Advances in Experimental Medicine and Biology 1073
Proteomics, Metabolomics, Interactomics and Systems Biology

José-Luis Capelo-Martínez *Editor*

Emerging Sample Treatments in Proteomics

Springer

Advances in Experimental Medicine and Biology

Proteomics, Metabolomics, Interactomics and Systems Biology

Advances in Experimental Medicine and Biology presents multidisciplinary and dynamic findings in the broad fields of experimental medicine and biology. The wide variety in topics it presents offers readers multiple perspectives on a variety of disciplines including neuroscience, microbiology, immunology, biochemistry, biomedical engineering and cancer research. *Advances in Experimental Medicine and Biology* has been publishing exceptional works in the field for over 30 years and is indexed in Medline, Scopus, EMBASE, BIOSIS, Biological Abstracts, CSA, Biological Sciences and Living Resources (ASFA-1), and Biological Sciences. The series also provides scientists with up-to-date information on emerging topics and techniques. 2018 Impact Factor: 1.760

More information about this series at http://www.springer.com/series/15040

José-Luis Capelo-Martínez
Editor

Emerging Sample Treatments in Proteomics

Springer

Editor
José-Luis Capelo-Martínez
BIOSCOPE Group, LAQV-REQUIMTE
Departamento de Química
Faculdade de Ciências e Tecnologia
Universidade Nova de Lisboa
Caparica, Portugal

Proteomass Scientific Society
Rua dos Inventores
Madan Park
Caparica, Portugal

ISSN 0065-2598 ISSN 2214-8019 (electronic)
Advances in Experimental Medicine and Biology
ISBN 978-3-030-12297-3 ISBN 978-3-030-12298-0 (eBook)
https://doi.org/10.1007/978-3-030-12298-0

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The word "proteome" was coined in 1994 by Marc Wilkins to define a group of proteins encoded by a genome and expressed in particular cells, tissues, organs, or organisms. A quarter of a century later, proteomics, the science that studies the proteome, has become one of the most important and popular scientific fields for the large-scale study of complex cell-based systems, largely boosted by the parallel development of affordable high-resolution mass spectrometers and powerful bioinformatics tools. As a current example, the hyphenation of ion-mobility and time-of-flight-based mass spectrometry with parallel accumulation-serial fragmentation software has made it possible to quantify about 1500 proteins in 20 minutes, a staggering throughput at an incredible speed [1].

Nowadays, the only limit to the wide range of proteomics applications is our imagination. The study of biological functions, understanding changes in cellular regulation mechanisms caused by disease states, discovery of biomarkers for disease diagnosis, prognosis, and the development of new drugs for therapeutic approaches can all benefit from the use of proteomics. Just 10 years ago, most proteomics studies relied on different but complementary tools, such as two-dimensional gel electrophoresis, chromatographic separation, and mass spectrometry. Protein quantification was, and in many cases still is, done using stable isotope labeling methods. Protein quantification by label-free shot-gun ion-mobility high-resolution mass spectrometry is soon likely to supersede isotope-based methods. Software for data collection and analysis has also greatly evolved, moving from data-dependent to data-independent acquisition.

However, despite such technological advances, the procedures used to handle samples in the laboratory are still complex, lengthy, and laborious. Sample collection and preservation and proteome extraction, purification, reduction, alkylation, and digestion are key steps that greatly influence the pipeline of mass spectrometry-based quantification, especially the reliability of the results. One of the most important steps in any protein identification or quantitation workflow is the digestion, or hydrolysis, of the proteome. This crucial step is traditionally performed with proteases, such as trypsin, for 12–48 hours. With the advent of the so-called ProteoSonic

method, the digestion of 96 proteomes can be performed in less than 5 minutes, an unprecedented speed of sample throughput [2].

The improvement of two critical steps remains to be tackled before protein quantification, using mass spectrometry, can be considered a routine technique. One is the development of software to handle the vast quantity of information generated with new ion mobility high-resolution mass spectrometers. The other, sample treatment, is the subject of this book. There is no universal treatment to extract the proteome from all types of samples. This is because the matrixes of cells from different tissues and organisms present different constraints due to their specific physical and chemical properties, so different approaches are required to successfully extract as many of the proteins as possible. For this reason, after a first chapter addressing the necessary tools of the trade of modern proteomics, eight chapters follow, focusing on different sources of biological samples, namely, saliva, plasma, serum, animal tissues, urine, feces, plants, and bacteria. Each chapter describes in detail the current sample treatments used daily by researchers in the analytics and proteomics communities, including the pitfalls and how to overcome them, and the pros and cons to consider.

I would like to thank Professor Daniel Martins-de-Souza, who kindly encouraged me to edit this book, and the Springer Editorial team, who kindly accepted my book proposal. I send my deepest gratitude to all chapter contributors and thank them for their valuable insights and shared interest in this vibrant field.

