ON THE ANALYSIS OF INTACT PROTEIN ISOFORMS BY MASS SPECTROMETRY
Jeremiah D. Tipton, John C. Tran, Adam D. Catherman, Dorothy R. Ahlf, Kenneth R. Durbin, and Neil L. Kelleher*
Department of Chemistry, Department of Molecular Biosciences, and The Chemistry of Life Processes Institute, Northwestern University, Evanston, IL 60208.
Running Title: Protein Forms and Top Down Analysis with Mass spectrometry
*Corresponding author E-mail: n-kelleher@northwestern.edu

SUMMARY
An organism’s diverse proteome arises from such events as single nucleotide substitutions at the DNA level, different RNA processing, and dynamic enzymatic posttranslational modifications. This review focuses on the measurement of intact proteins to describe the diversity found in proteomes. The field of biological mass spectrometry has steadily advanced, enabling improvements in the characterization of single proteins to proteins derived from cells or tissues. In this mini review, we discuss the basic technology for “Top Down” intact protein analysis. Further, examples of studies involved with the qualitative and quantitative analysis of full length polypeptides are provided.

INTRODUCTION TO INTACT PROTEIN MASS SPECTROMETRY

Goals of modern biological and biomedical research include characterizing, and ultimately, treating human disease. Different diseases may be characterized by DNA, RNA, and proteins in regards to observed malfunctions and phenotypes at the molecular level. The field of functional genomics has progressed to ‘rapid’ whole genome sequencing. Large scale mapping of genetic information catalogues mutations and polymorphisms that can translate into proteome variation. RNA processing increases protein variation through basic transcription or alternative splicing (Fig. 1). To determine the significance of shotgun proteomic type experiments, the identified proteins may often be sorted into “interaction networks” based on known interactions or changes in regulation upon different disease types or perturbations (Fig. 1). The analysis of an organism’s proteome presents a difficult challenge due to the complexity and often the distribution of gene products into different protein forms. In contrast to genomics and RNA analysis, comprehensive mapping of an organism’s proteome and its dynamic variation is further complicated by the enzymatic (or chemical) addition/deletion of post-translational modifications (PTMs).

The control and regulation of protein expression stems largely from the control of RNA expression from coded DNA, and such, the control of different isoforms. Isoforms, as has been recommended by IUPAC, refers to protein forms that have high sequence identity and arise from the same gene family or polymorphisms. Thus, the variation for protein isoforms arises from ‘genetic’ sources by this definition. The term “protein species” is suggested to refer to highly related protein forms that differ due to PTMs and alternative splicing (1, 2) (Fig. 1). Colloquially, most use the term ‘isoform’ or ‘variant’ to refer to a mixture of all these sources of molecular variation at the protein level. Here, we focus on modern protein analysis, and more specifically, the identification and characterization of whole protein molecules by mass spectrometry (MS). This has come to be known as “Top Down” (TD) intact protein analysis (3-5).

Proteins may also be routinely analyzed by “Bottom Up” proteomics (BU), where the proteins are digested with a protease prior to peptide detection and protein identification by peptide sequencing with tandem MS (Sup. Fig. 1). The digested samples produce a complex mixture of peptides between ~500 and 3500 Da that are usually separated by single mode or multidimensional chromatography. Online precursor mass measurement, along with the “fragmentation” spectrum by tandem mass spectrometry (MS/MS), typically allow for the inference of peptides by mapping to protein databases and probability of such. The peptides are matched to proteins along with the probability of the protein being a true identification. Without the aid of additional
information, often only the gene family can be identified by BU proteomics.

Further, several isoforms may be associated with a gene family and the individual isoforms may produce peptides with identical sequences (6). Although isoforms, sequence, and PTM information may be lost in the context of describing the full proteome, BU proteomics affords higher proteome coverage in comparison to TD proteomics (7). One way to convey this, in biological terms, is in the challenge of analyzing isoforms created by evolution in the process of microbial speciation. Using simple mass analysis of whole ribosomal proteins, the sequence variation imbedded in the whole protein’s mass can be used to create a new type of phylogenetic tree (8).

