

Chapter 1

Introduction to Proteomics: a Brief Historical Perspective on Contemporary Approaches

Jay J. Thelen

Abstract The field of proteomics has experienced numerous milestones over the course of the past 35–40 years. As an introductory chapter to this larger review text on plant proteomics, this article provides a cursory historical perspective on protein separation and identification techniques widely used in plant biochemistry laboratories today. In the past 10 years alone, advancements in techniques such as two-dimensional gel electrophoresis, mass spectrometry, and mass spectral data mining have made previously intractable proteomics problems almost routine by today's standards. In analyzing these various proteomics approaches I also discuss and project their utility for the next generation of proteomics research.

1.1 Introduction

Proteomics, or the high-throughput identification and analysis of proteins, is an emerging field of research facilitated by numerous advancements over the past 35–40 years in protein separation, mass spectrometry, genome sequencing/annotation, and protein search algorithms. Recognizing this trend in the physical and life sciences, the term “proteome” was first used by Wilkins et al. (1995) to describe the protein complement to the genome. Since the first use of this term its meaning and scope have narrowed. The host of post-translational modifications, alternative splice products, and proteins intractable to conventional separation techniques has each presented a challenge towards the achievement of the classic definition of the word (Chapman 2000; Westermeier and Naven 2002; Wilkins and Gooley 1998). The broad dynamic range of protein expression has also contributed to difficulties in efforts towards identifying every protein expressed in the life cycle of any given organism (Corthals et al. 2000). For example, identification of every protein expressed in plant leaves would never reveal proteins that are specifically expressed

Jay J. Thelen: University of Missouri-Columbia, Department of Biochemistry, Columbia, MO 65211, E-mail: thelenj@missouri.edu

in roots. Despite these limitations, hundreds if not thousands of proteins can be resolved, profiled and identified using the latest methods – a remarkable achievement given the recent genesis of this discipline.

The purpose of this chapter is to briefly introduce and provide a historical perspective on established proteomics concepts and methods that are being used in many plant biology laboratories today to comparatively profile protein expression and identify proteins. I will also attempt to provide a perspective on the future outlook of each of these approaches. This introduction will hopefully be useful for non-experts in the field of proteomics as an aid to comprehension of most of the terminology and jargon used in this highly technical field of life sciences research. The varied approaches to proteomics research can be generally classified as having one of two major objectives: (1) protein or peptide separation, and (2) identification and characterization of resolved proteins or peptides, typically by mass spectrometry. I will address these two aspects of proteomics research in the first two sections in this introductory chapter and then discuss general strategies for quantitative protein profiling.

1.2 Protein Separation and Detection for Proteome Investigations

Currently, there are three preferred methods for separation of complex protein or peptide samples: (1) denaturing polyacrylamide gel electrophoresis (PAGE) also referred to as sodium dodecyl sulfate polyacrylamide (SDS-PAGE); (2) two-dimensional (2-D) gel electrophoresis; and (3) liquid chromatography (LC) a general term that includes all forms of ion exchange, affinity, and reversed-phase chromatography (Hunter et al. 2002). There are of course other forms of protein separation, including preparative isoelectric focusing (protein separation according to native charge) and native or blue-native PAGE, to name but a few alternative techniques. Due to space constraints however, only SDS-PAGE and 2-D gel electrophoresis will be discussed here.

1.2.1 Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis

No protein separation technique is more widely used than SDS-PAGE, first reported by Laemmli in 1970. It would not be an exaggeration to state that nearly all contemporary laboratories performing life sciences research employ this technique. The widespread use of SDS-PAGE to separate proteins according to size can be attributed to its ease, reproducibility, and modest consumable and instrument expenses. Although an easy technique to perform, the resolving power of SDS-PAGE is somewhat limited. Mass spectrometry (MS) analysis of any single discreet SDS-PAGE protein band from a complex protein sample consistently reveals multiple proteins, frequently greater than ten (Phinney and Thelen 2005). However, for highly enriched

samples of low complexity (<10 unique proteins) SDS-PAGE may be suitable. In general, accurate quantitative analysis of SDS-PAGE protein bands from a complex sample is not feasible as the volume of any band is the collective composition of each unique protein in that band. However, as a pre-fractionation technique for alternative quantification strategies including chemical labeling (using stable isotope conjugates; Ramus et al. 2006) and perhaps label-free quantification using recently developed software tools (SIEVE, DeCyder MS), SDS-PAGE may find a new niche as a rapid, reproducible separation technique prior to MS quantification.