Caparica, Grande Lisboa, Portugal José-Luis Capelo-Martínez

References

1. timsTOF Pro – The new standard for shotgun proteomics. https://www.bruker.com/products/mass-spectrometry-and-separations/lc-ms/o-tof/timstof-pro/overview.html. Last accessed 24 March 2019
2. Jorge S, Araújo JE, Pimentel-Santos FM, Branco JC, Santos HM, Lodeiro C, Capelo JL (2018) Unparalleled sample treatment throughput for proteomics workflows relying on ultrasonic energy. Talanta 178:1067–1076. https://doi.org/10.1016/j.talanta.2017.07.079

Contents

Chapter 1
Proteomics: Tools of the Trade

Utpal Bose, Gene Wijffels, Crispin A. Howitt, and Michelle L. Colgrave

1.1 Introduction

Proteins are the functional traits of life, and they play critical roles in orchestrating various biological processes in order to define the cellular or organismal phenotype. Collectively the protein complement of a cell, organ or organism is known as the proteome. The proteome describes a dynamic system where the presence and abundance of proteins facilitate a multitude of processes and regulatory mechanisms. Individual proteins can carry out functions in living cells, or they can interact to form higher order structures and/or networks that work synergistically within a complex system [1]. The structure and function of proteins are altered by post-translational modifications, protein-protein interactions, their subcellular location and as a result of synthesis and degradation the progression of biological processes, such as cell proliferation, migration and development [2]. In order to understand the underlying mechanism of physiological processes, it is critical to be able to identify and quantify proteins and their interactions within these dynamic systems.

Sample preparation is an integral and crucial part of any proteomics experiment. The primary goal of sample preparation is to obtain high-quality protein samples which are representative of the proteome. Thus sample preparation is critical to maximising protein identification and quantitation rates in downstream analyses. Biological samples are complex, and they contain significant amounts of non-protein components that can interfere with protein analyses. To extract, identify, quantify and characterise proteins from these samples, diverse protein extraction

U. Bose · G. Wijffels · M. L. Colgrave (✉)
Agriculture and Food, CSIRO, St Lucia, QLD, Australia
e-mail: Utpal.bose@csiro.au; gene.wijffels@csiro.au; michelle.colgrave@csiro.au

C. A. Howitt
Agriculture and Food, CSIRO, Canberra, ACT, Australia
e-mail: crispin.howitt@csiro.au

J.-L. Capelo-Martínez (ed.), *Emerging Sample Treatments in Proteomics*, Advances in Experimental Medicine and Biology 1073, https://doi.org/10.1007/978-3-030-12298-0_1

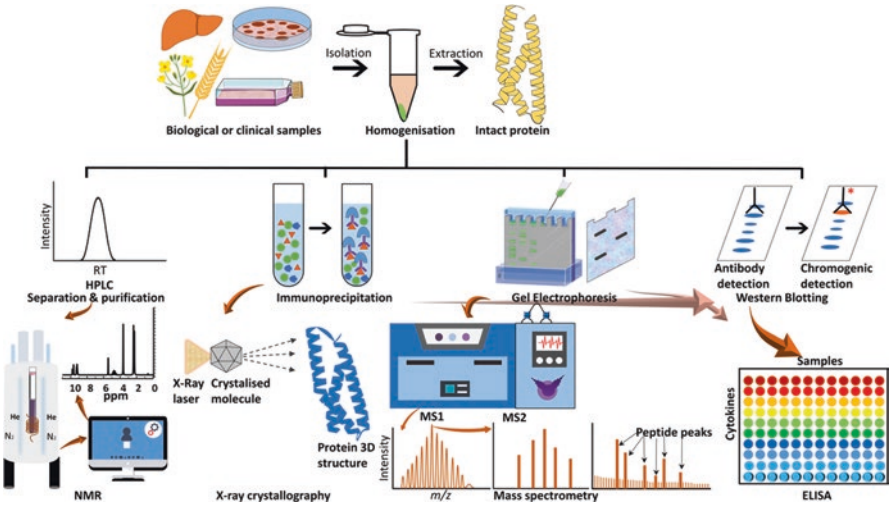

Fig. 1.1 Sample preparation for antibody-based techniques (e.g. immunoprecipitation, gel electrophoresis and Western blotting), structural proteomics (e.g. NMR and X-ray crystallography) and top-down MS (e.g. MALDI)

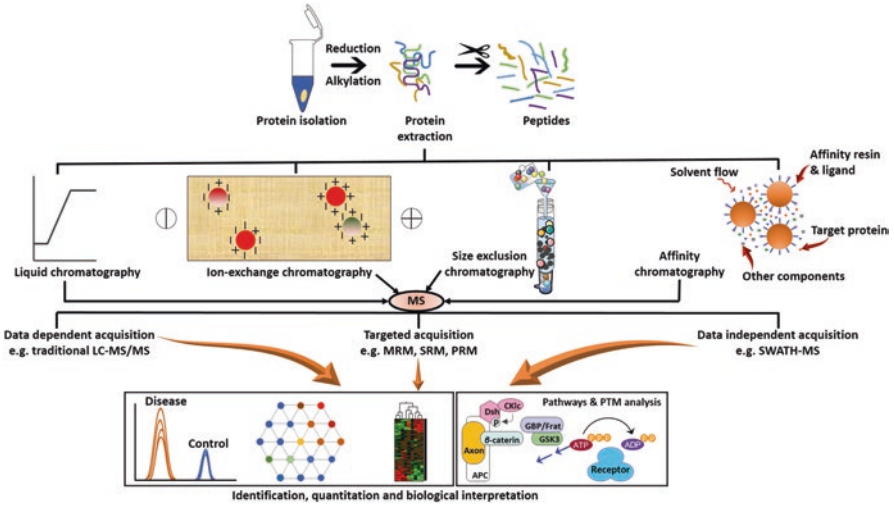

Fig. 1.2 Steps involved in MS-based proteomics sample preparation and data analysis

and purification techniques have been used to prepare samples for the different types of analytical methods. Proteomics experiments typically comprise at least four steps: (1) sample collection and storage (e.g. sample collection from biological or clinical specimen and storage at required temperature); (2) protein extraction (e.g. reagent-based or mechanical disruption); (3) protein clean-up or purification

