

Contents

Preface	ix
Acknowledgements	xi
1 Protein, proteome, and proteomics	1
1.1 Primary goals for studying proteomes	1
1.2 Defining the protein	3
1.3 Protein properties – attributes and values	5
1.4 Posttranslational modifications	14
1.5 Protein sequence databases	16
1.6 Identification and characterization of proteins	18
1.7 Two approaches for bottom-up protein analysis by mass spectrometry	20
1.8 Instrument calibration and measuring errors	25
Exercises	28
Bibliographic notes	28
2 Protein separation – 2D gel electrophoresis	31
2.1 Separation on molecular mass – SDS-PAGE	32
2.2 Separation on isoelectric point – IEF	35
2.3 Separation on mass and isoelectric point – 2D SDS-PAGE	36
2.4 2D SDS-PAGE for (complete) proteomics	37
Exercises	40
Bibliographic notes	41
3 Protein digestion	43
3.1 Experimental digestion	46
3.2 <i>In silico</i> digestion	50
Exercises	51
Bibliographic notes	52
4 Peptide separation – HPLC	53
4.1 High-pressure liquid chromatography, HPLC	54
4.2 Stationary phases and separation modes	56
4.3 Component migration and retention time	58
4.4 The shape of the peaks	59

4.5 Chromatography used for protein identification	60
4.6 Chromatography used for quantification	63
Exercises	63
Bibliographic notes	65
5 Fundamentals of mass spectrometry	67
5.1 The principle of mass spectrometry	67
5.2 Ionization sources	69
5.3 Mass analyzers	72
5.4 Isotopic composition of peptides	72
5.5 Fractional masses	73
5.6 The raw data	77
5.7 Mass resolution and resolving power	77
Exercises	79
Bibliographic notes	79
6 Mass spectrometry – MALDI-TOF	81
6.1 TOF analyzers and their resolution	82
6.2 Constructing the peak list	86
6.3 Peak list preprocessing	92
6.4 Peak list format	93
6.5 Automation of MALDI-TOF-MS	93
Exercises	94
Bibliographic notes	94
7 Protein identification and characterization by MS	97
7.1 The main search procedure	98
7.2 The peptide mass comparison	102
7.3 Database search and recalibration	103
7.4 Score calculation	105
7.5 Statistical significance – the P-value	111
7.6 Characterization	115
Exercises	116
Bibliographic notes	117
8 Tandem MS or MS/MS analysis	119
8.1 Peptide fragments	120
8.2 Fragmentation techniques	123
8.3 MS/MS spectrometers	124
8.4 Different types of analyzers	125
8.5 Overview of the process for MS/MS analysis	132
8.6 Fragment ion masses and residue masses	133
8.7 Deisotoping and charge state deconvolution	135
8.8 Precursor treatment	136
8.9 MS ³ spectra	139
Exercises	139
Bibliographic notes	140
9 Fragmentation models	141
9.1 Chemical approach	141
9.2 Statistical approach	142
9.3 Learning (collecting statistics)	144

9.4 The effect of amino acids on the fragmentation	149
Exercises	150
Bibliographic notes	151
10 Identification and characterization by MS/MS	153
10.1 Effect of operations (modifications, mutations) on spectra	154
10.2 Filtering and organization of the database	156
10.3 Scoring and statistical significance	157
Exercises	157
11 Spectral comparisons	159
11.1 Constructing a theoretical spectrum	159
11.2 Non-probabilistic scoring	160
11.3 Probabilistic scoring	164
11.4 Comparison with modifications	169
Exercises	176
Bibliographic notes	177
12 Sequential comparison – <i>de novo</i> sequencing	179
12.1 Spectrum graphs	180
12.2 Preprocessing	183
12.3 Node scores	184
12.4 Constructing the spectrum graph	185
12.5 The sequencing procedure using spectrum graphs	186
12.6 Combined spectra to improve <i>de novo</i> sequencing	189
Exercises	190
Bibliographic and additional notes	191
13 Database searching for <i>de novo</i> sequences	193
13.1 Using general sequence search programs	194
13.2 Specialized search programs	198
13.3 Peptide sequence tags	201
13.4 Comparison by threading	206
Exercises	208
Bibliographic notes	209
14 Large-scale proteomics	211
14.1 Coverage and complexity	211
14.2 Selecting a representative peptide sample – COFRADIC	212
14.3 Separating peptides into fractions	215
14.4 Producing MS/MS spectra	216
14.5 Spectrum filtering	217
14.6 Spectrum clustering	220
14.7 Searching the database	227
14.8 LIMS	227
Exercises	228
Bibliographic notes	228
15 Quantitative MS-based proteomics	229
15.1 Defining the quantification task	229
15.2 mRNA and protein quantification	230
15.3 Quantification of peaks	230

15.4	Normalization	231
15.5	Different methods for quantification	232
15.6	Label-free quantification	232
15.7	Label-based quantification	236
15.8	Variance-stabilizing transformations	240
15.9	Dynamic range	240
15.10	Inferring relative quantity from peptide identification scores	240
15.11	Absolute quantification methods	241
	Bibliographic notes	242
16	Peptides to proteins	243
16.1	Peptides and proteins	243
16.2	Protein identification using peptide masses: an example revisited	243
16.3	Minimal and maximal explanatory sets	246
	Bibliographic notes	247
17	Top-down proteomics	249
17.1	Separation of intact proteins	249
17.2	Ionization of intact proteins	249
17.3	Resolution and accuracy requirements for charge state determination and mass calculation	250
17.4	Fragmentation of intact proteins	251
17.5	Charges of the fragments	252
17.6	Protein identification	252
17.7	Protein characterization – detecting modifications	252
17.8	Problems with top-down approach	253
	Exercises	253
	Bibliographic notes	254
18	Standards	255
18.1	Standard creation	255
18.2	Standards from a proteomics perspective	256
18.3	The Proteomics Standards Initiative	260
18.4	Mass spectrometry standards	260
18.5	Modification standards	262
18.6	Identification standards	262
	Bibliographic notes	263
	Bibliography	265
	Index	277