PROTEOMICS

from protein sequence to function

Edited by
S R PENNINGTON • M J DUNN

Proteomics

From protein sequence to function

S.R. Pennington
Department of Human Anatomy and Cell Biology, University of Liverpool, UK

and

M.J. Dunn
Heart Science Centre, Harefield Hospital, Harefield, UK

First published 2001
Reprinted 2001

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

ISBN 1 85996 296 3

BIOS Scientific Publishers Ltd
9 Newtec Place, Magdalen Road, Oxford OX4 1RE, UK
Tel. +44 (0)1865 726286. Fax +44 (0)1865 246823
World Wide Web home page: http://www.bios.co.uk/

Published in the United States of America, its dependent territories and Canada by Springer-Verlag New York Inc., 175 Fifth Avenue, New York, NY 10010–7858, in association with BIOS Scientific Publishers Ltd.

Production Editor: Andrea Bosher
Typeset and illustrated by Paston PrePress Ltd, Beccles, UK
Printed by The Cromwell Press, Trowbridge, UK

Contributors

Aebersold, Ruedi. Department of Molecular Biotechnology, University of Washington, Box 357730, Seattle, WA 98195, USA

Blakemore, Steve. Genomics Unit, Glaxowellcome Medicine Research Centre, Gunnels Wood Road, Stevenage, SG1 2NY, UK

Cahill, Dolores J. Max-Planck-Institute of Molecular Genetics, Ihnestrasse 73, D-14195 Berlin-Dahlem, Germany

Cahill, Michael A. Department of Molecular Biology, Institute for Cell Biology, University of Tuebingen, Auf der Morgenstelle 15, Tuebingen, D-72076, Germany

Davison, Matthew. 13S28 Mereside, Alderley Park, Macclesfield, SK10 4TG, UK

Dunn, Michael J. National Heart and Lung Institute, Imperial College of Science, Technology and Medicine, Heart Science Centre, Harefield Hospital, Harefield UB9 6JH, UK

Eickhoff, Holger. Max-Planck-Institute of Molecular Genetics, Ihnestrasse 73, Berlin, D-14195, Germany

Evers, Stefan. F. Hoffmann-La Roche Ltd, PRPI-D, Building 69/16, Basel, CH-4070, Switzerland

Goodlett, David R. Department of Molecular Biotechnology, University of Washington, Box 357730, Seattle, WA 98195, USA

Görg, Angelika. Technische Universität München, Institut für Lebensmitteltechnologie und Analytische Chemie, FG Protemik, D-85350 Freising, Weihenstephan, Germany

Gray, Christopher. F. Hoffmann-La Roche Ltd, PRPI-D, Building 69/16, Basel, CH-4070, Switzerland

Humphrey-Smith, Ian. Department of Microbiology, Centre for Proteome Research and Gene-Product Mapping, Suite G12, National Innovation Centre, Australian Technology Park, Eveleigh, NSW, Australia

James, Peter. Wallenberg Laboratory, Lund University, Lund, Sweden

Jenkins, Rosalind. Department of Human Anatomy and Cell Biology, New Medical School, Ashton Street, Liverpool L69 3GE, UK

Johnson, Kevin. The Science Park, Melbourne, Cambridge SG8 6JJ, UK

Keatinge, Lucy. Department of Biological Sciences, Division of Environmental and Life Sciences, Maquarie University, NSW 2109, Australia

Klose, Joachim. Humboldt-Universität, Charité, Virchow-Klinikum, Institüt für Humangenetik, Augustenburger Platz, 13353 Berlin, Germany

Lehrach, Hans. Max-Planck-Institute of Molecular Genetics, Ihnestrasse 73, Berlin, D-14195, Germany