(e.g. chromatography, protein precipitation, enrichment or depletion); and (4) protein analyses (e.g. antibody-based techniques, mass spectrometry, structural studies). The optimal sample preparation method will be highly dependent on the analytical goal. For instance, sample preparation for liquid chromatography-mass spectrometry (LC-MS) proteomics will be vastly different to that used for nuclear magnetic resonance (NMR) structure elucidation experiments [3]. Less sample is required for LC-MS-based bottom-up proteomics, but the workflow will typically include a protein digestion step prior to identification or quantitation of the proteins; whereas a substantially higher amount of highly purified target protein is required for NMR-based structural elucidation experiments.

Proteins play critical roles in various cellular processes, and their complete characterisation is of great interest. However, it is technically challenging to determine the presence, abundance, modification state(s) and localisation of every protein in a system due to the complex environment in which they exist. An array of well-established methods and protocols including immunohistochemistry (IHC), immunoprecipitation (IP), Western blotting (WB) or enzyme-linked immunosorbent assay (ELISA), in vitro protein assays and in vivo imaging are used to study the presence, sequence, function and biochemical mode of action of one or a few specific proteins. With techniques such as mass spectrometry (MS), it is now possible to list the cast of protein players and using systems biology approaches start to elucidate the cellular and/or organismal priorities in a particular state or time point [1]. For example, LC-MS proteomics quantified ~11,000 proteins from human heart [4] or yeast (*Saccharomyces cerevisiae*) [5] within a few hours.

The ideal protein extraction and identification method should capture the most comprehensive repertoire of proteins, and ideally their abundances, while minimising protein degradation and/or the effect of non-protein interferences. This chapter will focus on sample pretreatment for proteomics experiments as it applies to commonly used methods and instruments used in proteomics experiments in the context of their principles and experiment-specific applications (Figs. 1.1 and 1.2). The overall considerations in proteomic sample preparation are given in Box 1.1.

Box 1.1 Considerations for the Protein Sample Preparation Strategy

- *Aim:* What is the ultimate goal of the project, identification, characterisation or quantitation?
- *Type:* What is the sample type? Eukaryotes, prokaryotes, plants, etc., vary in their cell/tissue structure requiring different approaches.
- *Scale:* How much sample do you have? Can you cover technical and biological replicates? Concentration or enrichment steps might be required. Selection of a sensitive method of analysis may overcome low sample amounts.
- *Localisation:* Where is/are the protein/s of interest localised? Reagent and protocol selection will define the success in isolating protein(s) from a

Box 1.1 (continued)
subcellular fraction or organelle. Subcellular fractionation can assist in enriching low-abundant components. Harsher solvents or strong detergents are often required to efficiently extract intrinsic membrane proteins.

- *Dynamic range:* Is/are the protein(s) of interest in a complex matrix? Is the aim to extract all proteins or a particular target? The detection and/or quantitation of low-abundance target proteins may require depletion, enrichment or purification steps and increase the amount of sample required due to inevitable losses and degradation.
- *Preservation:* Is protein structure or function important? Inhibitors (e.g. protease or phosphatase) may be added to preserve the proteins. Should thermal denaturation of enzymes be considered to limit post-collection changes?
- *Analysis platform:* What type of downstream analytical platform or methodologies will be used? Gels, WB, ELISA, MS, NMR, X-ray crystallography. The level of purity is dictated by the downstream analyses. For example, detergents and salts interfere with MS and need to be removed. Highly purified protein is often required for protein structure identification by NMR or X-ray crystallography.

1.2 Steps Involved in Proteomics Sample Preparation

1.2.1 Sample Collection

Sample collection from biological or clinical sources is the first step in proteomic sample preparation. Samples can be obtained from naïve source material such as whole tissues, body fluids or cell culture. Samples representing all kingdoms—animal, plants, fungi, or from the prokaryotic sources such as bacteria—are commonly investigated with each sample type presenting unique challenges in their analysis. During sample collection, the cells or tissues are removed from their origin, which can cause disturbances in homeostasis. This step can trigger the activity of various proteases and enzymes that are responsible for post-translational modifications (PTMs) leading to alteration or degradation of protein components. Numerous strategies have been used to reduce the chance of protein loss or modification. These include rapid freezing in liquid nitrogen, chemical- and/or temperature-induced denaturation of protein-modifying enzymes and the addition of protective or stabilising compounds such as reducing agents of protease inhibitors. Protein stabilisation may also achieved by thermal denaturation or protein precipitation. Should samples need to be preserved for future experiments, tissue samples or extracted proteins can be stored at −80 °C. However, the rapid extraction of proteins from frozen samples is preferred as some samples may be less stable than others.

