Table of Contents

Preface.

Acknowledgements.

1 Protein, Proteome, and Proteomics.
   1.1 Primary goals for studying proteomes.
   1.2 Defining the protein.
      1.2.1 Protein identity.
      1.2.2 Splice variants.
      1.2.3 Allelic variants - polymorphisms.
      1.2.4 Posttranslational modifications.
      1.2.5 Protein isoforms.
   1.3 Protein properties - attributes and values.
      1.3.1 The amino acid sequence.
      1.3.2 Molecular mass.
      1.3.3 Isoelectric point.
      1.3.4 Hydrophobicity.
      1.3.5 Amino acid composition.
   1.4 Posttranslational modifications.
   1.5 Protein sequence databases.
      1.5.1 UniProt KnowledgeBase (Swiss-Prot/TrEMBL, PIR).
      1.5.2 The NCBI non-redundant database.
      1.5.3 The International Protein Index (IPI).
      1.5.4 Time-instability of sequence databases.
   1.6 Identification and characterization of proteins.
      1.6.1 Top-down and bottom-up proteomics.
      1.6.2 Protein digestion into peptides.
   1.7 Two approaches for bottom-up protein analysis by mass spectrometry.
      1.7.1 MS - Peptide mass fingerprinting.
      1.7.2 MS/MS - Tandem MS.
      1.7.3 Combination approaches.
      1.7.4 Reducing the search space.
   1.8 Instrument calibration and measuring errors.
      1.8.1 Calibration.
      1.8.2 Accuracy and precision.
1.9 Exercises.
1.10 Bibliographic notes.

2 Protein Separation - 2D Gel Electrophoresis.
2.1 Separation on molecular mass - SDS-PAGE.
2.1.1 Estimating the protein mass.
2.2 Separation on isoelectric point - IEF.
2.3 Separation on mass and isoelectric point, 2D.
2.3.1 Transferring the proteins from the first to the second dimension.
2.3.2 Visualizing the proteins after separation.
2.3.3 Problems.
2.3.4 Excising the proteins.
2.4 2D SDS-PAGE for (complete) proteomics.
2.4.1 Identifying the proteins.
2.4.2 Quantification.
2.4.3 Programs for treating and comparing gels.
2.4.4 Comparing results from different experiments - DIGE.
2.5 Exercises.
2.6 Bibliographic notes.

3 Protein Digestion.
3.1 Experimental digestion.
3.1.1 Cleavage specificity.
3.1.2 Trypsin.
3.1.3 Chymotrypsin.
3.1.4 Other considerations for the choice of a protease.
3.1.5 Random cleavage.
3.1.6 Chemical cleavage.
3.1.7 In-gel digestion.
3.2 In silico digestion.
3.3 Exercises.
3.4 Bibliographic notes.

4 Peptide Separation - HPLC.
4.1 High Pressure Liquid Chromatography - HPLC.
4.2 Stationary phases and separation modes.
4.2.1 Reverse phase chromatography, RP.
4.2.2 Strong cation exchange chromatography, SCX.
4.2.3 Other types of chromatography for proteomics.
4.2.4 Tandem HPLC.
4.3 Component migration and retention time.
4.4 The shape of the peaks.
4.4.1 The width.
4.4.2 Asymmetry.
4.4.3 Resolution.
4.5 Chromatography used for protein identification.
4.5.1 Theoretical calculation of the retention time for reverse phase chromatography.
4.6 Chromatography used for quantification.
4.7 Exercises.
4.8 Bibliographic notes.

5 Fundamentals of Mass Spectrometry.
5.1 The principle of mass spectrometry.
5.2 Ionization sources.
5.2.1 MALDI - Matrix Assisted Laser Desorption Ionization.
5.2.2 ESI - Electrospray ionization.
5.2.3 Other ionization sources.
5.3 Mass analyzers.
5.4 Isotopic composition of peptides.
5.4.1 Estimating the charge.
5.5 Fractional masses.
5.5.1 Estimating one or two peptides in a peak complex.
5.6 The raw data.
5.7 Mass resolution and resolving power.
5.7.1 Isotopic resolution.
5.8 Exercises.
5.9 Bibliographic notes.