Lopez, Mary. 22 Alpha Road, Chelmsford, MA 01824-4, USA
Nordheim, Alfred. Department of Molecular Biology, Institute for Cell Biology, University of Tuebingen, Auf der Morgenstelle 15, Tuebingen, D-72076, Germany
Nordhoff, Eckhard. Max-Planck-Institute of Molecular Genetics, Ihnestrasse 73, Berlin, D-14195, Germany
O'Brien, John. Max-Planck-Institute of Molecular Genetics, Ihnestrasse 73, Berlin, D-14195, Germany
Oswald, Helmut. German Heart Institute, Augustenburger Platz, Berlin, D-13353, Germany
Packer, Nicolle H. Proteome Systems Ltd, 1/35–41 Waterloo Road, North Ryde, Sydney, NSW 2113, Australia.
Patterson, Scott D. Amgen Inc., Amgen Centre, MS 14-2-E, One Amgen Centre Drive, Thousand Oaks, California, CA 91320-1789, USA
Patton, Wayne F. Molecular Probes Inc., 4849 Pitchford Avenue, Eugene, OR 97402, USA
Pennington, Stephen R. Department of Human Anatomy & Cell Biology, New Medical School, University of Liverpool, Liverpool L69 3GE, UK
Pleissner, Klaus-Peter. Institute for Plant Genetics and Crop Plant Research, Corrensstr 3, Gatersleben, D-06466, Germany
Plomion, Christophe. Laboratoire de Génétique et Amélioration des Arbres Forestiers, INRA, France
Pritchard, Kevin. Cambridge Antibody Technology Ltd, The Science Park, Melbourne, Cambridge SG8 6JJ, UK
Quadroni, Manfredo. Institute of Biochemistry, University of Lausanne, Epalinges, Switzerland
Thiellement, Herve. Department de Biologie Vegetale, Universite de Geneve, 3 Place de l'Universite, Geneve 4, CH-1211, Switzerland
Trower, Michael K. Genomics Unit, Medicines Research Centre, Gunnels Wood Road, Stevenage SG1 2NY, UK
Valge-Archer, Viia. Cambridge Antibody Technology Ltd, The Science Park, Melbourne, Cambridge SG8 6JJ, UK
Wallace, Don M. Medicines Research Centre, Glaxo Wellcome Research and Development, Stevenage SG1 2NY, UK
Wegner, Susan. German Heart Institute, Augustenberger Platz 1, Berlin, D-13353, Germany
Wilkins, Marc. Proteome Systems Ltd, Locked Bag 2073, North Ryde, Sydney, NSW 1670, Australia
Zivy, Michel. Station de Génétique Végétale, Gif-sur-Yvette, France

Abbreviations

ABA	abscissic acid
AFLP	amplified fragment length polymorphism
ANS	1-anilino-8-napthalene sulphonate
APAF	Australian Proteome Analysis Facility
ARRM	Advanced Rapid Robotics Manufacturing
ASRI	ABA/water stress/ripening related protein
bis-ANS	bis(8-toluidino-1-naphthalene sulfonate)
BLAST	basic local alignment search tool
BSA	bovine serum albumin
CAD	collision-activated dissociation
Cam	Chloramphenical
CCD camera	charge-coupled device camera
cDNA	complementary DNA
CE	capillary electrophoresis
CG	candidate gene
CHAPS	3[(cholamidopropyl)dimethylammonio]-1-propane sulphonate
CID	collision-induced dissociation
CP	candidate protein
CS	Chinese spring
CSF	cerebrospinal fluid
CyA	cyclosporine A
CZE	capillary zone electrophoresis
dbEST	expressed sequence tag database
DD-PCR	differential display PCR
1-DE	one-dimensional polyacrylamide gel electrophoresis
2-DE	two-dimensional polyacrylamide gel electrophoresis
DGE	differential gene expression
DHFR	dihydrofolate reductase
2D-PP	two-dimensional phosphopeptide
DMAA	*N*,*N*-dimethyl acrylamide
DNA	deoxyribonucleic acid
DSN	database spot number
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme-linked immunosorbant assay
ESI	electrspray ionization
EST	expressed sequence tags

