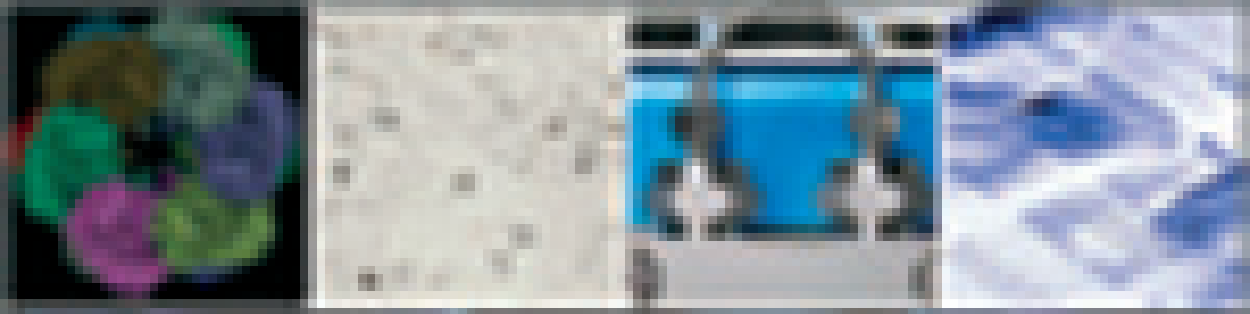
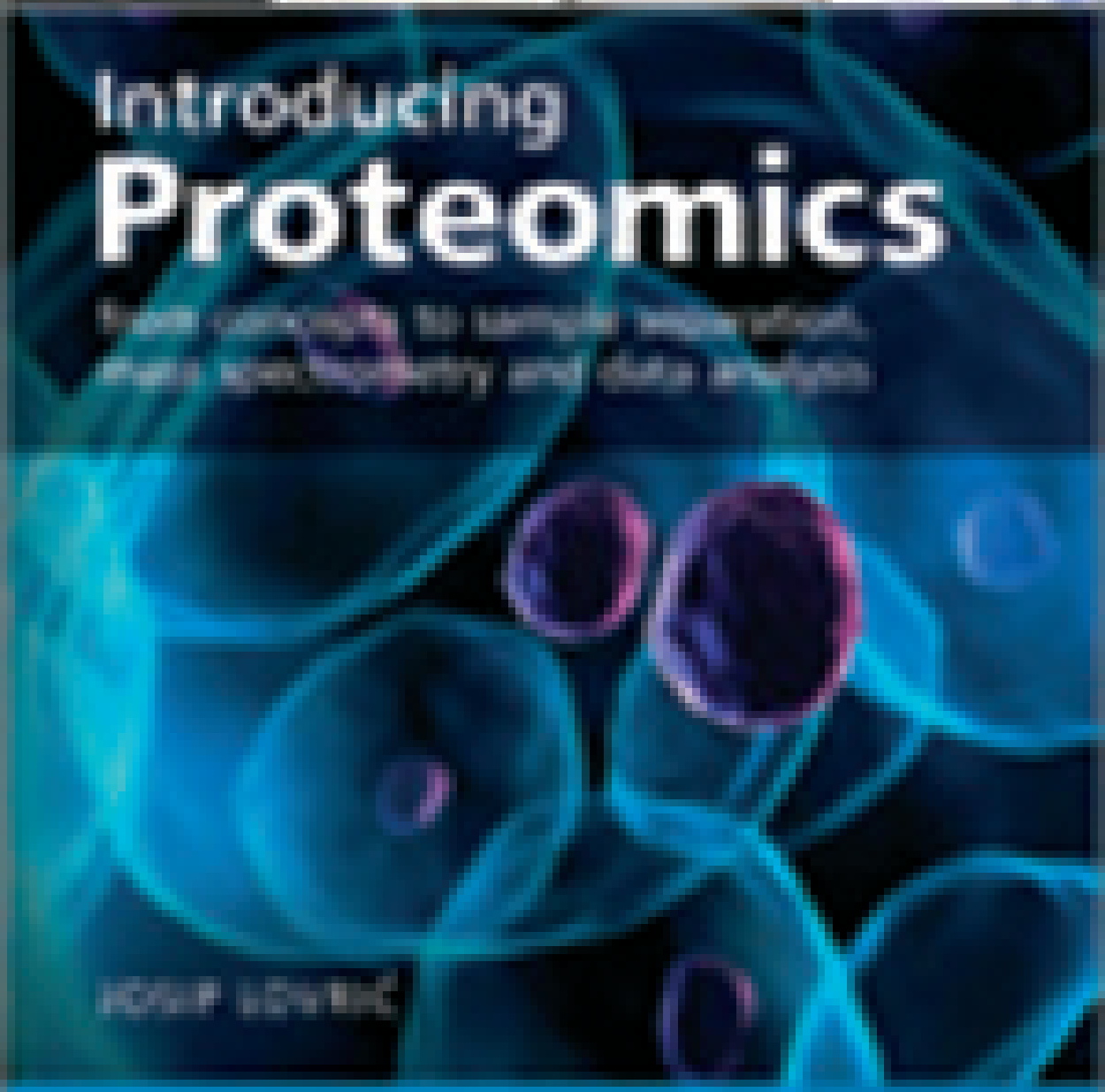