As mentioned above, TD has yet to reach the same proteome coverage as BU proteomics. This is further exacerbated by non-routine TD analysis of proteins above 50 kDa in a high-throughput discovery mode. However, intact protein analysis is more likely to result in measurement of variation from coding polymorphisms, alternative splicing, and diverse PTM changes, which has been demonstrated in several targeted studies on single proteins or those present in modest mixtures (9-12).

If we loosen our definition of TD to include the classical approach of 2-D gel electrophoresis (2D-GE), one can see that this strategy of molecular analysis has been used for quite some time (13). However, to identify proteins with 2D-GE, interesting “spots” are analyzed by BU proteomics. Thus, this mini-review will include some information on hybrid techniques, such as 2D-GE with BU proteomics for protein identification (14). However, much of the review in the technology section will describe separations that are “solution-based” platforms which maintain proteins in solution. This strategy preserves the proteins in their intact, mature form for mass measurement and direct sequencing in the mass spectrometer (5).

Covering the continuum of sample complexity - from single protein to protein complexes to whole cells and tissues - helps to define the workflows and platforms required for analysis. For “targeted” TD intact protein analysis, “high-sequence coverage” of single proteins or protein complexes can be obtained; whereas high proteome coverage can be obtained for global analysis of cellular and tissue lysate. Global proteome analysis is often associated with high-throughput technologies because several different components may be identified over a relatively short amount of time. “Targeted” TD analysis is not intended to mean low-throughput. In regards to global TD analysis, increasing the number of identified isoforms per experiment is desirable to describe multiple isoforms at the same time, affording the ability to perform experiments in discovery mode. Thus, high-throughput TD includes the analysis of proteomes from complex samples; whereas, targeted TD analysis is performed on single or simple mixtures of proteins (Fig. 2).

As the duty cycle for instruments has improved over the last 10 years, high-throughput TD proteomics has continued to become increasingly possible for discovery and quantitation experiments in a fashion that is starting to approach BU proteomics. The next section will expand upon the different separations, instrumentation, and informatics that have been developed to achieve TD in targeted or high-throughput contexts.

TECHNOLOGY FOR INTACT PROTEINS: IONIZATION, SEPARATIONS, INSTRUMENTATION, AND INFORMATICS

Many of the breakthroughs in biological MS - i.e. intact protein analysis - were achieved due to the advent of soft ionization techniques such as electrospray ionization (ESI) (15) and matrix assisted laser desorption ionization (MALDI) (16, 17). A basic description of these ionization techniques can be found in the supplemental section. Briefly, ESI allows for samples to be directly infused or analyzed online by reversed-phased liquid chromatography (RP LC). This unique capability allows for proteins to be maintained in the “solution phase” for MS analysis. Proteins may also be separated by different modes such as those based on charge, strong cation exchange or weak anion exchange, or those based on normal phase separations with hydrophilic interaction chromatography. Other reviews have outlined successful applications of different separation
modes (18, 19). Further, non-particle base preparative (solution) isoelectric focusing (20) and gel-eluted liquid fraction entrapment electrophoresis (21) have recently provided modes of separation similar to acrylamide gels, but result in recovery of proteins in the solution phase with high yields. In contrast to RP LC, each of the above techniques is usually performed off-line to MS because of incompatibility with ESI ionization. Thus, samples are separated with an orthogonal method such as RP LC for direct infusion ESI MS or by on-line RP LC ESI MS (22-25). Especially for complex protein samples, improved separations will always improve the MS analysis (19).

As mentioned above, intact proteins may also be separated by slab gel electrophoresis, although not considered TD analysis in the traditional sense. Slab gels are the cheapest and most common way to separate proteins by size and/or isoelectric point. Proteins are detected by staining dyes, covalently attached fluorescent dyes, radiolabeling, immunoblots, (26, 27) or identified by BU mass spectrometry. One of the main strengths of slab-gel intact protein separations is quantitation with 2-D difference gel electrophoresis (DIGE) (28). The drawback for this detection scheme is that PTM site location, full sequence, and/or full isoform information may be lost. Solution phase based TD proteomics also provides quantitation platforms with chemical and metabolic labeling schemes. Akin to fluorescent labeling for DIGE, analysis of yeast with $^{14}$N/$^{15}$N metabolic labeling may provide relative quantitation across different treatments (29). More recently, Muddiman’s group performed stable isotope labeling of amino acids (SILAC) with Aspergillus flavus and human embryonic stem cells to yield relative quantitation of whole proteins (30).