1.2.2 Protein Extraction

Protein sample extraction from biological or clinical sources can profoundly affect the experimental outcome. Generally, protein extraction methods can be divided into two main categories: gentle processes, for instance, chemical and osmotic lysis, liquid or liquid/liquid extraction and temperature treatments, and harsher processes, such as mechanical disruption via ultrasonication, grinding and/or pressure homogenisation. In chemical-based lysis processes, denaturing solutions (lysis buffers) are often used to disturb protein structural elements, leading to disruption of protein-protein interaction and protein unfolding and thus facilitating protein extraction. Strong denaturants (e.g. urea, guanidine, heat, detergents) and extremes of pH and ionic detergents (e.g. sodium dodecyl sulphate (SDS), deoxycholate and cetyltrimethylammonium bromide) are often used to solubilise and denature membrane proteins. Non-ionic or zwitterionic detergents (e.g. Triton X-100, NP-40, digitonin or CHAPS) have a lower micelle concentration and thus require a lower amount to solubilise proteins [6]. In cells and prokaryotes, lysozymes (enzyme) can be used to lyse erythrocytes and for the extraction of periplasmic proteins from *E. coli* by subcellular fractionation [7]. In addition to chemical-based approaches, liquid or liquid/liquid protein extraction methods have been included in the workflow to remove cell debris and as a preliminary clean-up step in proteomic sample preparation [8]. An appropriate combination of organic solvents can be selected to extract proteins based on their phase separation or partitioning. The extraction efficiency can be improved by altering the pH and temperature, varying the ionic strength and salt type and changing polymer or surfactant concentrations.

Although chemical-based extraction is often the first choice, mechanical or physical disruption methods are particularly useful in protein extraction from plant tissues. The most common methods in this category include grinding under liquid nitrogen (mostly for plant tissues), use of ultrasonic probes, lysis by osmotic shock, grinding with aluminium or sand, pressure cycling, use of bead beaters/devices and thermal lysis [9, 10]. Mechanical disruption methods can suffer from poor reproducibility in part resulting from non-homogeneity of the finished sample after processing. To mitigate this issue, chemical/solvent-based extraction is often used in combination with homogenisation and mechanical grinding to efficiently extract proteins from diverse sample types.

1.2.3 Protein Depletion

Biological and clinical samples represent an excellent opportunity to discover biomarkers for their future application as diagnostics or as therapeutic targets. However, it can be challenging to identify and consistently quantify low-abundance biomarker proteins in the presence of high-abundance proteins [11]. For example, human plasma and serum protein concentrations in the blood can span up to 12 orders of magnitude [12]. They contain a substantial amount of albumin and immunoglobulin

that can comprise up to 70% of the protein content [11]. Different protein depletion strategies, such as protein precipitation and antibody column-based approaches, have been developed to mitigate this problem, where the ultimate goal is to separate high-abundance proteins from the low-abundance proteins. Protein depletion methods, such as albumin precipitation [13, 14] or column-based immunodepletion [15], and commercial kits such as ProteoPrep® 20 (Sigma-Aldrich), Pierce™ Top (Thermo Fisher Scientific) and Aurum Affi-Gel Blue Mini Columns (Bio-Rad) have been developed to substantially reduce the high abundance proteins from blood serum and plasma in order to identify low-abundance proteins.

In plants, ribulose-1,5-bisphosphate carboxylase oxygenase (Rubisco) represents 30–50% of the total protein in green tissues, which imposes a consistent challenge for both gel-based and shotgun proteomics analysis [16]. Several strategies, such as polyethylene glycol (PEG) fractionation [17], protamine sulphate precipitation [18], Rubisco IgY affinity [19] and polyethyleneimine-assisted Rubisco clean-up [16], have been used to deplete Rubisco from plant samples. Although protein precipitation is commonly used to remove high abundant proteins from plant samples, it can be difficult to re-solubilise proteins after precipitation leading to significant losses of protein yield [20].

During the depletion process, it is possible to lose other proteins due to non-specific binding to the depleted proteins and resin surfaces [21]. For example, comparative experiments between non-depleted serum and serum depleted of the six most abundant proteins have shown that while depletion significantly increased the number of proteins analysed and identified, some of the proteins found in the non-depleted serum were not found in the depleted serum [22, 23]. Therefore, when optimising protein depletion steps, the depleted protein resin bound fraction should be analysed to confirm that target proteins are not inadvertently lost.

1.2.4 Protein Enrichment

Enrichment strategies aim to enhance the concentration of particular classes of proteins (protein families or modified proteins, e.g. phosphoproteins) concomitantly reducing sample complexity for subsequent analysis and identification. This process is vital when the target proteins represent a minor component or have been post-translationally modified. Enrichment can be achieved by subcellular fractionation wherein centrifugation is the most straightforward protocol employed. Proteins located in the subcellular components, such as the membrane, mitochondria or nucleus, can be isolated to individual components based on their weight, size and shape through repeated centrifugation and velocity and equilibrium sedimentation [24, 25]. Centrifugation is also applicable to fractionation of protein complexes. Solvent- and/or chemical-based protein precipitation can be another choice to enrich low-abundance plant seed proteins [26].