1.2.2 Two-Dimensional Gel Electrophoresis

Around the time at which SDS-PAGE was introduced, O'Farrell applied isoelectric focusing (IEF) to protein samples prior to SDS-PAGE to pioneer the concept of two-dimensional (2-D) gel electrophoresis (O'Farrell 1975). Although extremely powerful in its resolving capabilities, this method suffered from reproducibility issues owing to the casting, focusing, and extrusion of the fragile tube gels used for IEF. Over the years this procedure has been improved through the introduction in 1978 (Görg et al. 1978) and recent commercialization (Görg et al. 2000) of the immobilized pH gradient (IPG) strip, to replace IEF tube gels, which has resulted in a major resurgence in this technique.

Reproducibility, sample loading and resolution for 2-D gel electrophoresis have significantly improved with the introduction of the IPG strip in conjunction with commercial Peltier-cooled programmable focusing units for IEF (Görg et al. 2000). These advancements have made 2-D electrophoresis an attractive method for the separation of complex protein samples. Besides the impressive separation capabilities, another reason 2-DE is frequently preferred to LC-based approaches for protein separation is that a reproducible 2-DE proteome reference map is a static, visual entity. A fully annotated 2-DE reference map for a specific organ, tissue, cell, or organelle of interest is a valuable tool that can save time and money when 'landmarking' differentially expressed proteins in response to a treatment, mutation, or transgene introduction. Although 2-D electrophoresis suffers from well-publicized limitations, such as under-representation of membrane proteins (Wilkins et al. 1998; Santoni et al. 2000), this time-honored method is presently one of the preferred approaches for quantitative characterization of complex protein samples. The popularity of 2-DE will no doubt continue with recent technical developments such as sensitive and quantitative pre- and post-electrophoretic stains for total proteins, as well as phospho- and glycoproteins, as discussed herein.

1.2.3 Extracting Proteins From Plant Samples

Performing 2-D electrophoresis with plant samples can be a challenging endeavor, in part due to the high carbohydrate:protein ratio in most plant tissues. Direct grinding of samples in IEF extraction media, while generally sufficient for non-plant

cells, results in extensive streaking in the IEF dimension with plant tissues. Therefore, obtaining high quality 2D gels from plant tissues requires the removal of carbohydrates prior to IEF. Mooney et al. (2004) found that phenol partitioning of proteins followed by ammonium acetate/methanol precipitation was one consistent strategy with which to isolate proteins free of complex polysaccharides in mature soybean seeds. This approach was first employed for 2-DE by Hurkman and Tanaka (1986) for plant membrane proteins. This protein isolation procedure has also been successfully employed with developing seeds from soybean, castor, rapeseed, and Arabidopsis as well as purified oil bodies from rapeseed (Hajduch et al. 2005, 2006; Agrawal and Thelen 2006; Katavic et al. 2006). A comparison of phenol/methanol precipitation versus TCA/acetone extraction in grape berries revealed greater protein yield and spot resolution using the phenol extraction procedure (Vincent et al. 2006). Another investigation compared phenol/methanol extraction with two variations of TCA/acetone extraction in a range of tomato tissues and the fruits of banana, avocado, and orange (Saravanan and Rose 2004). This latter investigation reported that phenol extraction gave higher protein yields and greater spot resolution and intensity, particularly from fruits rich in polysaccharides. However, it was noted that the spot patterns were different amongst these extraction procedures, in part due to enhanced glycoprotein extraction with the phenol procedure. Despite the nearly universal success of the phenol extraction procedure for recalcitrant protein samples, other less time-consuming procedures have also proven successful with plant protein samples (discussed further in Chap. 2 by Hurkman and Tanaka).

