



Published in final edited form as:

Chem Rev. 2013 April 10; 113(4): 2343–2394. doi:10.1021/cr3003533.

## Protein Analysis by Shotgun/Bottom-up Proteomics

Yaoyang Zhang<sup>1,†</sup>, Bryan R. Fonslow<sup>1,†</sup>, Bing Shan<sup>1</sup>, Moon-Chang Baek<sup>1,2</sup>, and John R. Yates III<sup>1,\*</sup>

Yaoyang Zhang: zhangyy@scripps.edu; Bryan R. Fonslow: fonslow@scripps.edu; Bing Shan: bingshan@scripps.edu; Moon-Chang Baek: mcbaek@scripps.edu

<sup>1</sup>Department of Chemical Physiology, The Scripps Research Institute, La Jolla, CA 92037, USA

<sup>2</sup>Department of Molecular Medicine, Cell and Matrix Biology Research Institute, School of Medicine, Kyungpook National University, Daegu 700-422, Republic of Korea

### 1 Introduction

According to Genome Sequencing Project statistics (<http://www.ncbi.nlm.nih.gov/genomes/static/gpstat.html>), as of Feb 16, 2012, complete gene sequences have become available for 2816 viruses, 1117 prokaryotes, and 36 eukaryotes.<sup>1–2</sup> The availability of full genome sequences has greatly facilitated biological research in many fields, and has greatly contributed to the growth of proteomics.

Proteins are important because they are the direct bio-functional molecules in the living organisms. The term “proteomics” was coined from merging “protein” and “genomics” in the 1990s.<sup>3–4</sup> As a post-genomic discipline, proteomics encompasses efforts to identify and quantify all the proteins of a proteome, including expression, cellular localization, interactions, post-translational modifications (PTMs), and turnover as a function of time, space and cell type, thus making the full investigation of a proteome more challenging than sequencing a genome. There are possibly 100,000 protein forms encoded by the approximate 20,235 genes of the human genome,<sup>5</sup> and determining the explicit function of each form will be a challenge.

The progress of proteomics has been driven by the development of new technologies for peptide/protein separation, mass spectrometry analysis, isotope labeling for quantification, and bioinformatics data analysis. Mass spectrometry has emerged as a core tool for large-scale protein analysis. In the past decade, there has been a rapid advance in the resolution, mass accuracy, sensitivity and scan rate of mass spectrometers used to analyze proteins. In addition, hybrid mass analyzers have been introduced recently (e.g. Linear Ion Trap-Orbitrap series<sup>6–7</sup>) which have significantly improved proteomic analysis.

“Bottom-up” protein analysis refers to the characterization of proteins by analysis of peptides released from the protein through proteolysis. When bottom-up is performed on a mixture of proteins it is called shotgun proteomics,<sup>8–10</sup> a name coined by the Yates lab because of its analogy to shotgun genomic sequencing.<sup>11</sup> Shotgun proteomics provides an indirect measurement of proteins through peptides derived from proteolytic digestion of intact proteins. In a typical shotgun proteomics experiment, the peptide mixture is fractionated and subjected to LC-MS/MS analysis. Peptide identification is achieved by comparing the tandem mass spectra derived from peptide fragmentation with theoretical tandem mass spectra generated from *in silico* digestion of a protein database. Protein

\*Corresponding author: Prof. John R. Yates III, Phone: +1-858-784-8862. Fax: +1-858-784-8883, jyates@scripps.edu.

†These authors contributed equally to this work.

inference is accomplished by assigning peptide sequences to proteins. Because peptides can be either uniquely assigned to a single protein or shared by more than one protein, the identified proteins may be further scored and grouped based on their peptides. In contrast, another strategy, termed 'top-down' proteomics, is used to characterize intact proteins (Figure 1). The top-down approach has some potential advantages for PTM and protein isoform determination and has achieved notable success. Intact proteins have been measured up to 200 kDa,<sup>12</sup> and a large scale study has identified more than 1,000 proteins by multi-dimensional separations from complex samples.<sup>13</sup> However, the top-down method has significant limitations compared with shotgun proteomics due to difficulties with protein fractionation, protein ionization and fragmentation in the gas phase. By relying on the analysis of peptides, which are more easily fractionated, ionized and fragmented, shotgun proteomics can be more universally adopted for protein analysis. In fact, a hybrid of bottom-up and top-down methodologies and instrumentation has been introduced as middle-down proteomics.<sup>14</sup> Essentially, middle-down proteomics analyzes larger peptide fragments than bottom-up proteomics, minimizing peptide redundancy between proteins. Additionally the large peptide fragments yield similar advantages as top-down proteomics, such as gaining further insight into post-translational modifications, without the analytical challenges of analyzing intact proteins. Shotgun proteomics has become a workhorse for the analysis of proteins and their modifications and will be increasingly combined with top-down methods in the future.

In the past decade shotgun proteomics has been widely used by biologists for many different research experiments, advancing biological discoveries. Some applications include, but are not limited to, proteome profiling, protein quantification, protein modification, and protein-protein interaction. There have been several reviews nicely summarizing mass spectrometry history,<sup>15</sup> protein quantification with mass spectrometry,<sup>16</sup> its biological applications,<sup>5,17–26</sup> and many recent advances in methodology.<sup>27–32</sup> In this review, we try to provide a full and updated survey of shotgun proteomics, including the fundamental techniques and applications that laid the foundation along with those developed and greatly improved in the past several years.

## 2 Techniques

### 2.1 Protein extraction and isolation methods

Proteins are part of a complex network of interacting biomolecules that regulate their function and localization within the cell. Extraction and isolation of proteins from chemical and physical interactions with other biomolecules from specific cellular sub-compartments has become a critical step for their global analysis in a biological context. In some cases, physical and chemical interactions may otherwise inhibit the isolation or analysis of proteins of interest by LC-MS. The global analysis of membrane-embedded proteins is a prominent example. Isolation, solubilization, and proteolytic digestion of lipid-bound proteins have all proven to be essential steps in their shotgun proteomic analysis. The integration of multiple methodological advancements for analysis of membrane-bound proteins is described in the "Proteolytic Digestion and Chromatographic Separations" sections. Another noteworthy example of protein isolation from interactions with other non-protein biomolecules that improved proteomic analysis was recently demonstrated with transcription factors (TF). Under specific biological conditions, TFs form complexes with high affinity for DNA. Until recently their proteomic analysis was thought to be limited by their low abundance, but was instead due to inadequate disruption of TF-DNA interactions. The degradation of DNA using a combination of deoxyribonucleases improved the recovery of transcription factors from standard hypotonic-lysed cells, facilitating their targeted proteomic analysis.<sup>33</sup> Proteins have also been isolated from more rigid structures, such as bone tissue using hydrochloric acid, allowing for the identification of ~2,500 proteins with shotgun proteomics.<sup>34</sup>