Targeted strategies for enrichment include the use of IP, or through PTM-specific affinity binding, such as phosphoprotein enrichment [27]. These techniques are limited by the availability of suitable antibodies. During the binding process,

proteins or peptides bind to the antibody or receptor surface while unrelated molecules remain free in the sample. Washing steps aim to eliminate non-specific binding, before the bound proteins are eluted from the receptors.

Electrophoretic methods can be used to separate proteins based on their size, shape and charge. Some of the most common electrophoretic pre-fractionation methods include electrokinetic methodologies which rely on isoelectric focusing (IEF) steps [28]. Frequently chromatography-based techniques are used in combination with subcellular fractionation approach to maximise the protein yield. For instance, subcellular fractionation technique combined with "loss-less" high-pH reversed-phase fractionation was used to isolate and extract low-abundant regulatory proteins from the human heart, mouse liver and brain [4, 29, 30].

To enrich post-translationally modified proteins from cell or tissue extracts, metal-ion affinity chromatography, titanium dioxide chromatography, superparamagnetic nanoparticles, affinity tags and high-affinity antibodies can be used [27, 31–33].

1.2.5 Normalisation

Cellular proteins span a wide dynamic range in biological samples. The most abundant components often dominate analyses hiding the minor components. Decreasing the concentration range to match the dynamic range of the analytical platform assists with the identification of low abundance proteins. Use of a combinatorial hexapeptide-bead library, for example, commercially available as ProteoMiner by Bio-Rad [34], can minimise the protein concentration dynamic range in serum and plasma samples [35]. This technique can reduce the amount of high-abundance proteins without immunodepletion while preventing the loss of low-quantity proteins bound to high-abundance proteins. The approach has also been extended to enrich for glycopeptides and phosphopeptides, which enables the study of post-translational modifications [36]. Use of this method for proteomics applications has been reviewed [35].

1.2.6 Protein Clean-Up

Protein isolation and extraction, enrichment or depletion steps introduce chemicals, such as detergents, salts, buffers and chaotropic agents, to improve protein solubility. The presence of these chemicals and biomolecules that coexist in the sample, e.g. lipids, can interfere with subsequent analysis, and their removal is not a straightforward process. Numerous methods have been described to remove detergents, chemicals and non-protein molecules from biological and clinical samples. For example, solid-phase extraction (SPE, e.g. StageTips and ZipTips), protein precipitation (e.g. trichloroacetic acid (TCA)/acetone), filters (ultrafiltration), chromatography-based techniques (high-pressure liquid chromatography (HPLC), fast protein

liquid chromatography (FPLC), size-exclusion, membrane, ion-exchange chromatography) and affinity-based techniques are frequently applied in labs [37–42].

In SPE-based techniques, compounds are separated based on their chemical and physical properties, which determines their distribution between a mobile liquid phase and a solid stationary phase. After protein (or peptide) binding, the remaining compounds are washed away and the bound proteins are subsequently eluted from the solid phase by altering the mobile phase. For example, StageTips contain very small disks made of beads that are available in a variety of chemistries, reversed-phase, cation-exchange or anion-exchange surfaces, embedded in a Teflon mesh, which bind molecules based on their chemical nature [40]. Although SPE-based approaches have been used in proteomics workflows, they are sample limited (dependent on surface area/volume), are typically only capable of delivering crude fractions from complex matrices, and may not give complete recovery required for high-sensitivity experiments [43].

Precipitation buffers are used to remove interfering molecules from the proteomic samples [44]. TCA/acetone maintains the solubility of non-protein molecules while precipitating most proteins. Although protein precipitation is a rapid, simple and inexpensive method, the inherent drawback for this method is the fact that proteins may not fully resolubilise after TCA/acetone precipitation. Additionally, this precipitation solution is unable to extract the maximum number of proteins from recalcitrant plant tissues [26, 44].

Chromatography-based techniques are often included in the proteomic workflow in order to enrich for target proteins and reduce sample complexity. HPLC can be used online to separate the proteins/peptides and thereby increase the depth of coverage. When employed offline, HPLC can purify or enrich proteins/peptides of interest, typically using gradient elution and offline fraction collection. HPLC can be tailored to particular protein classes, e.g. polar or non-polar protein separation. Additionally ion-exchange chromatography is a complementary, but orthogonal, technique capable of separating proteins on their charge state via ion-ion interactions between the protein analyte and the stationary phase [45]. Additional to these techniques, size-exclusion chromatography (SEC) can be used to separate proteins based on their molecular size. SEC is a commonly used technique, because of the diversity of the molecular weights of proteins in biological tissues and extracts. SEC has been employed for many roles, including buffer exchange (desalting) or removal of non-protein contaminants (DNA, viruses) [10].

Affinity chromatography, also known as affinity purification, is often used to characterise protein complexes and large protein interaction networks [46]. This technique depends on the utilisation of sample-specific antigen-antibody recognition and binding. The antibody is chemically immobilised or ‘coupled’ to a solid support so that when a complex mixture is passed over the column, those molecules having the specific binding affinity to the antibody are captured. Other sample components are washed away before the bound targets are eluted from the support, resulting in their enrichment from the original sample. Affinity-tagged-based recombinant protein purification can be used where proteins of interest are simply expressed in-frame with an epitope tag (at either the N- or C-terminus), which is then used as an affinity handle to purify the tagged protein (the bait) along with its interacting partners (the prey) [46].

