

Contents

1	Introduction to the Proteome.	1
	<i>Keith L. Williams and Denis F. Hochstrasser</i>	
1.1	Proteome: a new word, a new field of biology	1
1.2	The proteome and technology	5
1.2.1	Thinking in two dimensions	6
1.2.2	Further dimensions in protein analysis	8
1.2.3	Information and the proteome	9
1.3	Looking towards new frontiers	10
	Acknowledgments	11
	References	11
2	Two-Dimensional Electrophoresis: The State of the Art and Future Directions.	13
	<i>Ben R. Herbert, Jean-Charles Sanchez and Luca Bini</i>	
2.1	Introduction	13
2.2	Sample preparation	15
2.2.1	Increasing protein solubility with chaotropes and surfactants	15
2.2.2	The choice of reducing agent	16
2.2.3	Removal of nucleic acids	16
2.3	Sample loading on IPG gels	18
2.3.1	Sample application during IPG rehydration	18
2.4	Low abundance proteins	19
2.4.1	Cell fractionation and protein prefractionation	20
2.4.2	High protein loads combined with narrow pH gradients	21
2.4.3	Sensitive detection	22
2.5	Basic proteins	24
2.5.1	Separation of basic proteins on IPG gels	24
2.6	Protein quantitation	25
2.7	Future directions for 2-D PAGE	26
2.7.1	Simplifying the IPG-SDS-PAGE interface	27
2.7.2	Fluorescent protein detection	28
2.7.3	High throughput 2-D PAGE	28
2.8	Conclusion	29
	Acknowledgments	29
	References	30

3	Protein Identification in Proteome Projects	35
	<i>Marc R. Wilkins and Andrew A. Gooley</i>	
3.1	The purpose of protein identification	35
3.2	An overview of protein identification strategies.	36
3.3	Primary attributes for protein identification	38
3.3.1	Protein species of origin	38
3.3.2	Protein isoelectric point	38
3.3.3	Protein apparent mass and mass	40
3.3.4	Protein N- and C-terminal sequence tags	41
3.3.5	Extensive N-terminal protein sequence	45
3.4	Secondary attributes for protein identification	46
3.4.1	Attributes of peptides from mass spectrometry	46
3.4.2	Protein amino acid composition	54
3.5	Cross-species protein identification	56
3.6	Future developments and conclusions	60
	Acknowledgments	60
	References.	61
4	The Importance of Protein Co- and Post-Translational Modifications in Proteome Projects	65
	<i>Andrew A. Gooley and Nicolle H. Packer</i>	
4.1	Introduction.	65
4.2	An overview of modifications: what are they and where do they occur?.	68
4.3	Modifications that influence protein charge on 2-D PAGE	70
4.4	Analysis of co- and post-translational modifications after 2-D PAGE.	70
4.4.1	Presentation of the modified protein.	70
4.4.2	Detection of co- and post-translational modifications	74
4.4.3	Analysis of co- and post-translational modifications	76
4.4.4	Mass spectrometry perspectives in the analysis of protein co- and post-translational modifications	83
4.5	Future directions	86
	Acknowledgments	86
	References.	86
5	Proteome Databases	93
	<i>Amos Bairoch</i>	
5.1	Introduction.	93
5.2	Protein sequence databases.	93
5.2.1	SWISS-PROT	95
5.2.2	TrEMBL	100
5.2.3	Completeness and non-redundancy issues	100
5.2.4	Specialised protein sequence databases	102
5.3	Nucleotide sequence databases.	103
5.3.1	Hidden treasures	107
5.4	Pattern and profile databases	108