Traditionally, TD intact protein analysis has been most successful with ESI Fourier transform ion cyclotron resonance (FTICR) MS. FTICR MS is still currently the highest resolution instrument available for sample analysis (31). The analytical advantages of FTICR MS for large biomolecules have been reviewed elsewhere (9, 19, 31). FTICR MS can provide mass error measurement of less than 2 ppm for proteins up to 25 kDa. This mass accuracy is often required for unambiguous characterization, especially when the mass shift may be associated with deamidation ($\Delta = 1$ Da on a $\sim$15 kDa protein) (32), the difference between acetylation and tri-methylation ($\Delta = 0.0363$ Da on fragments) (33), or reduced/oxidized disulfide bonds ($\Delta = 2$ Da on a $\sim$14 kDa protein) (34). Further, when analyzing fragment ion spectra (i.e. dissociation of the intact protein into fragments for mass measurement), high mass accuracy is desirable for high confidence, and high resolving power is advantageous for resolving overlapping fragment peaks (35). FTICR MS has set the standard for the requirements for accurate and precise mass measurement. Recently, a second high resolution instrument, the LTQ-Orbitrap, akin to FTICR technology but lacking a superconducting magnet, has been applied to the analysis of intact proteins (25, 36 -38).

Although not typically associated with high-resolution intact protein analysis, Time-of-flight (TOF) mass spectrometers have also been applied to characterize proteins. For example, MALDI-TOF analysis for protein mapping has been developed for such applications as bio-defense (39). Thus, intact protein mass spectra may be compared across sample types by MALDI-TOF. Another application includes tissue imaging, by which MALDI-TOF can provide highly sensitive and specific spatial analysis of biomolecules based on mass (40 – 42). The experiments add another dimension to the investigation of protein arrangements in all types of biological tissues, such as diagnosis from resections (43).

Each of the different mass analyzers described above may fragment proteins through different methods based on activation using thermal heating (collisions with gas) or electron capture/transfer. Fragment ions report on the protein primary sequence and PTM site location. One of the first successful implementations of protein fragmentation includes electron capture dissociation (ECD) and FTICR MS on ubiquitin and cytochrome c (44). ECD fragments proteins by ultra-fast dissociation of the amino acid backbone, thus generating several product ions along the backbone of the protein. Often, ECD
produces the most sequence information for characterization of intact proteins. Further, ECD is usually associated with single or simple mixtures of proteins.

Not mentioned in the above mass analyzer section are linear quadrupole ion trap (LTQ) or quadrupole (Q) mass analyzers. LTQs (45) and quadruples are not normally associated with intact mass measurement as stand-alone mass analyzers; however, the development of hybrid instruments (two mass analyzers placed in-line), i.e. LTQ-FTICR, LTQ-Orbitraps, Q-TOF’s, TOF/TOF (46) have improved throughput by thermal activation with collisional induced dissociation (CID). Ion trap hybrid instruments fragment proteins in the LTQ or Q prior to transfer to a second higher resolution mass analyzer. Electron transfer dissociation (ETD) (47) is another fragmentation method that is akin to ECD but takes place in a LTQ as opposed to ICR cells. Also, higher energy collisional activated dissociation (HCD) has been implemented on LTQ-Orbitraps (48).

In comparison to ECD, thermal activation produces fewer fragments with increased amino acid specificity. An early implementation of thermal activation include CID fragmentation by Loo and coworkers in 1990 (49) Other early forms of protein fragmentation methods include nozzle skimmer, infrared multiphoton (50, 51) and surface off-resonance irradiation dissociation (52). During the last 10 years, all of the described fragmentation methods have been improved. For example, funnel skimmer dissociation, akin to nozzle skimmer dissociation (53, 54) and CID (55) continue to be implemented across instrument types; however, fragmentation is still limited to specific amino acids and often only on the proximity of the N- or C-terminus of proteins.