1.3 Analytical Tools for Proteomics

A range of tools and methods are available to obtain information about proteins present in biological samples (Figs. 1.1 and 1.2). These diverse methods yield different types of information, and the choice of method depends on experimental requirements and/or the analytical goal. For example, the gel electrophoresis (GE) and WB techniques are applied to obtain macroscopic information about a protein, such as molecular weight, potential for dimerisation and oligomer formation, isoforms, relative abundance and protein conformation. MS offers great advantages in protein identification and quantitation (both relative and absolute), although antibody techniques such as ELISA are also capable of absolute quantitation. NMR spectroscopy and X-ray crystallography are most commonly employed for protein structural analysis. In this section, the commonly used analytical techniques will be discussed with reference to sample preparation methods that are used to maximise the value of each approach. The approaches include gel electrophoresis (1D; 2D; and difference in-gel electrophoresis (DIGE)), antibody-based techniques for targeting specific proteins or classes (WB, ELISA, IHC and antibody columns), MS-based high-throughput proteomics (top-down, middle-down and bottom-up) and the common strategies for data analysis and structural proteomics (NMR and X-ray crystallography).

1.3.1 Gel Electrophoresis

Gel electrophoresis (GE) separates protein or peptide mixtures on the basis of electric field strength, the net charge on the molecule, size and shape of the molecule, ionic strength and properties of the gel matrix (e.g. viscosity, pore size) [47]. Polyacrylamide and agarose are two background matrices commonly used in electrophoresis, where matrices serve as porous media and behave like a molecular sieve. Generally, GE can be performed within one dimension (SDS-polyacrylamide gel electrophoresis (PAGE) and native PAGE) and two dimensions (2D-PAGE and DIGE). After proteomic separation by GE, the proteins can be visualised by using various staining methods such as Coomassie dye-, silver-, zinc-, fluorescent- and functional group-specific stains [48].

One-dimensional (1D) electrophoresis separates proteins by mass, commonly used in the lab for comparative analysis of multiple samples. SDS-PAGE can be run in two modes: denaturing and reducing the protein with a discontinuous buffer system that separates the proteins based on their mass and non-denaturing PAGE, also known as native PAGE, which separates proteins according to their mass/charge ratio. In the first method, protein samples are typically heated in the presence of SDS and dithiothreitol to denature the protein and the ionic detergent binds tightly to the protein to yield uniform negatively charged moieties. When a current is applied, all SDS-bound proteins within a sample migrate towards the positively charged electrode based on their mass. Proteins with known mass, known as

markers, are run as a reference alongside samples in order to obtain the measurement of the size of the experimental protein. In native PAGE, proteins are run in an alkaline buffer, and they are separated according to the net charge, size and shape of their native structure. Unlike denaturing SDS-PAGE, no denaturants are used in this method, and thus subunit interactions within a multimeric protein are generally retained and information can be gained about the quaternary structure. Although this technique is used to prepare purified and active proteins, proteins with the same molecular weight cannot be separated and it is unsuitable for molecular weight determination.

Two-dimensional PAGE (2D-PAGE) is used for separation of complex protein mixtures by the independent parameters of isoelectric point and molecular weight [49]. The first dimension separates proteins according to their native isoelectric point using isoelectric focusing in a pH gradient. The second dimension separates the proteins based on their mass obtained from SDS-PAGE. 2D-PAGE provides the highest resolution for protein analysis and is an important technique in proteomic research, where the resolution of thousands of proteins on a single gel is sometimes necessary.

DIGE is a technique used to measure differential expression of proteins between two samples and a 1:1 mixture of the two samples on a single gel. This method is semi-quantitative and has three steps: the samples (individual or mixed) are each tagged with a fluorescent dye, run on a 2D-PAGE gel. The fluorescent images generated from scans of each dye of the same gel are superimposed over each other. In this way, DIGE allows relative quantitation of proteins by comparing each test sample to the control (mixture). The method can be extended to run additional samples all compared back to a universal standard as the control. Once protein spots of interest are identified, duplicate gels are run on a preparative scale and the spots of interest are matched, excised and subjected to in-gel digestion and LC-MS analysis for protein identification.

1.3.2 Antibody-Based Techniques

Antibody-based techniques commonly used include WB, ELISA and IHC. These techniques are highly dependent on the availability and quality of the antibody reagents and appropriate protein standards. Typically recombinant proteins are used as positive controls and standards. The antibody reagents may be monoclonal or polyclonal antibodies. Furthermore, good secondary antibody and amplification reagents to improve the detection of the antigen-bound primary antibody are critical. Generally secondary antibodies are anti-species immunoglobulin antibodies, for example, a rabbit anti-mouse immunoglobulin antibody. The secondary antibodies are commonly conjugated to enzymes such as horseradish peroxidase or

alkaline phosphatase, which can be used with chromogenic, fluorogenic and chemiluminescent substrates. Biotinylated secondary antibody used in conjunction with enzyme labelled streptavidin can give excellent amplification of binding signal. All antibody-based techniques can suffer high backgrounds due to non-specific binding of any of the reagents. Frequently blocking buffers incorporating such reagents as non-fat dry milk, normal serum, BSA (bovine serum albumin), gelatin and one or more gentle surfactants are employed to reduce non-specific binding.