Application of traditional subcellular isolation techniques, primarily sucrose gradient sedimentation and similar methodologies, from different cell types and tissues have allowed for global analysis of proteins within subcellular compartments. Studies of the differential expression and trafficking of proteins between subcellular compartments are important to understanding proper cellular function. Characterization of the principal subcellular proteomes, nuclear and cytosolic, has been performed on yeast,<sup>35–36</sup> leukemia cells,<sup>37</sup> a bronchitis virus cellular model,<sup>38</sup> and an autoimmunity cell model.<sup>39</sup> A study of mouse liver used protein correlation profiling to map ~1,400 proteins to 10 subcellular localizations with validation by enzymatic assays, marker protein profiles, and confocal microscopy.<sup>40</sup> Further studies have teased apart finer subcellular compartments, such as the nucleolus. Significant changes to the nucleolar proteome were found upon infection with the Influenza A virus.<sup>41</sup> Many studies have begun to interrogate the energy-producing organelle within the cell, the mitochondrion. Mitochondrial proteins have been identified from *C. elegans*,<sup>42</sup> rat liver,<sup>43</sup> and leukemia cell lines<sup>44</sup> by coupling biochemical sedimentation purification with shotgun proteomic analysis. Similarly, the protein post-translational modifications phosphorylation and carbonylation have been identified from mitochondrial preparations of murine heart and skeletal muscle, respectively.<sup>45–46</sup> A quantitative comparison of mitochondrial proteins among rat heart, liver, and muscle tissues found essentially the same proteins with tissue-specific abundances.<sup>47</sup> Combining many of these methodologies, approximately 22 protein abundance changes were found among subcellular fractions due to myocardial ischemia.<sup>48</sup> Another highly studied subcellular fraction using proteomic methods is the synaptosome. Studies on the synaptosome have implications in many aspects of neuroscience research. The synapse is the biochemical communication junction unique to neuronal cells and can be isolated within a synaptosomal preparation from multiple differential centrifugation steps. As with all subcellular proteomic studies discussed thus far, the quality of the data is heavily dependent on the purity of the subcellular preparation. In the case of the synaptosome, correlation-profiling has been employed to validate post-synaptic density proteins.<sup>49</sup> Similar quantitative proteomic methodologies have facilitated the characterization of post-synaptic density proteins between rat forebrain and cerebellum,<sup>50</sup> hippocampal synaptosome protein changes in CAM kinase II mutant mice,<sup>51</sup> and synaptosomal protein dynamics and spatiotemporal dynamics of synaptosomal and non-synaptosomal mitochondrial proteins during brain development.<sup>52–53</sup> Quantitative phosphoproteomic analysis of KCl-activated synaptosomes found a uniquely regulated phosphorylation site on the glutamate receptor subunit GluR1.<sup>54</sup> Further advances in subcellular fractionation and proteomic methodologies will aid in understanding the complex dynamics of proteins among cellular sub-compartments.

## 2.2 Protein depletion and equalization methods

Protein dynamic range is the largest challenge that faces proteomics technology development. Currently, all steps within an LC-MS proteomics pipeline are protein abundance-dependent. Thus, adjustment of protein concentration dynamic range has become an option for improving comprehensiveness through improved analysis of low abundance proteins. Two main approaches are used for protein dynamic range adjustment: (1) selective depletion of known high abundance proteins and (2) selective equalization of protein dynamic range using combinatorial ligand libraries. In particular, both of these strategies have proven most useful for the analysis of plasma and other clinical samples. Specific applications of this will be described in the “Clinical Applications” section. Clinical samples for proteomic analysis are often more complex than model systems and present higher protein dynamic ranges, up to nine orders of magnitude.<sup>55</sup> Fortunately, only a few proteins are extremely abundant, such as serum albumin in the case of plasma, and thus can be specifically removed or depleted prior to LC-MS analysis. Chemical-based approaches can selectively precipitate abundant proteins, usually albumin, from plasma to improve

proteomic depth and have been demonstrated with sodium chloride and ethanol,<sup>56</sup> acetonitrile,<sup>57</sup> the disulfide reducing agents DTT and TCEP,<sup>58</sup> and ammonium sulfate.<sup>59</sup> Antibody arrays against the highest abundance proteins have also improved proteomic coverage of clinical samples,<sup>60</sup> yet there remain significant shortcomings for the analysis of low abundance proteins.<sup>61</sup> Although relatively effective, another drawback to antibody depletion methodologies is the high cost of reagents (Sigma 20 protein single depletion kit - \$1,200). Additionally, since the depletion efficiency relies on the binding capacity of the antibodies only small sample amounts (8  $\mu$ L plasma) can be depleted. Ultimately this depletion capacity can limit the starting mass of low abundance proteins within the sample. Nonetheless, antibody depletion is a critical step for most clinical applications.

Protein abundance dynamic range adjustment using combinatorial ligand libraries is an alternative, cheaper, and more holistic approach. Bead-bound combinatorial peptide ligand libraries simultaneously deplete abundant proteins while enriching low abundance proteins<sup>62</sup> and can be performed for \$50 per experiment. The equalization strategy is the converse of depletion strategies in that proteins which bind to a bead-conjugated hexapeptide ligand are collected and used for analysis. The combinatorial hexapeptide library serves as a large collection of ligand epitopes for proteins to bind. Additionally, hexapeptide ligands are assumed to be at approximately equimolar amounts, allowing for equalization of protein concentration. An obvious drawback to this methodology is that for a protein to be retained for analysis it must have affinity for one of the millions of possible hexapeptide ligands that are represented on the bead library used.<sup>63</sup> High abundance proteins that saturate their equimolar hexapeptide ligands can be washed away, while low abundance proteins can be concentrated and enriched on the solid-phase beads, especially if they have high affinity for their associated ligand or ligands. The benefits of protein equalization with hexapeptide beads in conjunction with shotgun proteomics was demonstrated on a model human cell line.<sup>64</sup>

More recently, classic Michaelis-Menten enzyme kinetics were exploited to equalize proteomes in an unbiased fashion.<sup>65</sup> The methodology cleverly uses a protease, already used in shotgun proteomic pipelines, to selectively digest abundant proteins into peptides for removal with a molecular weight cutoff spin-filter. The remaining less abundant, undigested proteins are then also digested to completion, as routinely performed and described in the following section. With the abundant proteins depleted and the proteome equalized, dramatic improvements were observed in the total number of protein identifications and the sequence coverage and quantitation metrics of low abundance proteins. This strategy directly addresses one of the most daunting challenges of globally analyzing proteomes, protein abundance dynamic range, and presents a simple and versatile strategy to significantly improving shotgun proteomic analysis.

### 2.3 Proteolytic digestion methods

Analysis of proteins from their proteolytic peptides circumvents some of the challenges associated with intact protein separation, ionization, and MS characterization. A protein lysate is a highly heterogeneous mixture of proteins with diverse physicochemical properties. Purposefully increasing the complexity of a sample prior to analysis is somewhat counter-intuitive. However, selective protease digestion acts to normalize and compartmentalize the biochemical heterogeneity of proteins within a sample as peptides and may, in fact, create a less heterogeneous mixture when protein splice isoforms and post-translational modifications are considered. Additionally, with multiple representations of a protein as peptides the probability of sampling and identifying a peptide associated with a particular low abundance protein and/or post-translational modification increases.