FACS	fluorescence-activated cell sorter
FITC	fluorescein isothiocyante
FPR	formyl peptide receptor
FSH	follicle stimulating hormone
FT-ICR-MS	fourier transform ion cyclotron resonance mass spectrometry
Fus	fusidic acid
FWHM	full-width half maximum
GAPDH	glyceraldehyde-3-phosphate dehytrogenase
GPI	glycosyl phosphatidylinositol
hCG	human chorionic gonadotrophin
HCl	hydrochloric acid
HIS-tagged	histidine-tagged
HIV	human immunodeficiency virus
HPLC	high-performance liquid chromatography
HRP	horseradish peroxidase
HSP	heat shock protein
HUVEC	human umbilical vein endotheilial cells
ICAT	isotope-coded affinity tagging
ICC	immunocytochemistry
ICE	interleukin converting enzyme
ICL	instrument control language
IEF	isoelectric focusing
IgG	immunoglobulin G
IMAC	immobilized metal affinity chromatography
IMAGE	integrated molecular analysis of genomes and their expression
IMS	ion mobility spectrometry
IP	isoelectric point
IPG	immobilized pH gradient
IPG-DALT	2-DE with immobilized pH gradients in the first dimension
IPTG	isopropyl-β-D-thiogalactopyranoside
IR	infrared
IRMPD	infrared multiphoton dissociation
IT	ion trap
Kan	kanamycin
KCl	potassium chloride
LC	liquid chromatography
LCM	laser capture microdissection
LC-MS/MS	liquid chromatography tandem mass spectrometry
LEA	late embryogenesis-abundant
LH	luteinising hormone
LIMS	laboratory information management
LMW	low molecular weight
LOG	Laplacian of Gaussian
m/z	mass-to-charge ratio
MALDI-MS	matrix-assisted laser desorption-ionization mass spectrometry

MALDI-TOF	matrix-assisted laser desorption-ionization time-of-flight
MAP	mitogen-activated protein
MDPF	2-methoxy-2,4-diphenyl-3(2H)furanone
MeCN	acetonitrile
MIP	molecularly imprinted polymer
MPI	minimal protein identifier
Mr	relative molecular mass
mRNA	messenger RNA
MS	mass spectrometry
MS/MS	tandem mass spectrometry
NCBI	National Centre for Biotechnology Information
NEPHGE	non-equilibrium pH gradient electrophoresis
NP-40	Nonidet P40
OD	optical density
PBS	phosphate buffered saline
P/A	presence/absence
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDA	piperazine diacrylamide
ppm	parts per million
PQL	protein quantity loci
PS	position shift
PSD	postsource decay
PVDF	polyvinylidene difluoride
QCM	quartx crystal microbalance
QTL	quantitative trait loci
RAPD	random amplified polymorphic DNA
RCE	relative collision energy
RF	radio frequency
RFLP	restriction fragment length polymorphism
rHu G-CSF	human granulocyte colony-stimulating factors
RNA	ribonucleic acid
RP-HPLC	reversed-phase high performance liquid chromatography
RT-PCR	reverse transcriptase polymerase chain reaction
SAGE	serial analysis of gene expression
SAM	self-assembled monolayer
SB	sulphobetaine
SCA	synthetic carrier ampholytes
scFv	single chain variable fragment
SCX	cation exchange
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SELDI	surface-enhanced laser desorption/ionization
SIM	selected ion monitoring
Smz	sulfamethoxazole

SP	candidate protein
SPR	surface plasmon resonance
SSP	standard spot number
Str	streptomycin
TBP	tributyl phosphine
TCA	trichloroacteic acid
TEA	triethylamine
TEMED	*N*, *N*, *N*′, *N*′-tetramethylethylenediamene
Tet	tetracycline
TFA	trifluroracetic acid
TGF	transforming growth factor
THF	tetrahydrofuran
TIC	total ion count
TIGR	The Institute for Genomics Research
TLC	thin-layer chromatography
TM	transmembrane
Tmp	trimethoprim
TNF	tumor necrosis factor
TQ	triple quadrupole
tRNA	transfer RNA
UV	ultraviolet
VCAM	vascular cell adhesion molecule
WST	watershed transformation
www	World Wide Web
ZP	zona pellucida

Preface

Recent developments in analytical methods for protein characterization, and the growing rate at which whole genome sequencing projects are being completed, have combined to support the emergence and development of the new field of 'proteomics'. Proteomics, the study of protein expression and function on a genome scale, is the amalgamation of very many different experimental approaches ranging from the analysis of gene function by mRNA expression profiling with cDNA arrays, analysis of protein:protein interactions by genome-wide two hybrid screening, to more direct analysis of protein expression, sequence and structure.

Proteomics: from protein sequence to function covers many aspects of this emerging field. In particular, it provides a detailed overview of the current methods used to undertake two-dimensional polyacrylamide gel electrophoresis based measurement of protein expression and mass spectrometry based protein identification, and characterization and their associated informatics. The contributors include many leaders in the field including those who have played pivotal roles in the development of two-dimensional polyacrylamide gel electrophoresis, detection of proteins, gel image analysis, analysis of proteins by mass spectrometry, the generation and application of phage-displayed antibodies and the integration of methods to support proteomics. This book is intended as a guide to these and other methods for all who have an interest in proteomics: newcomers and experienced practitioners alike.