Introducing Proteomics

From concepts to sample separation, mass spectrometry and data analysis



Rob Beynon

Portland Press

Introducing Proteomics

Introducing Proteomics

From concepts to sample separation, mass spectrometry and data analysis

Josip Lovrić
Faculty of Life Sciences, University of Manchester

WILEY-BLACKWELL
A John Wiley & Sons, Ltd., Publication

Wiley-Blackwell is an imprint of John Wiley & Sons, formed by the merger of Wiley's global Scientific, Technical and Medical business with Blackwell Publishing.

Registered office: John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial Offices:
9600 Garsington Road, Oxford, OX4 2DQ, UK
111 River Street, Hoboken, NJ 07030-5774, USA
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Library of Congress Cataloguing-in-Publication Data
Lovrić, Josip.
Introducing proteomics : from concepts to sample separation, mass spectrometry, and data analysis / Josip Lovrić.
p. ; cm.
Includes bibliographical references and index.
ISBN 978-0-470-03523-8 (cloth) – ISBN 978-0-470-03524-5 (pbk.)
1. Proteomics. I. Title.
[DNLM: 1. Proteomics. QU 58.5]
QP551.L875 2011
572′.6 – dc22

2010029400

A catalogue record for this book is available from the British Library.

This book is published in the following electronic formats: ePDF ISBN 978-0-470-67021-7.

Typeset in 9/12pt Times-Roman by Laserwords Private Limited, Chennai, India
Printed in Singapore by Markono Print Media Pte Ltd.

First Impression 2011

To my family, I hope they'll have me back . . . . . .

A narućito naknađujem ovo knjigu za moje roditelje.

Contents

Preface

The term proteomics was coined in the mid-1990s by the Australian (then post-doctoral) researcher Marc Wilkins. The term arose in response to the spirit of the day; researchers working in genetics developed genome-wide approaches and were very successful at the time. Researchers working on proteins rather than genes also felt that the time was right for a more holistic approach – rather than working on a single protein at a time, many (if not all) proteins in a single biological system should be analysed in one experiment. While surely the will was there and some good foundational work was done at the time, it still took about another five years before technologies were developed far enough, so that proteomics became a concept that could deliver some (but still not all) of the answers researchers hoped to be able to get by using it.

Historically proteomics was driven mainly by researches coming from the field of 2D gel electrophoresis. These 'bluefingers' joined forces with experts in mass spectrometry and bioinformatics. It was the combination of these fields together with the genomic revolution that created the first proteomic approaches. These were inevitably studies using 2D gel electrophoresis in combination with mass spectrometry of 'isolated' spots, often using MALDI-ToF mass spectrometry. In the beginning the development of the ionization technologies of MALDI, ESI and nano ESI were critical steps to allow the mass spectrometric analysis of biological material with reasonable sensitivity.