In general, proteolytic enzymes differ by their specificity for cleaving the amide bonds between individual residues in a protein. Commonly used proteases with their biochemical specificity and applications are listed in Table 1. The cleavage is carried out through hydrolysis of the amide bond before or after a specific residue, residues, or combination of residues. Trypsin has become the gold standard for protein digestion to peptides for shotgun proteomics. Trypsin is a serine protease which cleaves at the carboxyl side of arginine and lysine. This sequence-specific information has been used to filter identified peptides. However, high accuracy mass spectrometers have reduced the importance of this filtering criterion and allowed for identification of non-tryptic protein sequences and post-translational modifications where trypsin cleavage is inhibited.<sup>66–68</sup> For low complexity samples, such as protein complexes, the combination of both highly selective and non-selective proteases improves protein and post-translational modification coverage.<sup>69–70</sup> For complex proteomic samples, the utilization of a combination of highly selective proteases improves protein and proteome coverage and sensitivity by creating complementary peptides.<sup>71–72</sup> These multi-protease analyses are performed in parallel, as a systematic study showed that purposefully creating more peptides with two proteolytic enzymes reduced the number of proteins identified.<sup>73</sup> Similarly, parallel analysis of a tryptic digestion and subsequent proteolytic digestion with Glu-C after size-based isolation of long tryptic peptides improved protein, proteome, and phosphorylation identification coverage.<sup>74</sup> Related studies also illustrated the effects of protease biases on global phosphorylation identification studies based on identified phosphorylation motifs.<sup>75–77</sup>

Digestion efficiency has been optimized based on a number of reaction conditions. In particular, adequate solubilization and unfolding of all proteins in a complex mixture is important to provide a protease access to cleavage sites. The use of organic solvents during trypsin digestion has been shown to improve digestion efficiency based on peptide identifications, protein sequence coverage, and trypsin specificity on protein complex purifications<sup>78–79</sup> and complex proteomic mixtures.<sup>80</sup> SDS remains one of the best protein solubilizers, but is detrimental to LC-MS peptide sensitivity, as are most traditional surfactants, so strategies for removal are continually developed. SDS works well with protease digestions because it is a strong chaotropic agent, or chaotrope – a substance which denatures and disrupts the structure of macromolecules. A chaotrope swapping strategy was previously demonstrated<sup>81</sup> for removal of SDS while maintaining protein denaturation and was recently reintroduced as beneficial with higher sensitivity mass spectrometers.<sup>82</sup> As an alternative to pre-digestion removal of SDS, a KCl precipitation strategy was demonstrated post-digestion to provide the benefits of SDS-assisted proteolytic digestion without the detrimental effects during LC-MS analysis.<sup>83</sup> A number of LC- and MS-compatible surfactants have also been developed and evaluated to improve protein identification comprehensiveness.<sup>80,84</sup> Generally, the commercially-available MS-compatible surfactants (e.g. ProteaseMAX, Invitrosol, Rapigest, PPS Silent Surfactant) have an acid labile moiety within the surfactant structure so it can be degraded during or after digestion and prior to LC-MS. The degradation creates components which do not co-elute with peptides during LC-MS, making them compatible with common shotgun proteomics methods. Two less-common, volatile surfactants which can be evaporated prior to LC-MS have also proven useful for solubilization and digestion of membrane-bound proteins: perfluorooctanoic acid<sup>85</sup> and 1-butyl-3-methyl imidazolium tetrafluoroborate.<sup>86</sup> As an alternative to surfactants, trifluoroethanol has proven useful for concurrent protein extraction and denaturation for mass-limited samples where sample clean-up is usually detrimental to sensitivity.<sup>87</sup> Avoidance of chaotropes for protein solubilization through the use of ammonium bicarbonate alone was proposed to improve detection limits in an analysis of a few hundred to thousand cancer cells.<sup>88</sup>

Modification of protein digestions using physical methods has also contributed to improved digestion efficiency and proteomic coverage. Digital microfluidics allowed for automation of sample clean-up and protein digestion steps for MALDI to improve sensitivity.<sup>89–93</sup> Covalent and dynamic immobilization of trypsin within microreactors,<sup>94–95</sup> on microparticles<sup>96</sup> and nanoparticles<sup>97</sup> and the use of focused highly-intensity ultrasound<sup>98–99</sup> and microwave<sup>100–103</sup> heating have improved the kinetics of tryptic digestion, reducing digestion time. The immobilization of proteases on microwave-absorbing microspheres and nanoparticles further improved speed and efficiency.<sup>104–105</sup> Microwave heating of proteins under acidic conditions selectively cleaves proteins at aspartic acid residues, creating complementary peptides of similar length to trypsin digestion.<sup>106–107</sup> Sequential microwave-assisted acid hydrolysis and overnight protease digestion proved useful for the digestion of extremely thermal- and biochemically-stable proteins from the hyperthermophile *Pyrococcus furiosus*.<sup>108</sup> Raising the pressure of proteolytic digestion has also improved the efficiency of proteolytic digestion,<sup>109</sup> presumably by improved unfolding of proteins and greater mixing, and allowed for online digestion within an LC-MS pipeline.<sup>110</sup>

## 2.4 Protein separation and fractionation methods

Two dimensional polyacrylamide gel electrophoresis (2D-PAGE) is a powerful method for the separation of complex mixtures of proteins for proteomic analysis.<sup>111,112</sup> This method is based on molecular mass and charge and is capable of separating several thousands of intact proteins on a single gel. Following separation on 2D-PAGE, proteins can be identified as intact proteins or peptides using MALDI-MS or with LC-MS after an in-gel digestion.<sup>113</sup>

A combination of detergents (such as SDS and CHAPS) and chaotropes (such as urea and thiourea) have generally been used for solubilization and denaturation of protein samples prior to separation on the 2D gel.<sup>114,115</sup> Solubilized proteins are separated by isoelectric focusing (IEF), where a gradient of pH is applied to a gel, followed by application of an electric potential. Development of immobilized pH gradient (IPG) strips, which are far simpler to use than carrier-ampholyte-driven pH gradients, has led to greater popularity of the 2D method in the proteomics field. Enhanced reproducibility and generation of various pH gradient types, such as basic gradients,<sup>116</sup> non-linear pH gradients,<sup>117</sup> or narrow pH gradients<sup>118</sup> using plastic-supported IPG strips have been reported.

After 2D-PAGE separation, proteins can be detected by Commassie Brilliant Blue or silver staining. Despite greater sensitivity, silver staining does not show linearity of signal and is less compatible with MS, therefore Commassie Brilliant Blue is most commonly used.<sup>119</sup> A formaldehyde-free silver staining method that is compatible with MS has been developed.<sup>120</sup> The Pro-Q Diamond and Pro-Q Emerald staining methods have been developed for phosphoproteins and glycoproteins, respectively.<sup>121,122</sup>

Use of 2D-PAGE is labor-intensive and time-consuming, with a low dynamic range and significant gel-to-gel variation.<sup>123</sup> Introduction of 2D difference gel electrophoresis (DIGE) has overcome some of the drawbacks associated with use of 2D gel electrophoresis, and provides more accurate and sensitive results. This method utilizes the fluorescent property of Cy3 and Cy5 N-hydroxysuccinimidyl ester cyanines, which show different excitation fluorescent spectra at different wavelengths. Typically, two samples are labeled with fluorescent dyes Cy3 and Cy5, respectively, and an internal standard is labeled with Cy2. The three samples are mixed and analyzed on one gel,<sup>124</sup> which permits measurement of protein amount in each sample relative to the internal standard in which the amount of each protein is known. Software such as DeCyder can be used for the detection of spots, and for normalization and analysis of data, which can increase quantitative accuracy and speed. Gel spots can also be digested by trypsin, and analyzed by MALDI-ToF, which generates a peptide mass fingerprint, or LC-MS/MS, which can provide peptide sequence information.