1.3.3 Western Blotting (WB)

WB is used to visualise a single protein or group of related proteins by using the ability of an enzyme- or fluorescence-conjugated primary antibody to bind to its specific antigen. Commonly labelled secondary antibodies are used to enhance the binding signal. The WB can be a rapid method to ascertain the presence of target proteins in complex protein mixtures and samples without investing much time in optimisation. Additionally it can inform if the target protein is well preserved or degraded in stored samples. The WB technique is inherently qualitative, but researchers often use this method semi-quantitatively to assess changes of apparent expression of the protein(s) of interest over time or as result of a perturbation. This method has three steps: gel electrophoresis (as described above) to separate the proteins according to size and/or charge; transfer to a membrane; and the antigen-antibody interaction and its detection with deposition of the chromogenic product at the site of antibody binding.

Samples for gel or WB analysis are prepared from a range of biological tissues and/or cells by the range of techniques described above (physical, chemical) with few limitations as the first stage of analysis (SDS-PAGE) is highly tolerant to chemicals and interfering components. Protease inhibitors are often added to the sample to prevent protein degradation during sample preparation. Prior to loading on the gel, protein concentrations of the samples are measured by one of many available assays (Bradford; Lowry; or bicinchoninic acid (BCA)) to ensure equivalent loadings of the amount of total protein, which is important if comparing samples. The denaturation of the sample proteins inherent in the SDS-PAGE methodology can efficiently expose linear epitopes making them available for binding to the primary antibody. Once the proteins are separated in the gel, the proteins are transferred to nitrocellulose or polyvinylidene fluoride (PVDF) membranes. Blocking reagents are used to minimise non-specific binding by the antibodies. The antibody binds to the antigen after a short incubation. In direct detection, an enzyme or biotin is linked directly to the primary antibody whereas indirect detection involves the use of a labelled secondary antibody.

1.3.4 ELISA

ELISA is a sensitive method used to detect the presence or quantify the abundance of an antigen. ELISAs are mostly conducted in 96- or 384-well microtiter plates. There are three main constructs of the ELISA: indirect, sandwich and competitive. For the indirect ELISA, the protein of interest, whether purified, enriched or in a crude preparation, is bound through adsorption, directly to the surface of the well, and subsequently detected by a combination of primary and labelled secondary antibodies. This simple approach can be a quick means of ascertaining the presence of a protein in a complex mixture or follow its enrichment through various separation protocols. The major limitation is low sensitivity; thus the target protein needs to be at relatively high abundance, for example, in the range of 0.2–5 μg/mL.

In the sandwich ELISA, a primary antibody is immobilised on the solid surface of a microtiter well. This antibody is often referred to as the 'capture' antibody. After a blocking step to eliminate any protein unadsorbed sites, the sample is added (usually diluted in detergent-containing buffers), and the target protein is allowed to bind to the primary antibody. After washing the unbound material, a second primary antibody is added and allowed to bind the bound antigen. This second primary antibody, which can be referred to as the 'detection' antibody, may be labelled and thus detected directly. Alternatively, the 'detection' antibody may be detected with labelled secondary antibodies if the two 'capture' and 'detection' antibodies are derived from different species (e.g. mouse and rabbit antibodies). Proteins of very low or unknown concentration in the sample can be detected because the 'capture' antibody essentially concentrates the low-abundance analyte. With the unbound material washed away, the binding of 'detection' antibody and subsequent amplification reagents are not hindered by interfering sample constituents.

Competitive ELISA can take many formats. The simplest is whereby the competitor antigen or protein is initially bound to the well, and after the blocking step the primary antibody is incubated with the sample with unknown amounts of the analyte or target protein. The well-bound antigen and the in-solution antigen derived from the sample compete for the primary antibody. Unbound material is washed away, and a labelled secondary antibody is added to the wells, which recognises the primary antibody in this case, captured by the well-adsorbed antigen. Unlike the other ELISAs and WB, decreased signal indicates high relative abundance of the target protein in the sample.

If high-quality preparations of the target protein are available to generate standard curves, both the sandwich and competitive ELISA formats can provide excellent quantitation.

1.3.5 Immunohistochemistry

IHC detects molecules (proteins) of interest within tissues using antibodies. Probes or enzymes conjugated to primary antibodies that bind the target molecule (or secondary antibodies that recognise the primary antibody) report the location of

targets in tissue or cell samples for imaging. This method can help to visualise, identify and quantify the high-resolution distribution and localisation of specific proteins within cells and within their proper histological context [50]. Generally, this technique has two steps: firstly, sample preparation which broadly involves tissue collection and perfusion, fixation, embedding, sectioning and mounting, de-paraffinisation and epitope (antigen) retrieval, quenching/blocking endogenous target activity, blocking non-specific sites and sample labelling and, secondly, visualisation, interpretation and quantitation of the resultant protein expression [51].

