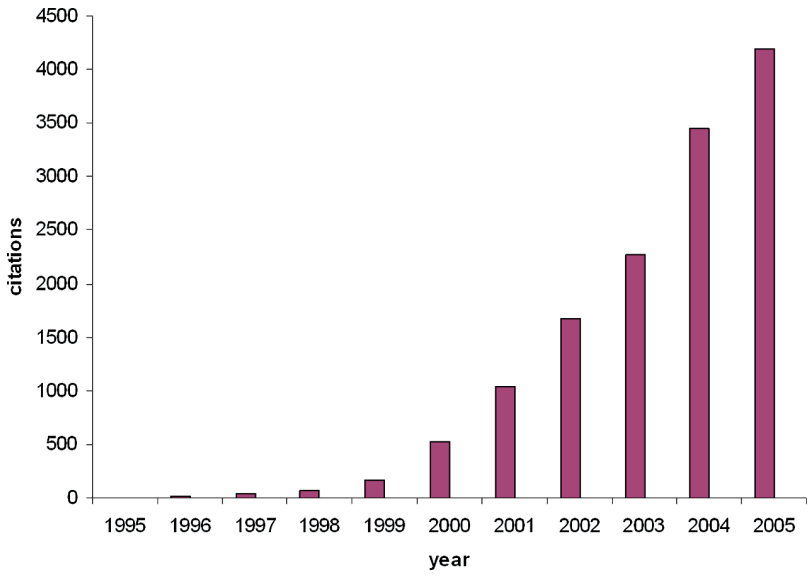

Fig. 1.1 Publications in the field of proteomics and proteome research have grown rapidly in the last 10 years. This was measured by querying the NCBI PubMed database for each year with the words 'proteome' or 'proteomics'. Note, however, that some articles may have been counted twice by this approach

1.1.2 Could Things Have Been Different?

So, would the world have been a different place had the term 'proteome' not been coined? Some commentators have argued that a combination of technical advances in separations technology (gel-based and chromatography-based), in mass spectrometry, and the explosion of information available from genome sequencing efforts have largely driven an increased interest in protein chemistry (Blackstock 2004).

While this is certainly true, it may be argued that the new language has brought renewed focus and legitimacy to protein chemistry that had previously been absent, largely due to the enormous shadow cast by genomics and other nucleic acid based approaches. The new language has also influenced biochemical thinking to move from a one-protein-at-a-time perspective to a more global view. Linguistically, it has been argued that thought cannot exist without language.[2] The proteome and proteomics are examples of this, as are other -omic words which were coined thereafter.[3] The new language and terminology has already helped a gamut of analytical technology to find its place in science and literature. New language in other fields will likewise legitimise emerging technology, focus thinking and also assist the funding of research in these areas.

1.2 Proteomics Is Technology-Driven

If we are to ask what has been done well in proteomics to date, one would have to pay particular attention to the development and dissemination of new technology. In a 10-year period, there have been a number of significant advances that, together, have transformed protein chemistry into the science of proteomics. Importantly, it has been a combination of conceptual breakthroughs and technical advances in separations techniques, mass spectrometry, protein chemistry and bioinformatics which have made this possible. The flood of nucleic acid sequence and genomic information, made available in sequence databases, was another essential co-requisite.

1.2.1 Protein Separations

Initially, proteomics researchers had a goal of visualising all proteins from a proteome on a single, or perhaps one acidic range and one basic range (2-D) polyacrylamide gel. This was happening in the late 1980s, and there was enormous excitement about the possibility of being able to see all

[2]Ferdinand de Saussure, Professor of Linguistics at Geneva University 1901–1913.
[3]See Chitty (2006) for an amusing list of new -omic words.

proteins in a proteome. However, it did not take long to realise that the separation and visualisation of all proteins from a proteome was not a straightforward task. In the mid-1990s, the availability of the first genome sequences and predicted proteomes allowed theoretical 2-D gels to be calculated, showing where each protein spot should be found (Urquhart et al. 1998). This revealed a bimodal distribution of proteins, with the majority of proteins having isoelectric point (pI) 4–6.5 and another group of proteins having pI 8–12. Most proteins had a mass of less than 100 kDa. The comparison of these theoretical maps with experimental 2-D gel separations immediately highlighted shortcomings with 2-D gels in that they were poor in resolving very acidic, very basic or very high mass proteins. A meta-analysis of proteins seen on 2-D gels and those predicted theoretically from genomes of *Escherichia coli*, *Saccharomyces cerevisiae* and *Bacillus subtilis* highlighted two additional issues (Wilkins et al. 1998). The first was that hydrophobic proteins were largely absent from the 2-D gels and that low-abundance proteins present at less than 1,000 copies per cell were likely to be undetectable, owing to limitations on the loading capacity and staining sensitivity of the 2-D gel process.