High resolution separation of intact proteins based on protein size is a unique characteristic of 2D-PAGE. This feature can also be used to identify proteins which are degraded under specific conditions such as apoptosis. In one such study, cell extracts were treated with and without the proteases granzyme B<sup>125</sup> and caspase 3<sup>126</sup>, then run on 2D PAGE to find differences. By taking advantage of the change of pI induced by various modifications such as phosphorylation, 2D-PAGE can also be applicable to study of PTMs. Phosphorylation primarily changes the pI of proteins which results in a shift in the mobility pattern in a horizontal direction on the gel. The modified form shows an extra spot, which does not migrate together with the principal spot, and modified peptides can be identified using MS. 2D gel-based proteomics is a mature technology that has been employed in proteomics for over 3 decades and is still useful to study bacterial proteomics with low complexity,<sup>127</sup> intact protein with post-translational modifications<sup>128</sup> and as a micropreparative tool,<sup>129</sup> among other things.

The development of LC-MS/MS technologies has led to a reduction in 2D gel-based separations, but 1D-SDS-PAGE remains a standard separation method for complex protein mixtures based on their molecular weight.<sup>130</sup> Separated complex protein mixtures are cut into 10–20 gel slices and each gel band is digested by in-gel trypsin method.<sup>131</sup> The methodology has been standardized as GeLCMS (gel-enhanced LC-MS) to represent the hybrid gel and LC-MS analysis.<sup>132</sup> Compared to gel-free methods, this method has some important advantages, including differentiating splice isoforms and degraded proteins and removal of low molecular weight impurities such as detergents and buffer components, which are detrimental for MS analysis. However, both sample loss to the gel and residual SDS, can limit sensitivity and comprehensiveness. The other, orthogonal separation from 2D gel electrophoresis, isoelectric focusing, has also been exploited for protein separation prior to LC-MS. An immobilized pH gradient strip in the Off-Gel™ electrophoresis (OGE) system focuses proteins based on their pI, allowing for their recovery in the liquid-phase.<sup>133</sup> Improvements from protein fractionation by OGE and bottom-up proteomic analysis were demonstrated with a human plasma sample.<sup>134</sup> A comparative GeLCMS analysis with a solution-phase IEF separation identified 25% more proteins and improved reproducibility to 96%.<sup>135</sup> Free-flow electrophoresis followed by RPLC was used to fractionate proteins from a model cell line and human serum prior to MS/MS identification as peptides.<sup>136–137</sup> Chromatofocusing<sup>138</sup> and size-exclusion chromatography (SEC)<sup>139</sup> have also been used prior to shotgun proteomic workflows to improve comprehensiveness. Combining solution IEF, RPLC, and 2DIGE, as an intact protein analysis system (IPAS), was used to quantitatively profile the human plasma proteome.<sup>140</sup> A similar methodology which followed solution IEF and RPLC, using the Proteome Lab™ PF 2D platform, by LC-MS analysis fractionated and identified proteins from human plasma gel-free.<sup>141</sup>

## 2.5 Peptide ionization, separation, and fractionation methods

Shotgun proteomics has become a widely used method for multiple reasons. One notable reason is the amenability of common analytical chromatography methods for separation of peptides prior to ESI-or MALDI-MS/MS. Since the initial demonstrations of online solid-phase extraction, strong cation exchange fractionation, and nanoflow liquid chromatography for direct ESI with homemade columns,<sup>8</sup> numerous academic and commercial configurations have been reported which further automate and improve the process. As a sign of the wide-spread need and benefits, commercial supplies and comparable configurations are available from Agilent Technologies Inc., New Objective Inc., and Shotgun Proteomics Inc.

The two main ionization methods used for charging and transferring peptide into the gas-phase in shotgun proteomics are nanoelectrospray (nESI)<sup>142</sup> and MALDI.<sup>143</sup> nESI has been generally coupled with RPLC separations and MALDI with gel-based separations. MALDI

is most often performed by transferring proteins or peptides from a gel to a support or substrate,<sup>144</sup> but methodologies were developed to perform MALDI directly from 2D gels to simplify the process and improve reproducibility.<sup>145–146</sup> There has also been significant cross-over of separation and ionization methods in both application and development of new ionization strategies. These include coupling HPLC<sup>147</sup> and CE<sup>148</sup> to MALDI in online<sup>149</sup> and offline<sup>150–151</sup> peptide fractionation modes. Collection of peptides from HPLC onto MALDI plates allowed for exclusion of redundant peptides from replicate analyses.<sup>152</sup> Due to the higher tolerance to surfactants than ESI, LC-MALDI was more comprehensive and complementary to gel and standard LC-ESI methods.<sup>153</sup> Desorption ESI (DESI), where ESI can be used to vaporize deposited peptides from a substrate, has been described for bottom-up proteomic analysis.<sup>154–155</sup> Other similar methods which combine DESI and MALDI have also been demonstrated,<sup>156–157</sup> including the capability to perform chemical reactions.<sup>158</sup> Laserspray ionization (LSI) has also produced ESI-like peptide spectra from a standard MALDI plate.<sup>159</sup> Capacitive charging of a number of substrates has been used to ionize peptides through electrostatic-spray ionization (ESTASI).<sup>160</sup> Peptides have even been vaporized and ionized from a solid-state small molecule matrix solely by introduction to vacuum, yielding ESI-like peptide charge states.<sup>161</sup> Similarly, new ESI-like ionization methods have either completely removed the electrical potential or changed the way it is applied to improve sensitivity and selectivity. Simple placement of the outlet of a HPLC column into the heated mass spectrometer inlet under vacuum, deemed solvent-assisted inlet ionization (SAII), vaporizes peptides with similar charge states to ESI and greater sensitivity.<sup>162</sup> SAII has shown to maintain chromatographic peak shape and is compatible with  $\mu\text{L}/\text{min}$  and  $\text{nL}/\text{min}$  flow rates.<sup>163–164</sup> Through application of an AC voltage, inductive ESI can control the charge state of peptides.<sup>165</sup> Introduction of an inverted post-column makeup flow during gradient elution HPLC improved ESI signal stability throughout the separation.<sup>166</sup> Continued improvements to peptide ionization methods will surely aid in shotgun proteomics, likely through improved sensitivity and expansion of MS-compatible liquid-phase separations.

Peptide chromatography and mass spectrometry are dramatically simpler both theoretically and experimentally than at the protein level, making methods to analyze proteins at the peptide level more straightforward. However, rigorous LC separation of peptides is critical for detection of peptides from a complex proteomic mixture for a number of reasons. Generally, in order for a peptide to be sampled for fragmentation its precursor intensity must be both above background noise and more abundant than other peptide ions measured simultaneously. Resolution of peptides of similar masses is essential in order to acquire distinct, non-mixed fragmentation spectra. This criterion is best met by separation in the liquid phase through chromatography since differences in charge and hydrophobicity can be exploited. Adequate chromatographic separation is also crucial for effective electrospray ionization (ESI). Although ESI is a powerful, relatively unbiased method for introducing peptides into the gas phase, it can easily be crippled by ionization suppression. Attempting to electrospray many analytes simultaneously, whether peptides or contaminants, results in ionization of only the most hydrophobic molecules.<sup>167</sup> This concept has resonated in the shotgun proteomics community over the years and has been reconfirmed recently in both untargeted and targeted peptide analyses.<sup>168–169</sup> In fact, exploiting this concept by increasing the hydrophobicity of peptides through alkylation of primary amines has increased ionization efficiency and signal intensity.<sup>170</sup> Thus resolution of peptides prior to ESI-MS/MS is of utmost importance. The challenge of adequate separation of peptides from a complex mixture becomes particularly evident from an examination of a yeast proteome, which, with  $\sim 7000$  gene products, is far less complex than a comparable human sample. A theoretical *in silico* digestion generated ca. 300,000 peptides.<sup>171</sup> This number of peptides is already nearly 10 times greater than the number that is commonly identified in a single comprehensive proteomic analysis. The most complex human samples, such as plasma, can

have a concentration dynamic range as high as 12 orders of magnitude with potentially 25,000 gene products<sup>55</sup> and even model organisms like *C. elegans* and *D. melanogaster* are predicted to have nearly as many genes as humans.<sup>172–173</sup> The first comprehensive proteomic analyses of model organisms using state-of-the-art proteomics technology, which exploited multidimensional separations of peptides, identified 2,388 (*S. cerevisiae*),<sup>174–175</sup> 10,631 (*C. elegans*),<sup>176</sup> and 9,124 (*D. melanogaster*)<sup>177</sup> proteins. More recent analyses with both improved peptide separations and more sensitive mass spectrometers have boosted protein identification numbers for yeast above 4000 proteins.<sup>178–179</sup>