Together with advances in all fields concerned, it was major developments in gel free, hyphenated peptide separation technologies that allowed proteomics to prosper in more recent times. Recent developments in gel based proteomics were confined mainly to more convenient sample handling and more pre-fabricated devices and most important computer based image analysis and new protein dyes, allowing for less variable results in a shorter time with less manual input. 2D HPLC in combination with tandem mass spectrometry is a hallmark of the development of hyphenated technologies. Modern proteomics is driven by the development of ever improving software to deal with the huge amount of data generated, allowing better and more efficient data mining; new mass spectrometers allowing new imaging approaches or qualitatively better approaches through improvements in versatility, accuracy, resolution or sensitivity. Also developments in labelling reagents and affinity matrices allow more intelligent approaches, more tailored to specific questions, such as quantitative analyses and analyses of phosphorylations and other posttranslational modifications. Nano-separation methods become more routine and combination of multi-dimensional separation approaches become feasible, allowing 'deeper' views of the proteome. And if all these developments were not enough, there is a plethora of more specialized developments, like the molecular scanner (Binz *et al*., 2004), MALDI imaging mass spectrometry for tissues, organs and whole organisms such as the mouse or rat (Caldwell and Caprioli, 2005) or Laser Capture Microdissection (Jain, 2002) which enables proteomics analysis from just a dozen of cells (Nettikadan *et al*., 2006). In a book like this it is impossible to do justice to all these developments, and they will be mentioned as we go along, especially in Chapter 5 on strategies in proteomics. Sadly, some fields such as 3D structural analyses have had to be omitted.

Next to complete 'work floors', the mass spectrometers and separation devices (e.g. nano HPLC, free flow electrophoresis equipment) that come with the territory represent the biggest capital investment for laboratories getting involved in proteomics, ranging from some US$ 160 000 to more than a million dollars per item. In the early

days of proteomics, many developments were driven by scientists rather than industries. Since 2000, proteomics has become big business, with the potential for companies to sell hundreds of mass spectrometers instead of a dozen a year to the scientific community.

Away from 'classical' approaches there have been huge developments in very diverse fields such as protein fluorescent staining, chemical peptide modification, ultra-accurate mass spectrometers, microscope assisted sample collection, improved sample treatment, isobaric peptide tagging and of course bioinformatics, to name just a few, that have opened up a whole range of new possibilities to tackle biological problems by proteomic approaches. It is this diverse group of fields that contribute towards making proteomics such a vibrant and interesting field, on the one hand, but also a field that may seem difficult to get started in, on the other hand.

This is where we aim to place this book: to give an introduction to the complete field of proteomics without delving too deeply into every single area within it, because for most of these areas there is excellent specialist literature available.

In this respect the book aims primarily to give a basic understanding of the most important technologies. At the same time it intends to allow the reader to develop an understanding of the possibilities, but also the limitations, of each of the technologies or their combinations. All this is presented with the aim of helping the reader to develop proteomic approaches that are suited to the needs of their specific research challenge.

WHO WILL BENEFIT FROM THE BOOK

This book is aimed at diverse groups of potential users. In the academic world it is written easy enough to be useful and aimed at undergraduate students to give an introduction to the field of proteomics; so many biochemical/physical principals are explained at that level.

On the other hand, this book will also be useful for postgraduate students and more senior researchers in academia and industry. While it brings an overview and an explanation of principles to postgraduate students who may be about to start to work on a proteomic project, it will also explain the possibilities and limitations of a potential proteomics approach for a principal investigator and give them an idea of the sort of financial and intellectual commitments necessary.

It will be a useful tool for experienced researchers in the field of proteomics to 'catch up' on areas that were outside their focus for a while or have developed only recently. It may also help scientists to understand the needs of a certain approach and help them with their planning; be it for starting collaborating with someone in the field of proteomics, or to help such a collaboration to be successful or for writing a new grant in this field.

While this book does not contain recipes or manuals for instruments, it will be of great benefit in helping people to get trained practically in the field, since it explains all the major principles and puts them in a wider perspective.

I hope it will also help researchers from (apparently) distant areas of research to develop new approaches and identify fields in which further research into technologies might be necessary and possible to help proteomics to become and remain one of the sharpest tools in the box of biological and medico-pharmaceutical research.