Since that time, a series of important technical advances have been made to help us see more proteins in the proteome. The latest advances associated with 2-D electrophoresis are discussed in Chap. 2. Broadly speaking, a number of strategies have been adopted. These include the running of narrow pI range gels to 'zoom in' on a particular region of the proteome, the fractionation of samples into either biologically (e.g. organelles) or physicochemically distinct fractions (e.g. membrane proteins) that can then be analysed appropriately, the enrichment or depletion of proteins of interest from a sample, along with new solubilisation and gel running techniques to assist in the analysis of the more difficult proteins. Importantly, fractionation has provided an avenue to load more of the relevant portion of samples of interest onto 2-D gels, thus assisting in the detection of lower-abundance proteins.

To completely bypass many of the challenges of working with complex mixtures of proteins, a conceptually different strategy emerged for protein analysis in proteomics. Called 'shotgun proteomics', probably inspired by the shotgun DNA sequencing approaches that were developed by Venter et al. (1998), it involves taking complex mixtures of proteins or indeed a whole proteome, and digesting all proteins to peptides with endoproteinases of known specificity. The resulting mixtures of peptides, which are physicochemically more homogenous than their parent proteins although greater in number, are then analysed using 2-D liquid chromatography and tandem mass spectrometry. Peptide fragment data are matched against sequence databases (Wolters et al. 2001) to determine the proteins present in a sample. Whilst this approach has limitations, notably the loss of protein isoforms (see Chap. 5), it provides an alternative to gel-based analyses for the separation and identification of large numbers of proteins from a proteome.

1.2.2 Mass Spectrometry

The last 20 years has brought astonishing advances in mass spectrometry technology. These advances have helped establish the science of proteomics. Mass spectrometers, whilst remaining expensive, now have remarkable mass accuracy and resolution, can analyse femtomolar quantities of peptides and proteins, and are increasingly automated. Two means of ionisation of proteins and peptides are in widespread use, electrospray ionisation and matrix-assisted laser desorption/ionisation, and these are teamed with a variety of mass analysers and detectors (see Chap. 3).

Mass spectrometers have all but superseded Edman degradation as the method of choice for protein identification. Two techniques, namely peptide mass fingerprinting and peptide fragmentation, can be used. Peptide mass fingerprinting has been used in a number of massive projects, for example more than 20,000 proteins were analysed as part of a large-scale analysis of yeast protein complexes (Gavin et al. 2002). However, peptide mass fingerprinting is losing favour to higher-confidence peptide fragmentation approaches that are able to fragment multiple peptides from the same protein. Nevertheless, it should be noted that mass spectrometers typically do not sequence peptides or proteins *per se*. They instead allow us to infer sequences by matching peptide fragmentation data against sequence databases. Routine *de novo* sequencing remains complex and is thus a work in progress (see Chap. 3).

In addition to protein identification, a myriad of new mass spectrometry approaches have been developed for the quantitative analysis of two or more samples. Such comparisons are of great scientific interest for the detection of biomarkers and the understanding of the multiplicity of changes that can occur when a proteome is perturbed by intrinsic or extrinsic forces. Previously, the comparison of protein expression in two or more samples was done by 2-D gel electrophoresis and computer image analysis (see Sect. 4.2). This approach has been successfully used in a large number of studies and remains widespread. The newer mass spectrometry based approaches are a significant advance and essentially use different stable isotopes to label proteins from two or more samples (Gygi et al. 1999). The samples are then mixed together and co-analysed. The high mass accuracy of the mass spectrometers allows the isotopic variants to be separated and relative quantitation to be undertaken. This concept has now been developed in a number of different ways (see Sect. 4.3) and whilst not perfect is providing a new means to undertake comparative analysis of two or more complex samples.

A final area in which mass spectrometry is now playing a major role is in the characterisation of proteins. Post-translational modifications of proteins are of increasing interest as they are key to the control and modulation of many processes inside the cell. Our recent appreciation of their roles in protein–protein interaction networks, whereby interactions between many proteins require the presence of certain post-translational modifications (Pawson and Nash 2003), is providing even greater impetus for their study.