6 Mass Spectrometry - MALDI-TOF.
6.1 Time-of-flight analyzers and their resolution.
6.1.1 Time-to-mass converter.
6.1.2 Producing spectra.
6.1.3 Ionization statistics.
6.2 Constructing the peak list.
6.2.1 Noise.
6.2.2 Baseline correction.
6.2.3 Smoothing and noise reduction.
6.2.4 Peak detection.
6.2.5 Example.
6.2.6 Intensity normalization.
6.2.7 Calibration.
6.3 Peak list preprocessing.
6.3.1 Monoisotoping and deisotoping.
6.3.2 Removing spurious peaks.
6.4 Peak list format.
6.5 Automation of MALDI-TOF-MS.
6.6 Exercises.
6.7 Bibliographic notes.
7 Protein Identification and Characterization by MS.
7.1 The main search procedure.
7.1.1 The experimental data.
7.1.2 The database - the theoretical data.
7.1.3 Other search parameters.
7.1.4 Organization of the database.
7.2 The peptide mass comparison.
7.2.1 Reasons why experimental masses may not match.
7.3 Database search and recalibration.
7.3.1 The search program MSA (Mass Spectra Analyzer).
7.3.2 Aldente.
7.4 Score calculation.
7.4.1 Score components.
7.4.2 Scoring scheme examples.
7.4.3 Identification from a protein mixture.
7.5 Statistical significance - the P-value.
7.5.1 A priori probability for k matches.
7.5.2 Simulation for determining the P-value.
7.5.3 A simple Mascot search.
7.6 Characterization.
7.7 Exercises.
7.8 Bibliographic notes.
8 Tandem MS or MS/MS Analysis.
8.1 Peptide fragments.
8.2 Fragmentation techniques.
8.3 MS/MS spectrometers.
8.3.1 Analyzers for MS/MS.
8.4 Different types of analyzers.
8.4.1 TOF/TOF.
8.4.2 Triple quadrupole (Triple quad).
8.4.3 Ion trap (IT).
8.4.4 Fourier Transform Ion Cyclotron Resonance (FT-ICR).
8.4.5 Combining quadrupole and Time of flight - Q-TOF.
8.4.6 Combining quadrupole and ion trap - Q-TRAP.
8.4.7 Combining TOF and Ion trap.
8.4.8 Combining Linear ion trap with Orbitrap.
8.4.9 Characteristics and performances of some type of analyzers.
8.5 Overview of the process for MS/MS analysis.
8.6 Fragment ion masses and residue masses.
8.7 Deisotoping and charge state deconvolution.
8.8 Precursor treatment.
8.8.1 Precursor mass correction.
8.8.2 Estimating the charge state of the precursor.
8.9 MS3 spectra.
8.10 Exercises.
8.11 Bibliographic notes.

9 Fragmentation Models.
9.1 Chemical approach.
9.1.1 The mobile proton model, MPM.
9.2 Statistical approach.
9.2.1 Constructing the training set(s).
9.2.2 Spectral subsets.
9.3 Learning (collecting statistics).
9.3.1 Fragmentation Intensity Ratio (FIR).
9.3.2 Linear models.
9.3.3 Use of decision trees.
9.4 The effect of amino acids on the fragmentation.
9.4.1 Selective fragmentation.
9.5 Exercises.
9.6 Bibliographic notes.

10 Identification and Characterization by MS/MS.
10.1 Effect of operations (modifications - mutations) on spectra.
10.1.1 Comparison including modifications.
10.2 Filtering and organization of the database.
10.3 Scoring and statistical significance.
10.4 Exercises.

11 Spectral Comparisons.
11.1 Constructing a theoretical spectrum.
11.2 Non-probabilistic scoring.
11.2.1 Number and intensities of matching peaks or intervals.
11.2.2 Spectral contrast angle.
11.2.3 Cross-correlation.
11.2.4 Rank based scoring.
11.2.5 SEQUEST scoring.
11.3 Probabilistic scoring.
11.3.1 Bayesian method - SCOPE.
11.3.2 Use of log-odds - OLAV.
11.3.3 Log-odds decision trees.
11.4 Comparison with modifications.
11.4.1 Zone modification searching.
11.4.2 Spectral convolution and spectral alignment.
11.5 Exercises.
11.6 Bibliographic notes.