Hybrid instruments have also resulted in gains to ‘scan speed’ for intact protein analysis. For example, a linear octopole was placed in-line with a FTICR to detect 101 whole proteins from 5 to 59 kDa, with many PTMs and protein forms identified from the methanogen, Methanosarcina acetivorans (56). Another implementation included CAD with a LTQ-FTICR mass spectrometer. This system was similar to the linear octopole; however, the number of ions transferred into the ICR (or Orbitrap) can be controlled (57, 58). Another advantage of the improved duty cycle of the new generation of hybrid instruments becomes particularly important as the resolution of intact protein separation has improved with reversed phase nano-liquid chromatography (RP nLC i.e., flow rates at 300 – 500 nL/min with capillary columns). Currently, high throughput mass spectrometric analysis with on-line nLC can provide routine analysis of proteins up to 35 kDa (22, 23). LTQ-Orbitraps have yet to reach the same intact mass plateaus as ICR (in terms of resolving power), however they are very promising for TD proteomics analysis (59).

On-line fragmentation of intact proteins can also be achieved with TOF instrumentation (60). Resemann et al. reported the ability to unambiguously sequence a 13 kDa fragment of a variable region of an antibody through the use of MALDI in-source decay in a hybrid TOF/TOF instrument (61). Hybrid TOF/TOF instruments (two TOF mass analyzers) have also been found to be useful for some proteins between 5 and 15 kDa (62). A second type of TOF hybrid instrument includes Q-TOF research performed by McLuckey’s group (63). They were able to analyze fragmentation spectra with relatively lower resolving power through reducing the complexity of fragmentation spectra by collapsing the charge states of fragments to mainly singly charged species though the use of fast ion-ion reactions (64). All of the described instrumentation has generated new types of data to be extracted and analyzed. The complexity of the data has resulted in new programs for sequencing proteins.

Several advances in protein identification and instrumentation control have also improved the throughput of TD intact protein analysis. Intact and fragment masses are characterized by mapping calculated sequence tags or fragment ions to databases with known protein sequences or by heuristic analysis. ProSight was the first program to provide algorithms for single and high-throughput protein analysis (65). More recently, other algorithms such as big mascot (53) and PIITA (66) have been developed to provide general software packages for intact protein analysis. TOF/TOF hybrid instruments have also found
development of software for interpretation of small proteins in regards to food borne pathogens (62). In regards to instrument control software, fraction collection followed by direct infusion MS with a robot provided automated control of all steps from sample addition to data analysis. In short, fractions were collected off-line with a split-flow directed towards the instrument for intact mass profiling. The information from the chromatogram generated a feedback loop for the robot to infuse the sample with intelligent mass spectrometric data acquisition that includes mass isolation and fragmentation. Fragment ions can then be “signal averaged” for increased sensitivity and mass accuracy with direct infusion ESI FTICR (67, 68).

Other deviations of this experiment include Lubman et al. who achieved high separation resolution by coupling solution isoelectric focusing and non-porous RPLC to MALDI-TOF (69). This pseudo 3D separation afforded mass resolution (from the TOF analyzer) that surpassed that of SDS-PAGE. Further the group also provided identifications through off-line digestion of the fractions collected from the front end separation. In regards to other forms of automation, Paša-Tolić and coworkers have combined intact protein profiling by on-line nLC FTICR MS and split-flow fraction collection. The integrated workflow was also implemented to identify phosphorylation sites with BU proteomics as a second dimension to TD (70).

APPLICATIONS OF INTACT PROTEIN MASS SPECTROMETRY

High-throughput characterization of isoforms and PTMs may reveal incorrect database predictions. For example Ferguson et al. detected 99 proteins from Methanosarcina acetivorans, of which 15 were shown to have translational start sites that were mispredicted. Also, the study identified 5 unannotated proteins and another set of proteins from incorrect reading frames (71). In another high-throughput study on human primary leukocytes harvested from leukoreduction, Roth et al. was able to report on “the proteotype” of the diverse human proteome by high-throughput characterization of over 133 proteins (greater than 600 unique intact masses), 32 of which had PTMs, cSNPs, or were detected as proteolysis products (72). Other studies continue to improve lower limits for characterization. For example, with only 500 ng of material, Lourette et al. were able to profile and semi-quantitate over 250 oxidized and nitrated calmodulin forms in activated macrophages through nanoLC-MS (73). There are other fine examples of high-throughput characterization of intact proteins and isoforms; however, we now turn to more targeted studies.