5.4.1	PROSITE	109
5.4.2	BLOCKS.....	110
5.4.3	PRINTS.....	112
5.4.4	Pfam	112
5.4.5	Other protein domain databases	112
5.4.6	Towards a unified database of protein families and domains.....	114
5.5	2-D PAGE databases.....	114
5.5.1	SWISS-2DPAGE	115
5.5.2	Other 2-D PAGE databases	116
5.6	Three-dimensional structure databases.....	116
5.6.1	PDB.....	117
5.6.2	Swiss-3DImage.....	119
5.6.3	DSSP, HSSP and FSSP.....	119
5.7	Post-translational modification databases.....	119
5.7.1	O-GLYCBASE.....	120
5.8	Genomic databases	120
5.8.1	OMIM.....	121
5.9	Metabolic databases	123
5.9.1	Some specific metabolic databases.....	124
5.10	Conclusions.....	125
5.10.1	How to access all of this data	125
5.10.2	What future for protein databases?.....	126
	References.....	129
	Colour Figures.....	133
6	Interfacing and Integrating Databases	149
	<i>Ron D. Appel</i>	
6.1	Introduction.....	149
6.2	Past data integration techniques (3300 B.C. – 1993 A.D.).....	150
6.2.1	Data exchange on the Internet before WWW.....	151
6.2.2	Hard media data distribution.....	152
6.2.3	Local database integration	153
6.2.4	Discussion of standard techniques	154
6.3	Current integration: a <i>Cyber-Encyclopaedia of the Proteome</i>	154
6.3.1	Cross-references	154
6.3.2	The World-Wide Web	156
6.3.3	Integrated databases on the World-Wide Web	159
6.3.4	Federated proteome databases	162
6.3.5	Discussion of current integration	166
6.4	Integration of databases and analytical methods	166
6.5	The future of proteome database integration	168
6.5.1	Intelligent dedicated search engines.....	168
6.5.2	Enhanced database integration	169
6.5.3	Automation of database construction	170

6.5.4	Advanced data processing over the Internet.....	170
6.5.5	Security and confidentiality	171
6.6	Concluding remarks	172
	Acknowledgments	172
	References.....	173
7	Large-Scale Comparative Protein Modelling.....	177
	<i>Manuel C. Peitsch and Nicolas Guex</i>	
7.1	Introduction.....	177
7.2	Methods and programs	178
7.2.1	Identification of modelling templates.....	178
7.2.2	Aligning the target sequence with the template sequence(s)	179
7.2.3	Building the model	179
7.2.4	Model refinement	179
7.2.5	Assessing the model quality	180
7.2.6	Application of protein models	181
7.2.7	Sequence-to-structure workbench.....	182
7.3	Modelling of Proteomes	182
7.3.1	Large-scale protein modelling	182
7.3.2	Model Repository	183
7.4	Examples.....	184
7.4.1	Obtaining a 3-D structure from the SWISS-MODEL Repository	184
7.4.2	Modelling the Fas-ligand protein using the Swiss-PdbViewer and SWISS-MODEL	184
7.5	Concluding remarks	185
	References.....	186
8	Clinical and Biomedical Applications of Proteomics.....	187
	<i>Denis F. Hochstrasser</i>	
8.1	Introduction.....	187
8.2	Biotechnology, genomics and medicine to date	188
8.3	The application of proteomics to medicine.....	189
8.4	Disease diagnosis from body fluids	190
8.4.1	Glycoproteins in body fluids.....	191
8.5	Proteomics, toxicology and pharmaceuticals	193
8.5.1	Retinoic acid	195
8.5.2	Glycosylation	196
8.5.3	Phosphorylation	197
8.5.4	Drug adducts and vehicles	198
8.5.5	Other pharmacological treatments	198
8.5.6	Hormones	199
8.6	Generation of complexity in proteomes	200
8.7	Proteomics and cancer	200
8.7.1	Cell cycle and cancer theory.....	200
8.7.2	Sample preparation	201

8.7.3	Tissue of origin	202
8.7.4	Metastatic potential and cancer treatment	206
8.7.5	Data analysis and automatic diagnosis	206
8.7.6	A return to cell cycle and oncogene products	208
8.8	Current limitations and future directions of proteomics for medicine . .	210
	References	212
9	Biological Applications of Proteomics	221
	<i>Keith L. Williams and Vitaliano Pallini</i>	
9.1	Scope of this chapter	221
9.2	Proteome maps	221
9.2.1	Mollicutes	222
9.2.2	Prokaryotes	222
9.2.3	Eukaryotes	223
9.3	Applications of proteome technology	226
9.3.1	Tracking complexity	226
9.3.2	Immunogenic proteins	227
9.3.3	Improved agricultural products	228
9.3.4	Value added agricultural products	228
9.3.5	Quality control	229
9.4	Concluding remarks	230
	References	231
10	Conclusions	233
	<i>Denis F. Hochstrasser and Keith L. Williams</i>	
10.1	Scope of this chapter	233
10.2	Large-scale science: genomics, combinatorial chemistry and proteomics	233
10.3	Future methodologies in proteomics	234
10.4	Other views of the future of proteomics	236
	References	237
	Index	239