Early and more recent applications of mass spectrometry for proteomics used either gel-based or offline LC fractionation of peptides. Although many of these methodologies are still useful for targeted and global proteomic studies, the largest gains in sensitivity from peptide separations were initially demonstrated and still used today by coupling nano-electrospray<sup>142</sup> to reverse-phase nanoflow liquid chromatography (nLC). A representative base peak chromatogram is shown in Figure 10a. Peptides were loaded directly onto a homemade nLC capillary column, separated based on hydrophobicity, and were directly electrosprayed from the capillary tip into the mass spectrometer.<sup>180–181</sup> The sensitivity of this methodology, which avoids the use of auto samplers, was dramatically redemonstrated in the phosphoproteomic analysis of ~10,000 cells.<sup>182</sup> Employing proven chromatographic technological improvements has also benefited in shotgun proteomic analysis. Reducing the column inner diameter and eluent flow to low nanoliter per minute rates while increasing the column length has produced similar results to longer multidimensional separations through both increased peak capacity and ionization efficiency.<sup>183</sup> High temperature RPLC has proven essential for the separation and identification of hydrophobic peptides from membrane-embedded proteins.<sup>184</sup> Sub-zero RPLC reduced back-exchange and improved dynamic range in amide hydrogen/deuterium exchange experiments.<sup>185</sup>

Cutting-edge separation technology has begun to further impact the capabilities of shotgun proteomics, including the use of ultrahigh pressure liquid chromatography (UPLC) developed by Jorgenson.<sup>186</sup> Reductions in chromatographic resin particle size, and thus peptide peak width, have increased peptide identification efficiency, sensitivity, and reproducibility.<sup>187</sup> A capillary column frit commonly used in proteomics was designed specifically for UPLC analysis.<sup>188</sup> A detailed study of synthetic peptides illustrated the capabilities of UPLC to separate peptide isomers and conformers.<sup>189</sup> Coupled with improvements to mass spectrometer speed, UPLC separation efficiency has facilitated identification of 78% of the validated yeast proteome with a single 4 hr separation dimension and six replicate runs<sup>179</sup> and over 1400 human proteins in less than a half hour.<sup>190</sup> Superficially porous HPLC particles have also been shown to improve separation efficiency, but have yet to be tested on complex peptide mixtures with MS/MS.<sup>191–195</sup> Monolithic columns have provided separation benefits without the backpressure restrictions from smaller particles. As a result, extremely long columns (2–4 m) running an ~8 hr separation facilitated the identification of 22,196/2,602 peptides/proteins from *E. coli*, 41,319/5,970 from HeLa cells, and an astounding 98,977/9,510 from human induced pluripotent stem cells.<sup>196–198</sup>

The aforementioned peptide separations have relied almost exclusively on RPLC prior to ESI, but two alternative analytical separations provide different peptide retention mechanisms with ESI-compatible buffers. HILIC separates peptides based on their hydrophilic interactions with an ionic resin and has found most application in peptide fractionation and PTM analysis.<sup>199</sup> An organic to aqueous gradient, inverse to RPLC, generally inverts peptide retention order. HILIC-ESI-MS of peptides has been performed with packed<sup>200</sup> and monolithic<sup>201</sup> columns, but has yet to be exploited for shotgun proteomic analysis. ERLIC is

a specific form of HILIC, using a weak anion exchange (WAX) resin. Unlike RPLC, peptides are retained under two separation modes. Early in the organic to aqueous gradient hydrophilic interactions dominate, as in HILIC and inversely to RPLC. However, as the aqueous content of the elution buffer is increased, basic peptides electrostatically repel the WAX resin while acidic peptides are retained until their hydrophilic interaction with the WAX resin is disrupted late in the gradient. These superimposed separation mechanisms with ERLIC distribute peptides over the gradient better than RPLC and outperform it based on peptide and protein identifications by higher confidence spectral matching of larger peptides.<sup>202</sup> Capillary electrophoresis has also reemerged as a complementary, more sensitive, and viable option in shotgun proteomics, largely due to improvements in electro spray interfaces.<sup>203–209</sup>

Fractionation of peptides prior to nLC-ESI to improve comprehensiveness was initially performed online with SCX resin,<sup>8,10,175,210</sup> minimizing sample losses from transfers intrinsic to offline fractionation and auto samplers. Addition of WAX resin with the SCX resin in a mixed-bed format exploited the Donnan effect and increased peptide recovery, reduced carryover between salt elutions, improved separation orthogonally, and led to identification of twice as many peptides and phosphopeptides in separate analyses.<sup>211</sup> The use of WAX alone with a pH-based elution also improved reproducibility and reduced carryover between fractions.<sup>212</sup> Other online systems have performed a RPLC separation at high pH prior to the low pH nLC-MS dimension to improve peak capacity in the first fractionation dimension.<sup>213–214</sup> Offline fractionation has allowed for further optimizations, the use of MS-incompatible buffers, and the combination of separations that can't be directly interfaced. Offline fractionation of peptides by SCX with UV detection allowed for optimization of sample loading for each subsequent nLC-MS run.<sup>215</sup> Preparative electrophoretic methods that were first applied at the protein level, have also been used for fractionating peptides such as free-flow electrophoresis<sup>216</sup> and IPG strips in a traditional<sup>217</sup> and OGE setup.<sup>134,218</sup> Many new fractionation methodologies have the benefit of identifying PTMs in a less biased manner than specific enrichment methods and may facilitate global co-identification and co-analysis of multiple relevant biological PTMs. In combination with Lys-N protease digestion, SCX fractionation has become possible to identify multiple PTMs simultaneously from complex mixtures including acetylated N-terminal peptides; singly phosphorylated peptides containing a single basic (Lys) residue; peptides containing a single basic (Lys) residue; and peptides containing more than one basic residue.<sup>219</sup> The use of zwitterionic (ZIC)-HILIC for prefractionation of peptides at median pH was found as a complementary alternative to SCX and at low pH, also a method to fractionate peptides into classes of PTMs with the more common proteases trypsin and Lys-C.<sup>220</sup> Continued work with ZIC-HILIC resins is improving sensitivity while maintaining comprehensiveness of analysis, identifying 20,000 peptides and 3,500 proteins from 1.5  $\mu\text{g}$  of HeLa cell lysate.<sup>221</sup> ERLIC fractionation of peptides identified more proteins and peptides, particularly of basic and hydrophobic character, than SCX<sup>222–223</sup> and this strategy facilitated simultaneous analysis of phosphorylation and glycosylation from rat kidney.<sup>224</sup> Inverting RPLC and ERLIC separation order has allowed for assessment of the extent of asparagine deamidation on peptides.<sup>225</sup> A similar separation mode, hydrophilic SAX, was also found to be highly orthogonal to RPLC.<sup>226</sup> Further combination of fractionation separations in three dimensional configurations has continued to improve comprehensiveness and sensitivity.<sup>227–230</sup> The use of an online three-dimensional separation of peptides has allowed the identification of proteins present at approximately 50 copies per cell from analysis of the entire yeast proteome.<sup>231</sup> This sensitivity gain from an added dimension of fractionation correlated to an estimated detection limit of 65 amol. Further improvements to prefractionation and enrichment of post-translationally modified peptides will undoubtedly improve identification comprehensiveness and lead to new biological discoveries.