1.2.4 In-Gel Detection of Proteins

Protein quantitation and detection is an area of proteomics research that has changed dramatically over the past 10 years. Although several methods for protein detection have been reported (for a comprehensive treatise, see Allen and Budowle 1999), silver and Coomassie Brilliant Blue (CBB) staining methods have historically been the preferred methods for in-gel protein detection. The two chemical forms of CBB, G-250 and R-250, differ in their sensitivity, quantitative linear range, and destaining properties. Since CBB G-250, also referred to as colloidal CBB, outperforms the R-250 variant on all counts it is generally recommended for proteomics applications. Silver staining is at least 10-fold more sensitive than colloidal CBB, with a reported detection range of 0.1–1 ng (Ochs et al. 1981; Shevchenko et al. 1996). However, silver staining is plagued by problems such as inferior reproducibility, poor linear dynamic range, and non-quantitative negative-staining of some modified proteins (Westermeyer and Naven 2002; Wilkins and Gooley 1998; Görg et al. 2000), all of which complicate downstream quantitation and spot matching. Broad dynamic range fluorescent protein stains including SyproRuby™, Deep Purple™, and ruthenium II, which have detection

sensitivities around 10–20 ng (Rabilloud et al. 2000, 2001; Steinberg et al. 2000; Chevalier et al. 2004) are promising, but expensive, alternatives to Coomassie and silver staining as general protein stains. Other, more specific fluorescent stains include Pro-Q Diamond and Pro-Q Emerald (both manufactured by Invitrogen), which are specific for phosphoproteins and glycoproteins, respectively (Steinberg et al. 2000, 2001). As with any of these specialized commercial stains, expense is a concern when implementing their use. Modified protocols for both Sypro Ruby and Pro-Q Diamond have demonstrated that multi-fold dilutions of these commercial stains are possible without compromising linear dynamic range or sensitivity (Krieg et al. 2003; Agrawal and Thelen 2005). With any of these fluorescent stains, manual spot picking from gels can be a challenging task. Generally, “over-staining” of these gels with Coomassie or silver is required to accurately determine spot location (Agrawal and Thelen 2006). For further discussion on staining, see Chap. 2 by Hurkman and Tanaka.

1.3 Protein Identification using Mass Spectrometry

Analysis of any analyte by MS requires ionization of that molecule and its entrance into the gas phase. For large macromolecules such as proteins and peptides this was long considered a Herculean task, analogous to making elephants fly (Fenn 2003). The development and commercialization of two different “soft” ionization approaches, electrospray ionization (ESI) and matrix assisted laser desorption ionization (MALDI) enabled large macromolecules such as proteins to be analyzed either in flowing, liquid solution or in a dry, crystalline state, respectively (Fenn et al. 1989; Tanaka et al. 1988; Karas and Hillenkamp 1988). The importance of these developments was appreciated by the scientific community and in 2002 led to John Fenn and Koichi Tanaka sharing the Nobel prize in Chemistry. In ESI, high voltage is applied to a flowing solution containing the analyte as it passes through a narrow bore needle. As the solution of charged molecules evaporates, the emitted droplets shrink into smaller droplets and shortly thereafter enter the gas phase. In MALDI, a laser is fired at a stainless steel plate under vacuum. Dried on that plate is the analyte, co-crystallized with “matrix” – a small organic molecule that absorbs the pulsed UV laser light. Although the mechanism is not completely understood, the laser light pulsed on the co-crystallized matrix-analyte layer results in the vaporization of matrix and the associated analyte. Early work with MALDI necessitated high vacuum conditions, which were easily configured with time of flight (TOF) mass analyzers; however, the emergence of atmospheric pressure MALDI has allowed this ionization approach to be configured with many other types of mass analyzers. For an in-depth discussion of mass analyzers a recent review is available (see Domon and Aebersold 2005).

Plant Proteomics

Samaj, J.; Thelen, J.J. (Eds.)

2007, XXIX, 369 p., Hardcover

ISBN: 978-3-540-72616-6