The book begins with two important chapters: one covers the integration of genomics and proteomics and the second describes current approaches to the measurement of mRNA expression. Increasingly, data from proteome analyses are being integrated with those from other DNA based approaches and so these chapters form a vital platform for the subsequent chapters that describe various aspects of the proteomics workflow. Chapters on two-dimensional polyacrylamide gel electrophoresis, protein detection, mass spectrometry based protein characterization including a chapter on recent developments in mass spectrometry that support quantitative analysis of protein expression and image analysis follow. Much of the proteomics workflow is at present laborious, time consuming and generates data on a scale that requires the application of software for data management. These are the subjects of two chapters on approaches to automation of the proteomics workflow. Although no one can deny that two-dimensional gel electrophoresis currently provides the most powerful platform for the analysis of protein expression in both simple and complex organisms, it does have significant limitations and so a chapter on potential alternatives to the technique follows. Further chapters describe the application of proteome analysis to drug development; the use of phage antibodies as tools for proteomics; glycobiology and proteomics; the establishment

of proteomics facilities in academic laboratories and the use of proteomics in plant genetics and breeding.

Clearly, there are many other aspects of proteomics that could have been included; we hope that the selection we have chosen (which inevitably reflects our own interests) will provide a strong foundation for those wishing to learn more about proteomics. We are very aware that we have not included the exciting developments, both practical and bioinformatic, in the elucidation of protein tertiary structure; these would arguably require a book in their own right if they were to be covered in sufficient detail.

There have been many people who have helped to bring this book to completion - a feat that at times it seemed might not be achieved – and our sincere thanks go to all of them. We are very grateful indeed to the contributors, their interest and enthusiasm in the project has been much appreciated. We are especially grateful to Scott Patterson, Marc Wilkins and Peter James who provided invaluable advice and help. We also thank all the team at BIOS and those who supported our 'logistics' activities including Jenni Brown, Lisa Crimmins and Jane Hamlett.

The role of proteomics in meeting the post-genome challenge

The rapidly emerging field of proteomics has now established itself as a credible approach for furthering our understanding of the biology of whole organisms – from simple unicellular organisms to those as complex as man. The readily available experimental tools for measurement of protein expression by two-dimensional gel electrophoresis (2-DE), and for protein identification and characterization by mass spectrometry-based methods, have already made a significant impact on proteomics. The growing interest in the field looks set to accelerate the development and implementation of improved strategies for the analysis of protein expression and function on a genome-wide scale.

The advent of proteomics has, in substantial part, been dependent on the success of whole genome sequencing projects, not least because the completion of these projects has resulted in the more widespread appreciation that in themselves they reveal little about the biology of the organism. Instead, they provide an essential platform for a wide range of complementary experimental approaches that will support the characterization of the genes encoded within the genomes, and ultimately the understanding of how the products of these genes act together to regulate the activity of the organism. Thus, it seems evident that genomics and proteomics will best serve the community of biologists if these two fields synergistically develop their co-dependence.

The success of whole genome sequencing projects has been both remarkable and exciting. The first complete genome for a free-living organism was published in 1995 (Fleischmann *et al.*, 1995) and as of July 2000, there were 40 complete genome sequences in the public domain, with a further 127 prokaryotic and 95 eukaryotic genomes under analysis (wit.integratedgenomics.com/GOLD/). So far, chromosomes 5, 16, 19, 21 and 22 of the human genome have been completed and the release of the complete, corrected and mapped human genome has been predicted for 2003. This is likely to have its initial impact on medical diagnostics and indeed a Consortium to map single nucleotide polymorphisms (SNPs) has been initiated (see *Table 1* for websites). The list of complete genome sequences continues to grow at an accelerating pace and is being matched by the unprecedented public access to databases, search tools and associated electronic information sources that are available (see *Table 1*). The sequence information has already been exploited in ways that would have been almost inconceivable to most just a decade ago. For example, comparative genomics compares all the gene sequences of a particular organism with all other genomes in order to identify differences that may account for defined and important properties, such as pathogenicity. The new field of structural genomics aims to expedite the determination of protein structure via