1.3.6 Mass Spectrometry (MS)-Based Proteomics

MS-based proteomics represents a powerful tool for systems biology. It can provide a systematic, global, unbiased and quantitative assessment of proteins, including interactions, modifications and location, and can inform function [1, 6, 9]. Significant improvements in the sensitivity and the resolution of MS instrumentation and the associated advances in sample preparation and data analysis have dramatically advanced our understanding of the complex and dynamic nature of proteome. For example, MS-based proteomics can provide the quantitative measurement of the yeast proteome in 1 hour (of analytical time) [52]. New label-free approaches avoid some of the pitfalls and costs required to construct and analyse large quantities of tagged proteins or the generation of protein-specific antibodies [2]. Mass spectrometry can provide high-throughput analysis, offers high sensitivity and is capable of identifying and quantifying proteins spanning at least five orders of magnitude (dynamic range).

Based on the method of protein/proteome characterisation, MS-based proteomics approach can be divided into two main analytical streams: top-down and bottom-up approaches. Both methods have advantages and disadvantages for the identification of proteins and protein interacting partners, protein quantification and analysis of PTMs [53, 54]. Through top-down approach, intact proteins can be studied, which greatly simplifies sample preparation and reduces the mixture complexity as no proteolytic digestion is required [55]. Sample preparation for top-down proteomics also follows the classical proteomic sample preparation steps: protein isolation, extraction, denaturing, enrichment, gel electrophoresis and MS analysis. After extraction, chemical labelling can be included with soluble proteins for relative quantitation of samples by fluorescence (e.g. DIGE) or MS (e.g. isobaric tags for relative and absolute quantitation (iTRAQ) or tandem mass tags (TMT)) [56].

Top-down MS provides comprehensive information for the whole protein by detecting the range of PTMs (e.g. glycosylation, phosphorylation, methylation, acetylation) and sequence variants (e.g. mutations, polymorphisms, alternatively spliced isoforms) concurrently in one spectrum without a priori knowledge [57]. Although top-down MS can identify the intact proteins and their combination of modification events, this method is experimentally challenging for functional proteomics experiments due to the greater difficulty in analysing proteins in

comparison with peptides. Proteins may be distributed across a number of isoforms [1]. Top-down methods have greater sample requirements, both in terms of protein quantity and purity. As a result, better separation methods are applied to reduce the sample complexity, obtain pure samples and increase the dynamic range of detection in top-down MS. Several offline and independent separation methods such as capillary HPLC, hydrophilic interaction liquid chromatography (HILIC) or combination of on- or offline HPLC with capillary isoelectric focusing, chromatofocusing or ion exchange can be used to separate and purify intact proteins for top-down MS analysis [58]. Furthermore, gel-eluted liquid fraction entrapment electrophoresis [59] and multicompartmental electrolyser called membrane-separated wells for isoelectric focusing and trapping [59] can also be included in the sample preparation to separate proteins for top-down-based MS analysis.

Independent of whether the target is protein or peptide, the analyte must be ionised and desorbed into the gas phase before detection and fragmentation in MS-based proteomics. Two types of ionisation techniques are used to generate the *m/z* in the gas phase: matrix-assisted laser desorption ionisation (MALDI) and electrospray ionisation (ESI). In MALDI samples are mixed with a matrix, often an organic acid, deposited on a target plate and ablated by a pulsed laser. The matrix absorbs energy transferring it to the analyte, leading to the ionisation of the analyte and its transfer into the gas phase. Generally, sample preparation for MALDI is straightforward and less complex than other MS-based methods, but there are still some points to be considered for successful sample preparation. Although MALDI is more tolerant to detergents and/or salts such as trifluoroacetic acid (TFA) or ammonium acetate which is often added to the LC for better separation, the presence of these chemicals can still interfere with crystallisation and/or ionisation. Additionally, the presence of higher amount of lipids and glycerol can interfere with the crystallisation process. The matrix selection can be empirically determined for each compound through a literature search or prior knowledge. Samples extracted for MALDI can be stored lyophilised or in solvent for future use. MALDI is most commonly coupled with time-of-flight (TOF) mass analysis. MALDI-TOF-MS has found great application downstream of one- or two-dimensional gel electrophoresis and more recently as mass spectrometry imaging (MSI) enabling detection and localisation of proteins in tissue samples [60]. This technique can also be a useful tool to explore molecular distribution and co-localisation of proteins in tissues or cells [61].

ESI is another soft ionisation technique used in MS for producing charged ions from proteins, peptides and other non-volatile macromolecules. In LC-MS-based proteomics, ESI provides high voltage to impart charge to the solvated analyte before exiting the electrode. The charged ions are desolvated and thus transferred to the gas phase before undergoing transmission and mass analysis. Once the molecule is in the gas phase, the *m/z* ratio of molecules is determined by their trajectories in a static or dynamic electric field. Taking advantages of high resolution and mass accuracy, top-down proteomics often uses ESI coupled to either Fourier transform ion cyclotron resonance (FT-ICR) [62] or Orbitrap mass analysers [63].

The bottom-up approach, also known as shotgun proteomics, relies upon proteolytic digestion of proteins into peptides with subsequent fragmentation by tandem