Fraction collection from liquid chromatography prior to direct infusion may provide highly purified proteins for “targeted” characterization. For example, Ryan et al. characterized salivary proteins from the cystatin family, identifying signal peptide cleavages, disulfide bond, and phosphorylation locations after offline pre-fractionation. Further, three SNP’s on two members were detected for this protein family (74). In another study, over 150 possible sites for phosphorylation of cardiac myosin binding protein C (cMyBP-C) were identified by Ge et al. on the full length and truncated isoforms. By achieving unit mass resolution and high mass accuracy for the 115 kDa full form of cMyBP-C, they were able to identify multiple phosphoprotein forms (mono, di, etc.). Further, sites for phosphorylation were localized to Ser-283, Ser-292, and Ser-312 after fragmentation analysis. In comparison to the wild-type forms, recombinant truncated forms of cMyBP-C studied had dramatically altered phosphorylation profiles (75). In another series of studies, Ayaz-Gunner et al., confidently identified bisphosphorylation of Ser22 and Ser23 from mouse cardiac troponin (cTnI) in a wild type, rather than a transgenic mouse. Certain sites were shown to be phosphorylated in vitro that were not observed in vivo (Fig. 3, top left) (76). For a last example, Zhang et al., reported that swine cTnI had the same Ser22/Ser23 phosphorylation and N-terminal acetylation, but also revealed the localization of a novel V116A SNP genetic variant (77). Many TD studies have provided results on highly purified proteins or families of proteins. For example, therapeutic antibodies represent a class of single proteins which have a grown in
relevance to biologics and the biotechnology industry.

TD intact protein analysis, in regards to biologics, may be applied with other techniques to characterize the expressed antibody. Zhang et al. and Bondarenko et al. characterized monoclonal immunoglobulin gamma G with direct infusion and on-line reversed phase chromatography with an LTQ-Orbitrap. The result was the mass analysis of large antibodies, along with the detection of different glycoforms (Fig. 3, top right) (48, 78, 79). Similar experiments have also been performed on TOF instruments, however with less sequence information and at lower resolution (80). There are many other publications which tackle the challenging problems of therapeutics. A recent review reports on the analysis of biologics and antibodies by MS (81). TD intact protein analysis may aid in the early decision process to streamline costs associated with drug development, especially in regards to therapeutics and antibodies. Further, there are many other therapeutic proteins that must be characterized before FDA approval. In considering drug discovery, another class of proteins which may be characterized by TD intact protein analysis includes histones, their isoforms, and their dynamic posttranslational modifications. Histone MS, in cohort with other techniques, has informed the selection of different methyl- and acetyl- transferases as ‘drugable’ targets.

Histones - the protein building block of chromatin - are a class of complex protein isoforms (H1, H2A, H2B, H3, H4 plus all of the other possible forms of such) with dynamic PTM changes depending on the expression state of a cell. All histone families have been successfully characterized by TD to reveal protein forms and dynamic PTMs (82). Some of the many examples from our group include the gene-specific characterization of human histone H2B (83), the H2A gene family (84), quantitiation of human H4 isoforms (85), specific methylation of H4K20 (86, 87) and characterization and quantitiation of H3 modification from human cells (88). More recently, Martinez-Garcia et al. provided a direct picture of the change in methylation of H3.1, H3.2, and H3.3 based on MMSET (a histone methyltransferase) and the lack of or presence of such. (Fig. 3, bottom-left) (82). In regards to the other proteins associated with chromatin, Garcia et al. have also performed much work on histones and protein classes such as high mobility groups and the characterization of PTM dynamics (89). Further, Coon et al. have extended histone TD proteomics to analysis with Orbitraps and the identification of 74 unique histone forms (90). There are other examples of protein families and the TD analysis of their isoforms and PTM status. However, we turn our attention to plant proteome applications.