Table 1. Web addresses for a selection of sequence databases and analysis tools

Website	Organization	Information available
ensembl.ebi.ac.uk	European Molecular Biology Laboratory, Heidelberg, Germany/ Sanger Centre, Cambridge, UK	Annotation of human, mouse and worm genomes
genome.wustl.edu/gsc/	Genome Sequencing Center, Washington University School of Medicine, St. Louis, USA	Human and model organism sequencing projects, EST projects, protocols and technical help
www.hgmp.mrc.ac.uk/	Human Genome Mapping Project Resource Centre, Wellcome Trust Genome Campus, Cambridge, UK	Sequence databases and search engines, phylogenetic linkage analysis, links to useful websites
www.hgsc.bcm.tmc.edu/	Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, USA	Human, mouse and Drosophila sequencing projects, human transcript database
wit.integratedgenomics.com/ GOLD	Integrated Genomics Inc., Chicago, Illinois, USA	Monitors complete and ongoing genome sequencing projects and links to relevant sites and publications
www.jgi.doe.gov/	Joint Genome Institute, Walnut Creek, California, USA	Human and microbial sequencing and mapping, functional genomics programme
star.scl.genome.ad.jp/kegg	Kyoto Encyclopaedia of Genes and Genomes, Institute for Chemical Research, Kyoto University, Japan	Sequence databases and current knowledge on molecular interactions
www.ornl.gov/hgmis/	Life Sciences Division, Department of Energy, Oak Ridge National Laboratory, Tennessee, USA	Links to progress reports, publications, meetings etc, particularly with regards to the Human Genome Project
www.ncbi.nlm.nih.gov/genome/ seq/	National Center for Biotechnology Information, National Institutes of Health, Bethesda, Maryland, USA	Sequence, SNP and literature databases, tools for data mining, human and mouse genetic and physical maps
www.sanger.ac.uk/	Sanger Centre, Wellcome Trust Genome Campus, Cambridge, UK	Progress of the human sequencing project plus many of the other prokaryotic and eukaryotic projects
www.tigr.org/tdb/	The Institute for Genome Research, Rockville, MD, USA	Sequence, function and taxonomy databases for microbes, plants and humans
snp.cshl.org	The SNP Consortium Ltd., Cold Spring Harbour Laboratory, Cold Spring Harbour, New York, USA	SNP map of the human genome
www-genome.wi.mit.edu/	Whitehead Institute Center for Genome Research, Cambridge, Massachusetts, USA	Genetic and physical maps for human, mouse and rat plus human SNP database

X-ray crystallography, nuclear magnetic resonance and other methods to relate structure to gene sequence and function. Genome sequencing and allied projects have also spawned new approaches to mRNA expression analysis or transcriptomics (see Chapter 2; Chee *et al.*, 1996; Eisen *et al.*, 1998; Gerhold *et al.*, 1999) and techniques to undertake comprehensive approaches to the analysis of protein:protein interactions using the two-hybrid assay (Fields and Song, 1989; Fromont-Racine *et al.*, 1997; Lecrenier *et al.*, 1998). Notably, the use of DNA chips and microarrays for high throughput analysis of mRNA expression, sometimes on a genome-wide scale, is having a dramatic impact on the investigation of gene function (Hughes *et al.*, 2000; Young, 2000). More recent advances in microarray technology have increased still further the speed at which sequence information and differential gene expression data may be gathered (Bowtell, 1999; Young, 2000). Together, these approaches are likely to transform our ability to identify and hence target both the desirable and undesirable attributes of organisms.

The application of cDNA arrays to transcriptomics exploits several important features: the relative ease with which the analyses may be undertaken on a comprehensive scale; the ability to automate both the production, and the hybridization and scanning, of the arrays; and the availability of effective software for analysis of the results. As importantly, once mRNAs that alter in expression in response to the conditions under investigation have been identified, it is very straightforward to use the extensive and readily accessible tools of molecular biology to begin to elucidate the expression and function of the genes identified. However, the approach also suffers from several serious limitations, not least of which are the observations that (i) mRNA abundance does not always correlate well with protein abundance, (ii) the sensitivity and dynamic range of existing approaches are such that the lowest abundance mRNAs (potentially encoding the most important regulatory proteins) are not readily measured alongside the more abundant mRNAs, and (iii) the activity of the proteins encoded by mRNAs is regulated at several levels beyond their mRNA or protein expression by, for example, their subcellular localization and/or the extent to which they are post-translationally modified, neither of which are revealed by measurement of mRNA abundance. In addition, there are a number of important biological samples, particularly those that might be used for human diagnostics, such as urine, cerebrospinal fluid and blood plasma, that do not contain mRNA. Moreover, the analysis of mRNA expression in human biopsy and post-mortem samples is still a significant challenge given the potentially protracted time between the collection of the sample and the vulnerability of mRNA to degradation.