REFERENCES

Binz, P.A., Mueller, M., Hoogland, C. *et al*. (2004) The molecular scanner: concept and developments. *Curr Opin Biotech*, **15**, 17–23.

Caldwell, R.L. and Caprioli, M.R. (2005) Tissue profiling by mass spectrometry. *Mol Cell Proteomics*, **4** (4), 394–401.

Jain, K.K. (2002) Application of laser capture microdissection to proteomics. *Methods Enzymol*, **356**, 157–167.

Nettikadan, S., Radke, K., Johnson, J. *et al*. (2006) Detection and quantification of protein biomarkers from fewer than 10 cells. *Mol Cell Proteomics*, **5** (5), 895–901.

Acknowledgements

My sincere thanks go to all those individuals who helped in the making of this book, be it by advice, by letting me have their data or by helping in the production process of this book. Special thanks are due to Dr Alistair McConnell, Dr Chris Storey, Dr David Knight, Fiona Woods, Gill Whitley, Dr Guido Sauer, Haseen Khan and her amazing project team, Izzy Canning, Nicky McGirr, Dr Paul Sims and Dr Songbi Chen.

1

Introduction

1.1 WHAT ARE THE TASKS IN PROTEOMICS?

1.1.1 The proteome

In genomics, one of the main aims is to establish the composition of the genome (i.e. the location and sequence of all genes in a species), including information about commonly seen polymorphisms and mutations. Often this information is compared between different species and local populations. In functional genomics, scientists mainly aim to analyze the expression of genes, and proteomic is even regarded by some as part of functional genomics. In proteomics we aim to analyze the whole proteome in a single experiment or in a set of experiments. We will shortly look at what is meant by the word analysis. Performing any kind of proteomic analysis is quite an ambitious task, since in its most comprehensive definition the proteome consists of all proteins expressed by a certain species. The number of these proteins is related to the number of genes in an organism, but this relation is not direct and there is much more to the proteome than that. This comprehensive definition of the proteome would also account for the fact that not a single individual of a species will express all possible proteins of that species, since the proteins might exist in many different isoforms, with variations and mutations, differentiating individuals. An intriguing example are antibodies, more specifically their antigen binding regions, which exist in millions of different sequences, each created during the lifetime of individuals, without their sequence being predictable by a gene. Antibodies are also a good example of the substantial part played by external influences, which define the proteome; for example, the antibody-mixture present in our bodies is strictly dependent on which antigens we have encountered during our lives. But of course a whole host of more obvious external factors influence our proteome, but not the genome (Figure 1.1).

Furthermore, the proteome also contains all possible proteins expressed at all developmental stages of a given species; obvious examples are different proteins in the life cycle of a malaria parasite, or the succession of oxygen binding species during human development, from fetal haemoglobin to adult haemoglobin (Figure 1.2).

On top of all these considerations, there are possible modifications to the expression of a protein that are not encoded by the sequence of its gene alone; for example, proteins are translated from messenger RNAs, and these mRNAs can be spliced to form different final mRNAs. Splicing is widespread and regulated during the development of every single individual, for example during the maturation of specific cell types. Changes in differential splicing can cause and affect various diseases, such as cancer or Alzheimer's (Figure 1.3).

As if all this was not enough variability within the proteome, most proteins show some form of posttranslational modification (PTM). These modifications can be signs of ageing of the protein (e.g. deamidation or oxidation of old cellular proteins; Hipkiss, 2006) or they can be added in an enzymatically regulated fashion after the proteins are translated, and are fundamental to its function. For example, many secreted proteins in multicellular organisms are glycosylated. In the case of human hormones such as erythropoietin this allows them to be functional for longer periods of time (Sinclair and Elliott, 2004). In other cases proteins are modified only temporarily and reversibly, for example by phosphorylation or methylation. This constitutes a very important mechanism of functional regulation, for example during signal transduction, as we will see in more detail later. In summary, there are a host of

Introducing Proteomics: from Concepts to Sample Separation, Mass Spectrometry and Data Analysis, Josip Lovrić