Many of the described platforms may also be applied to plant proteome analysis, and more specifically membrane proteins. The analysis of integral membrane proteins is difficult due to their hydrophobic nature. Proteins of this subset are often analyzed with variations in mobile phase conditions to improve solubility (91). Some examples include the analysis of several chloroplast proteins resulting in the mapping of a variety of PTMs and enzymatic cofactors (92, 93). Additionally, 11 integral membrane proteins identified from red algae photosystem II (a 750 kDa complex) resulted in the characterization of several PTMs (94). The use of organic solvents for membrane protein extraction followed by separation using HILIC has also been described (95, 96). These studies resulted in the identification of several integral membrane proteins from mitochondria.

As a second platform for the analysis of the plant proteome, genetics coupled to 2-D Difference Gel Electrophoresis (DIGE), can be used to investigate the genetic architecture of complex phenotypes (Fig. 4) (97). Cilia and colleagues used 2-D DIGE to phenotype sister aphid F2 genotypes that segregated in their ability to transmit plant viruses (97). The study revealed phenotyping and virus transmission information. Results included the description of genetic heterogeneity in the aphid’s bacterial endosymbiont Buchnera aphidicola co-segregating with the virus transmission phenotype, and heritable bacterial and aphid protein isoforms linked to the virus transmission phenotype and to specific virus transmission barriers within the aphid. There are other fine DIGE/Bottom Up experiments performed on human and other species beyond the scope of
this review. As mentioned above, tissue imaging with MALDI-TOF is also another form of intact protein analysis.

Tissue analysis with MALDI-TOF results in the mapping of protein species for different clinical applications (Fig. 3, bottom right) (43). Intact protein imaging from tissue is often followed by TD or BU protein characterization, unless the protein of interest has known molecular weight a priori. For example, Schey et al. has imaged ocular lenses and retinal tissue from human, bovine and rabbit (98). This study resulted in spatial resolution and mapping the distribution of G-protein coupled receptor in the rabbit retina and an integral membrane protein (AQP0) in human and bovine lenses could be mapped. Imaging normally consists of the analysis of soluble proteins, however the preparation in their study also allowed for the analysis of integral membrane proteins (98). Imaging can also be demonstrated for tissues, as Cazares et al. compared tissue slices from benign and cancerous prostates and found that Mitogen-activated protein kinase 2 (MEKK2) was prevalent in the cancerous tissue (99). For an additional example for profiling cancer, Rauser et al. reported on the usefulness of MS imaging for breast cancer and the human epidermal growth factor receptor 2 in regards to decision making and therapeutics (100). To confirm the findings, the group also accessed TD and BU proteomics techniques to fully identify the proteins.

CONCLUSIONS

This review has briefly covered the broad field of intact protein analysis by “TD” MS. Measurement of intact proteins (isoforms and protein species) should result in stronger correlations between mass spectrometric data and complex phenotypes. To the extent that the above hypothesis is true, TD and its focus on analyzing protein molecules in an isoform-resolved manner will help make the interface between biological MS and translational biomedicine more efficient. TD protein analysis has been driven by biological diversity to have many different platforms. Further, examples were presented that illustrate the usefulness of different platforms. TD proteomics can reveal the rich isoform and PTM diversity found in nature and the human body.

REFERENCES

FOOTNOTES

* The authors would like to thank Ying Ge from the Human Proteomics Program and Department of Physiology, School of Medicine and Public Health, University of Wisconsin; Jennifer Zhang from Genentech Inc; Erin Seeley from Vanderbilt University School of Medicine; and Michelle Cilia from Functional and Comparative Proteomics Center Robert W. Holley Center for Agriculture and Health for aid in the generation of Figures 3 and 4. We further acknowledge the Institute for General Medical Sciences at the National Institutes of Health (NIH R01 GM 067193) and the Institute on Drug Abuse (P30 DA 018310) for funding.