## 2.6 Mass spectrometers and peptide MS analysis

In recent years, peptide analysis has driven the technological advances of new mass spectrometers and techniques. In particular, multiple hybrid instruments have been developed with different mass analyzers, ion optics, and fragmentation sources. These combinations have improved the accuracy of peptide precursor and fragment ion mass measurements and created more informative and complementary peptide fragment ions through various fragmentation methods. Common mass analyzers that have proven useful for peptide analysis from complex mixtures are the LIT, Orbitrap, FT-ICR, quadrupole (Q), and ToF. Table 2 describes their commonly-achieved analytical metrics for proteomics. Since each analyzer isolates and measures peptides masses using different mechanisms, each mass spectrometer represents a balance between sensitivity, speed, and accuracy. Most current mass spectrometers used for proteomics employ a LIT, which allows for both isolation and fragmentation of peptide ions. Data-dependent acquisition allows for primarily unbiased sampling and identification of peptides within a sample, and is performed as follows. Peptide ions are trapped within the LIT and subsequently scanned by increasing the radiofrequency voltage applied to the trap. An initial precursor scan is performed to identify abundant peptide precursor ion  $m/z$ 's. Sequential trapping and isolation of each individual abundant precursor ion is done by filling the trap and ejecting all but a population of ions within a mass window ( $\sim 3$   $m/z$ ) containing the peptide precursor  $m/z$ . The isolated ions are translationally excited prior to collisions. The collisions lead to conversion of translational energy to internal (vibrational) energy and then fragmentation. Fragment ions are scanned out to the mass analyzer and the process is repeated until abundant peptide precursor ions have been sampled. Often, a signal-to-noise threshold or maximum number of fragmentation scans is used to define how frequently the precursor ion and fragmentation scans are performed.<sup>232</sup> Benefits in sensitivity have been achieved through sampling lower abundance ions through data-independent acquisition of consecutive small (10–25  $m/z$ ) windows.<sup>233–236</sup> The modification of the initial 3-D QIT to a 2-D LIT improved sampling speed 2-fold,<sup>237–238</sup> and a dual-pressure 2-D LIT configuration further improved sampling speed and sensitivity.<sup>239</sup> Interfacing the 2-D LIT to FT-ICR<sup>240</sup> and, more recently, to the Orbitrap<sup>241</sup> improved the accuracy of peptide precursor measurements and the confidence in both identification and quantification of proteins and post-translational modifications. FT-ICR instruments still have the highest mass accuracy, but the Orbitrap is faster, more sensitive, and improving in resolution, particularly with its recent reduction in size.<sup>242</sup> The use of identified peptides,<sup>243–247</sup> external lock mass standards,<sup>248</sup> background ions,<sup>249–251</sup> and fragment ions<sup>252</sup> for online<sup>253</sup> and offline<sup>254</sup> mass calibration has further improved mass accuracy over the course of proteomic runs where drift can otherwise occur. Other routine and emerging options for performing data-dependent acquisition of peptides using hybrid instruments are the Q-ToF<sup>255–257</sup> and Q-Orbitrap.<sup>258–259</sup> Schematic diagrams of the common mass spectrometers used for shotgun proteomics are illustrated in Figure 2 and Figure 3. Representative MS data is shown in Figure 10a. Application of ion mobility configurations at the inlet of mass spectrometers has reduced background noise and improved sensitivity,<sup>260–261</sup> increased dynamic range,<sup>262–263</sup> improved proteome coverage,<sup>264</sup> and allowed for differential identification of isomeric peptides,<sup>265–266</sup> phosphopeptides,<sup>267–268</sup> and glycopeptides.<sup>269</sup> The use of ion mobility is beneficial since it provides a gas-phase separation of peptides and chemical noise. Separation is achieved through the differential mobility of ions based on charge, size, and conformation as they are passed through a carrier buffer gas.

Peptide fragmentation methods with different fundamental mechanisms have been implemented on mass spectrometers used for shotgun proteomics which provide complementary peptide fragment ions, proteome coverage, and quantification accuracy. The instrumental configurations for fragmentation are also indicated in Figure 2 and Figure 3.

The common fragmentation mechanisms employed are illustrated in Figure 4. The most common and robust fragmentation method used for peptide analysis is CID/CAD (collision-induced/activated dissociation). A representative CID MS/MS spectrum is shown in Figure 10b. LIT instruments use resonant-excitation, creating only single fragmentation events as the ions fall out of the excitation frequency upon fragmentation. The most prevalent ions formed are b- and y-type which result from fragmentation at the peptide bond, leaving the positive charge on either the N-terminal or C-terminal fragment, respectively. Beam-type collisional activation entails passing ions through a quadrupole, fragmenting ions until they reach their minimum potential energy, creating mostly y-type ions. Resonant-excitation has been primarily used on LITs and beam-type collisional activation on triple quadrupole (QQQ) instruments. Recently beam-type fragmentation, deemed higher energy C-trap dissociation or higher energy CID (HCD),<sup>270</sup> has also been demonstrated on LIT instruments with<sup>271</sup> and without<sup>272</sup> instrumental modifications and on hybrid instruments with dedicated collision cells.<sup>273</sup> Beam-type collision yields sequence-informative low mass peptide fragment ions and its implementation on hybrid mass spectrometers has most dramatically improved peptide quantification using low mass isobaric tag reporter ions,<sup>274–275</sup> further described in the “Quantification” section. Briefly, reporter ions are unique mass fragments from an amine-reactive, isotope-labeled chemical tag that can be used for relative quantification between protein samples. Prior to HCD, these low mass reporter ions were unreliably observed with CID and a pulsed Q dissociation (PQD) activation strategy improved reporter detection, but hampered protein identifications.<sup>276–277</sup> CID of phosphopeptides often results in preferential loss of the phosphate moiety due to the low critical energy for this fragmentation process.<sup>278</sup> Further collisional activation of peptide fragment ions<sup>279</sup> after phosphate neutral loss in the same or a subsequent fragmentation scan creates more sequence-informative fragmentation ions.<sup>280–282</sup> HCD fragmentation is an efficient alternative for phosphoproteomic studies,<sup>283</sup> yet a comparative study illustrated that the greater acquisition speed of CID provided larger data sets.<sup>284</sup> Like CID, electron capture dissociation (ECD)<sup>285</sup> and electron transfer dissociation (ETD)<sup>286</sup> induce random fragmentation along the peptide backbone, albeit from gas-phase reactions with either thermal electrons or fluoranthene, respectively, creating mostly c- and z-type ions as shown in Figure 4. The benefit of ECD was initially demonstrated for localization of labile  $\gamma$ -CO<sub>2</sub> and SO<sub>3</sub> modifications that were otherwise ejected prior to peptide backbone fragmentation by CID.<sup>287</sup> Early bottom-up proteomics studies with ECD-FTICR-MS automated acquisition of fragmentation<sup>288</sup> and found that larger peptides<sup>289</sup> of higher charge state<sup>290</sup> were more efficiently fragmented and identified than with CID. Sequential CID and ECD in a large-scale phosphoproteomics experiment illustrated CID is better for phosphopeptide identification, ECD is better for phosphorylation localization, and together they provide complementary information and high confidence for infrequently identified phosphopeptides.<sup>291</sup> ECD with FTICR-MS presented great potential for bottom-up proteomics, but was not easily compatible with the commonly used LIT since the thermal electrons quickly lose their energy and cannot be trapped. The use of a focused electron beam has been proposed as a solution,<sup>292</sup> but this problem otherwise spawned ETD, where a LIT-trappable anion is used as a one-electron donor to induce fragmentation by the same non-ergodic pathway as ECD.<sup>286</sup> An optimization study using four reagents illustrated ETD fragmentation can be as fast or faster than CID depending on peptide charge state; azulene was the best reagent due to its higher reactivity, fewer inefficient proton-transfer reactions, and low fouling of ion optic elements from a higher vapor pressure.<sup>293</sup> One of the notable advantages of ETD is the retention of phosphorylation moieties on fragment ions, a shortcoming of CID methods.<sup>294</sup> Exploiting this characteristic of ETD, a comparison of phosphoproteomic results using CID and ETD<sup>295</sup> indicated phosphate rearrangement during CID fragmentation was much less prevalent than anticipated from a targeted study of a small number of synthesized phosphopeptides.<sup>296</sup> On the downside, the efficiency of ETD is charge-state dependent, thus diminishing its success for peptides in low charge states. When