The current applications of cDNA arrays have other limitations. Whilst cDNA arrays may be readily available for those model organisms whose genomes have been sequenced, concerns still remain about the availability of tools (cDNAs) for the construction of arrays for non-model organisms for which a significant amount of biology is known. There are also organisms, such as *Plasmodium falciparum,* whose GC/AT ratio is such that the applicability/usefulness of cDNA arrays has yet to be established. Despite these limitations, the application of cDNA arrays and the use of

Hillenkamp, F. (1997) Analysis of proteins by direct scanning infrared-MALDI mass spectrometry after 2D PAGE separation and electroblotting. *Anal. Chem.* **69:** 2888–2892.

Eisen, M.B., Spellman, P.T., Brown, P.O. and Botstein, D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95:** 14 863–14 868.

Fields, S. and Song, O. (1989) A novel genetic system to detect protein-protein interactions. *Nature* **340:** 245–246.

Figeys, D., Gygi, S.P., Zhang, Y., Watts, J., Gu, M. and Aebersold, R. (1998) Electrophoresis combined with novel mass spectrometry techniques: powerful tools for the analysis of proteins and proteomes. *Electrophoresis* **19:** 1811–1818.

Fleischmann, R.D., Adams, M.D., White, O., *et al.* (1995) Whole-genome random sequencing and assembly of haemophilus-influenzae Rd. *Science* **269:** 496–512.

Fromont-Racine, M., Rain, J.C. and Legrain, P. (1997) Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nature Genet.* **16:** 277–282.

Gerhold, D., Rushmore, T. and Caskey, C.T. (1999) DNA chips: promising toys have become powerful tools. *TIBS* **24:** 168–173.

Hughes, T.R., Marton, M.J., Jones, A.R., *et al.* (2000) Functional discovery via a compendium of expression profiles. *Cell* **102:** 109–126.

Klose, J. (1999) Large-gel 2D electrophoresis. *Meth. Mol. Biol.* **112:** 147–172.

Lecrenier, N., Foury, F. and Goffeau, A. (1998) Two-hybrid systematic screening of the yeast proteome. *Bioessays* **20:** 1–5.

Pappin, D.J.C., Hojrup, P. and Bleasby, A.J. (1993) Rapid identification of proteins by peptide-mass fingerprinting. *Current Biology* **3:** 327–332.

Pennington, S.R., Wilkins, M.R., Hochstrasser, D.F. and Dunn, M.J. (1997) Proteome analysis: from protein characterisation to biological function. *Trends Cell Biol.* **7:** 168–173.

Perou, C.M., Jeffrey, S.S., van de Rijn, M., *et al.* (1999) Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc. Natl. Acad. Sci. USA* **96:** 9212–9217.

Roepstorff, P. (1997) Mass spectrometry in protein studies from genome to function. *Curr. Opin. Biotechnol.* **8:** 6–13.

Scheler, C., Lamer, S., Pan, Z., Li, X.P., Salnikow, J. and Jungblut, P. (1998) Peptide mass fingerprint sequence coverage from differently stained proteins on two-dimensional electrophoresis patterns by matrix assisted laser desorption/ionization-mass spectrometry (MALDI-MS). *Electrophoresis* **19:** 918–927.

Wilkins, M.R., Sanchez, J.-C., Gooley, A.A., Appel, R.D., Humphery-Smith, I., Hochstrasser, D.F. and Williams, K.L. (1995) Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnology* **13:** 19–50.

Yates, J.R. (1998) Mass spectrometry and the age of the proteome. *J. Mass Spectrometry* **33:** 1–19.

Young, R.A. (2000) Biomedical discovery with DNA arrays. *Cell* **102:** 9–15.

Contents