12 Sequential Comparison - de novo Sequencing.
12.1 Spectrum graphs.
12.1.1 A general spectrum graph.
12.2 Preprocessing.
12.3 Node scores.
12.4 Constructing the spectrum graph.
12.5 The sequencing procedure using spectrum graphs.
12.5.1 Searching the graph.
12.5.2 Scoring the derived sequences against the spectrum.
12.6 Combined spectra to improve de novo sequencing.
12.6.1 Use of two fragmentation techniques.
12.7 Exercises.
12.8 Bibliographic and additional notes.

13 Database Searching for De Novo Sequences.
13.1 Using general sequence search programs.
13.1.1 The main principle of FASTA and BLAST.
13.1.2 Changing the operation of FASTA/BLAST.
13.1.3 Scoring and statistical significance.
13.2 Specialized search programs.
13.2.1 OpenSea.
13.2.2 SPIDER.
13.3 Peptide sequence tags.
13.3.1 A general model for peptide sequence tag search programs.
13.3.2 Automatic extraction and scoring of sequence tags.
13.3.3 Database search.
13.3.4 Extending the sequence tag hits with flanking amino acids.
13.3.5 Scoring the PST matches.
13.3.6 Statistical significance.
13.4 Comparison by threading.
13.4.1 Use of suffix tree.
13.4.2 Use of deterministic finite automata.
13.5 Exercises.
13.6 Bibliographic notes.

14 Large-Scale Proteomics.
14.1 Coverage and complexity.
14.2 Selecting a representative peptide sample - COFRADIC.
14.3 Separating peptides into fractions.
14.4 Producing MS/MS spectra.
14.5 Spectra filtering.
14.5.1 Classifying good and bad spectra.
14.5.2 Use of the classifier.
14.6 Spectrum clustering.
14.6.1 Recognizing sibling spectra.
14.6.2 Clustering of sibling spectra.
14.6.3 Representative spectra for the groups.
14.6.4 De novo sequencing from representative PRM spectra.
14.7 Searching the database.
14.8 LIMS.
14.9 Exercises.
14.10 Bibliographic notes.

15 Quantitative Mass Spectrometry-Based Proteomics.
15.1 Defining the quantification task.
15.2 mRNA and protein quantification.
15.3 Quantification of peaks.
15.4 Normalization.
15.5 Different methods for quantification.
15.6 Label-free quantification.
15.6.1 Comparing spectra.
15.6.2 MALDI-TOF based methods.
15.6.3 SELDI-TOF based methods.
15.6.4 LC-MS quantification.
15.7 Label-based quantification.
15.7.1 MS-based labelled quantification.
15.7.2 MS/MS-based quantification.
15.8 Variance stabilizing transformations.
15.9 Dynamic range.
15.10 Inferring relative quantity from peptide identification scores.
15.11 Absolute quantification methods.
15.12 Bibliographic notes.
16 Peptides to Proteins.
16.1 Peptides and proteins.
16.2 Protein identification using peptide masses: an example revisited.
16.2.1 Extension to MS/MS derived peptide sequences instead of masses.
16.3 Minimal and maximal explanatory sets.
16.3.1 Minimal and maximal sets in peptide-centric proteomics.
16.3.2 Determining maximal explanatory sets.
16.3.3 Determining minimal explanatory sets.
16.4 Bibliographic notes.

17 Top-Down Proteomics.

17.1 Separation of intact proteins.
17.2 Ionization of intact proteins.
17.3 Resolution and accuracy requirements for charge state determination and mass calculation.
17.4 Fragmentation of intact proteins.
17.5 Charges of the fragments.
17.6 Protein identification.
17.7 Protein characterization - detecting modifications.
17.8 Problems with top-down approach.
17.9 Exercises.
17.10 Bibliographic notes.

18 Standards.

18.1 Standard creation.
18.1.1 Types of standards.
18.2 Standards from a proteomics perspective.
18.2.1 Creation of test samples.
18.2.2 Data standards in proteomics.
18.2.3 Requirements for data standards.
18.2.4 Problems with data standards.
18.3 The Proteomics Standards Initiative (PSI).
18.3.1 Minimal reporting requirements.
18.4 Mass spectrometry standards.
18.5 Modification standards.
18.6 Identification standards.
18.7 Bibliographic notes.

Bibliography

Index.