The abbreviations are: PTMs, post-translational modifications; IUPAC, International Union of Pure and Applied Chemistry; MS, mass spectrometry; TD, Top down; BU, Bottom Up, MS/MS, tandem mass spectrometry; 2D-GE, 2 dimensional gel electrophoresis (2D-GE); ESI, electrospray ionization; MALDI, matrix assisted laser desorption ionization; RP LC, reversed-phased liquid chromatography; DIGE, 2-D difference gel electrophoresis; SILAC, stable isotope labeling of amino acids; FTICR, Fourier transform ion cyclotron resonance; TOF, Time-of-flight; ECD, electron capture dissociation; LTQ, linear quadrupole ion trap; Q, quadrupole; CID, collisional induced dissociation; ETD, electron transfer dissociation; HCD, higher energy collisional activated dissociation; RP nLC, reversed phase nano-liquid chromatography; PIITA, precursor ion independent top-down algorithm; cSNPs, Single Nucleotide Polymorphisms within cDNA sequences; DIGE, 2-D Difference Gel Electrophoresis.

FIGURE LEGENDS

Figure 1. *Top Left:* Pictorial representation of a homologous pair of chromosomes. A single locus is sequenced to reveal molecular variation, such as a heterozygous genotype from a single nucleotide polymorphism in the human population. *Center:* RNA variations arise from such events as alternative splicing (or RNA-editing). *Right:* Hypothetical picture of a protein network. Underneath a single node can lie whole populations of isoforms that are not characterized well with mainstream proteomic methods. These isoforms arise from a combination of molecular variation including single nucleotide polymorphism, alternative splicing, and post-translational modifications (PTMs). Depending on the state of the cell or tissue, isoform specific responses can be masked when analyzed by proteomics based on tryptic digestion.

Figure 2. The diagram displays a continuum from single proteins analyzed in targeted studies to full proteomes and conveys the current state of technology where complete characterization of highly modified proteins usually correlates with lower throughput. For interrogation of complex mixtures, multidimensional separations are required for sample preparation to address the classic “front end” problem in MS. As for Bottom Up, analyzing larger numbers of proteins in high-throughput mode enables a large number of protein identifications but often requires a trade off in the quality of molecular characterization (e.g., PTMs or SNPs are not localized to a specific site).

Figure 3. Four examples of intact protein analysis are presented. The full findings are described in the main text. The data at top left resulted in the characterization of cardiac troponin (cTnI) with several phosphorylations (76). The top right spectra include the analysis of antibodies with LTQ-Orbitrap, and the final deconvoluted spectra, including the detection of glycan forms (48). The bottom left presents the analysis of histones which were derived from MLL cell lines and had changes in the MMSET (a histone methyltransferase) expression (82). Up-regulation of MMSET expression increased the methylation states of H3.1, H3.2, and H3.3. The bottom right panel presents the capability of imaging whole tissue using whole protein MS, with the spatial resolution combined with detection of protein across the MW range of 1000 to 30,000 Da (43).
Figure 4. Left side is a full DIGE analysis of CYDV-RPV transmission-competent phenotype. Boxes (1,2,3) highlight changes between *Schizaphis graminum* and the CYDV-RPV vectors. More information on the observed difference may be found in reference 97. This work represents an example of the concept of “digital” or “qualitative” protein biomarkers, where the simple observation of a protein form (and not its level of expression) correlates tightly to an overall organismal or cellular phenotype.
Figure 1

DNA

RNA

Protein

Alternative Splicing

Single Nucleotide Polymorphism

T → G

Group 1

Group 2

Protein Interaction Map

Variations in Gene Product

Dynamic PTMs

Isoforms
Increasing Sample Complexity

Single “Complex” Proteins

Protein Families/Complexes

Whole Cell / Tissue

Characterization-Mode (off-line, targeted)

Biologics

Immunoprecipitation

Overexpression

Figure 2

Throughput-Mode (on-line, LC-MS)

Protein-Protein Interactions

Organelle Analysis

Sample Fractionation

Global Proteomics

Multidimensional Separations

Overexpression

Immunoprecipitation

Biologics
DIGE for Plant Proteomics and Virus Transmission by Aphids

Figure 4