coupled with Lys-N for proteolytic digestion, ETD fragmentation has been used to generate straightforward peptide sequence ladders,<sup>297</sup> which is useful in *de novo* sequencing as discussed in the “Bioinformatics” section. Integration of ETD with Orbitrap mass spectrometers<sup>298–299</sup> and application to large scale proteomics studies has demonstrated complementary identifications of peptides, proteins, and proteomes to CID fragmentation. ETD with supplementary CID fragmentation (ETCaD) of doubly protonated peptide precursor radicals,  $[M+2H]^{2+}$ , was developed to minimize non-dissociative electron transfer (ETnoD) processes.<sup>300</sup> This concept was also reported concurrently along with the charge-reduction of larger peptides with ETD prior to CID (CRCID) for identification of post-translational modifications.<sup>301</sup> From these characterizations, a decision tree-driven approach was demonstrated to select the appropriate fragmentation method based on peptide charge state and *m/z* and improve proteome coverage<sup>302</sup>, and has since been implemented with other fragmentation methods with further improvements.<sup>303–305</sup> Exploiting the anion fragmentation capabilities of ETD facilitated identification of the previously under-sampled acidic proteome.<sup>306</sup> The combination of ETD with infrared (IR)-activated ions outperformed both ETD and CID for higher peptide charge states.<sup>307</sup> The key improvements were reduction of the ETnoD process with preactivation and the creation of primarily odd electron  $z^*$ -type and even electron *c*-type product ions. Implementation of IR fragmentation alone<sup>308</sup> has also shown improved reporter-based relative quantification of peptides.<sup>309</sup> Ultimately, these improvements to fragmentation and identification of larger peptides may make a dramatic contribution to middle-down proteomics methodologies.

### 3 PTM analysis by mass spectrometry

The eukaryotic proteome is much more diverse than the corresponding genome due to two stages of regulations: post-transcription and post-translation. Post-transcriptional regulation generates multiple mRNA splicing patterns, which are subsequently translated into different proteins. The proteins can be further post-translationally modified by covalently adding some chemical units or changing the structures of the amino acids themselves. Therefore, the mapping of PTMs is an important dimension to help describe the whole proteome.

PTMs play key roles in almost all biological processes. Hundreds of PTMs have been found to occur on different amino acids. The typical approach of detecting PTMs using mass spectrometry requires enrichment techniques, due to relatively low PTM levels. PTMs are measured by mass shifts caused by the modification (Table 3). The modified residue site can be further interpreted by its site-specific fragments. In this review, we will focus on the most biologically relevant PTMs, including phosphorylation, ubiquitination and glycosylation.

#### 3.1 Phosphorylation

Addition of a phosphate group to a protein can influence many properties of the protein, including protein folding, activity, interaction with other proteins, and localization or degradation; thus, phosphorylation plays essential roles in regulation of nearly all biological phenomena, including proliferation, differentiation, apoptosis, and cellular communication.<sup>310,311</sup> Defects in regulation of reversible phosphorylation controlled by protein kinases/phosphatases can be the cause of various diseases, including cancer, diabetes, chronic inflammatory diseases, and neurodegenerative diseases.

In order to understand signaling networks in normal and pathogenic mechanisms in various diseases, development of an analytical method for identification of phosphorylated proteins and phosphorylation sites is essential. Phosphorylation is often a sub-stoichiometric process; at a given time point, not all copies of a given protein are present in a phosphorylated state. Therefore, highly sensitive methods for isolation, detection, and quantification of low abundant phosphorylation sites are required. In the past, phosphorylation analysis for

identification and localization of the modified amino acid was conducted primarily by radiolabeling<sup>312</sup> with <sup>32</sup>P combined with capillary electrophoresis, amino acid analysis, or Edman radiosequencing.<sup>313</sup> Recently, with vast and ongoing improvements in sample preparation, separation, enrichment, instrumentation, and data analysis, the field of phosphoproteomics has shown substantial expansion, which has resulted in more extensive and confident identification and quantification of phosphorylation sites. In fact, use of proteomic approaches that employ MS can lead to identification and quantification of thousands of phosphorylation sites from a single sample in the femtomole or even attomole range.<sup>314,315,316</sup> Current methods for study of phosphorylation using LC-MS/MS are reviewed below and illustrated in Figure 5.

### 3.1.1 Enrichment of phosphopeptides—Immobilized Metal Ion Affinity

Chromatography (IMAC) with metal ions including Fe(III), Ga(III), Al(III), and Zr(III), has been widely used for enrichment of phosphopeptides.<sup>317,318–319</sup> A number of factors, including binding, washing, and elution affect the efficiency of the IMAC procedure.<sup>320</sup> In order to eliminate the binding of non-phosphopeptides in IMAC, acidic conditions for loading and washing have been used; phosphopeptides can be then eluted using alkaline conditions.<sup>321</sup> Esterification of peptide carboxyl groups resulted in improved specificity of phosphopeptide enrichment<sup>322</sup> using IMAC and a method using Zr or Ti, which prevents binding of acidic peptides, has been developed.<sup>323</sup>

An alternative to IMAC enrichment of phosphopeptides uses TiO<sub>2</sub>, which has shown a higher capacity and better selectivity for phosphorylated peptides.<sup>324</sup> TiO<sub>2</sub> precolumns, TiO<sub>2</sub>-based HPLC chips, TiO<sub>2</sub> tips,<sup>324, 325</sup> and other metal oxide enrichment methods using ZrO<sub>2</sub><sup>326</sup> or Ga<sub>2</sub>O<sub>3</sub>,<sup>327</sup> have been used for purification of phosphopeptides. To improve binding selectivity of phosphopeptides on TiO<sub>2</sub>, competitive binders, such as 2,5-dihydroxybenzoic acid (DHB) and phthalic acid,<sup>325,328</sup> were used in the buffers during enrichment. To increase recovery of phosphopeptides, multiple rounds of consecutive incubation using fewer TiO<sub>2</sub> beads than the optimum ratio have been suggested.<sup>329</sup> In addition, various methods for elution of trapped phosphopeptides, such as ammonium bicarbonate with 50mM ammonium phosphate (pH 10.5), ammonia solution (pH 10.5–11), or step gradients from pH 8.5 (100mM triethylammonium bicarbonate) to pH 11.5 (3% ammonium hydroxide) have been used.<sup>330,331</sup> A method combining IMAC and TiO<sub>2</sub> (called SIMAC)<sup>332</sup> for simultaneous detection of mono-phosphorylated and multi-phosphorylated peptides has been introduced. Mono-phosphorylated peptides are eluted by acidic buffers from IMAC and multi-phosphorylated peptides by basic elution, resulting in increased sensitivity of detection and reduced sample complexity. Recently, polymer-based metal ion affinity capture (PolyMAC) using polyamidoamine dendrimers multifunctionalized with titanium ions and aldehyde groups has been introduced, showing high selectivity, fast chelation times, and high phosphopeptide recovery compared with current strategies based on solid phase nanoparticles.<sup>333</sup>

Fractionation of protein samples is still required even if methods employing IMAC and TiO<sub>2</sub> are used for enrichment of phosphopeptides. Several chromatographic methods, including SCX,<sup>315</sup> SAX,<sup>334</sup> HILIC,<sup>335</sup> and ERLIC<sup>336</sup> have been used in an effort to increase the efficient identification of phosphorylation sites by prefractionation. The combination of SCX fractionation and TiO<sub>2</sub> enrichment is a popular method. Tryptic peptides at pH 2.7 usually carry a net charge of +2 (N-terminal amino group and C-terminal arginine or lysine side chain); however, the presence of a phosphate group decreases the net charge state by one. Using SCX fractionation at pH 2.7, phosphopeptides can be separated from non-phosphorylated peptides<sup>337</sup> and analysis of these fractions using LC-MS/MS resulted in identification of more than 2,000 phosphopeptides from the nuclear fraction of HeLa cell.<sup>315</sup> An acidic, then highly acidic double SCX fractionation strategy separated and selected for

basic phosphopeptides.<sup>338</sup> pI-difference also allows separation of methylated phosphopeptides from the methylated nonphosphopeptides by in-solution isoelectric focusing<sup>339</sup> and this concept was also applied for IPG strips pH 3–10 as used in the first dimension of 2-DE.<sup>340</sup> When HILIC, is used, non-phosphorylated peptides are eluted in the early fractions followed by elution of peptides with single and multiple phosphorylated sites. The presence of the phosphate group increases the peptide's hydrophilicity, and thus retention time.<sup>341</sup> Exploiting a different separation mechanism with SAX, acidic phosphorylated phosphopeptides can be depleted for increased identification of basophilic kinase substrates.<sup>342</sup>

An excess of  $\text{Ca}^{2+}$  can be used for precipitation of phosphopeptides.<sup>343–344</sup> Calcium phosphorylation precipitation method for identification of phosphoproteome was combined with a subsequent IMAC method in rice embryonic cells, resulting in identification of 242 phosphopeptides representing 125 phosphoproteins.<sup>345</sup>  $\text{Ba}^{2+}$  was also used for precipitation of phosphopeptides and this method combined with MudPIT identified 1,037 phosphopeptides from 250  $\mu\text{g}$  of HeLa cells nuclear extract.<sup>344</sup> Furthermore, coverage of the phosphoproteome was extended with three  $\text{Ba}^{2+}$  ions concentrations. Immunoprecipitation based on anti-phosphotyrosine antibodies can be also used for enrichment of phosphotyrosine proteins/peptides.<sup>346,347</sup> In general, antibodies for phosphoserine and phosphothreonine are less specific than anti-phosphotyrosine antibodies and are not frequently used for enrichment. In addition, the recently repurposed hydroxyapatite (HAP) chromatography resin efficiently enriched phosphopeptides based on the higher affinity of multi-phosphorylated peptides towards HAP surfaces.<sup>348</sup> HAP has allowed for efficient analysis of mass-limited complex samples with single-step microcolumn enrichment coupled to MudPIT.<sup>349</sup>

**3.1.2 MS analysis of phosphopeptides**—Due to the electronegativity of the phosphoryl group, as well as the labile nature of the phosphoester bond, phosphopeptides are less efficiently ionized and fragmented in comparison to unmodified peptides. Therefore, MS analysis of phosphopeptides is more difficult than it is for non-phosphorylated peptides. Results of MS/MS obtained from phosphopeptide analysis under general CID conditions demonstrated loss of labile metaphosphoric acid or phosphoric acid.<sup>350</sup> To improve phosphopeptide identification, selection and analysis of ions with neutral loss can be performed with additional CID for generation of an MS3 spectrum.<sup>351</sup> In addition, treatment with phosphatase can increase confidence in phosphopeptide identification and localization of phosphorylation sites through parallel MS/MS analyses of two samples before and after treatment with phosphatase.<sup>352–353</sup>

Alternatively, use of electron capture dissociation (ECD)<sup>354</sup> and electron transfer dissociation (ETD)<sup>286</sup> can induce backbone fragmentation for generation of c- and z-type ions which retain an intact form of PTM such as a phosphorylation group. ETD is compatible with LTQ-Orbitrap, ion trap, and LTQ-FT-ICR mass spectrometers, and is suitable for LC-MS/MS analysis of complex phosphopeptide mixtures.<sup>355</sup> Because the two fragmentation modes are complementary, combination of CID and ETD/ECD for enhancement of the number of identified phosphopeptides has been suggested.<sup>356</sup>

**3.1.3 Identification of phosphorylation sites**—For identification of phosphorylation sites in phosphopeptides, the identification of the amino acid sequence and mapping of phosphorylation sites must be determined accurately by hand or analysis software. In particular, if a phosphopeptide contains consecutive amino acids of serine, threonine, or tyrosine, precise MS/MS spectra generated by cleavage both N-terminal and C-terminal of the identified phosphorylation site are required. Various algorithms, such as SEQUEST and Mascot, can be used for identification of phosphorylation sites; however, validation of the

assigned phosphorylation sites should be performed manually. Ambiguity in site assignment can occur because fragment ions don't adequately define the site or a peptide that is phosphorylated at different sites within the same sequence could co-elute and co-fragment providing evidence for both sites within the spectrum. Various software programs, such as Debunker<sup>357</sup> and Ascore,<sup>337</sup> have been developed to assist in evaluation and validation of phosphopeptides and phosphorylation site assignments, respectively. In addition, more specialized database resources, such as Phospho.ELM,<sup>358</sup> PHOSIDA,<sup>359</sup> Scansite,<sup>360</sup> KinasePhos,<sup>361</sup> and PPSP<sup>362</sup> can also provide tools for prediction and annotation which aid in identification of potential kinases associated with phosphorylation sites.

### 3.2 Ubiquitination

Ubiquitin is a highly conserved 76 amino acid small protein, which exists in all eukaryotic cells. Ubiquitination is reversible, ATP-dependent, and catalyzed by a ubiquitin E1 (activating) - E2 (conjugating) - E3 (ligating) cascade,<sup>363-365</sup> resulting in a covalent isopeptide bonds between a glycine residue at the carboxy-terminal of ubiquitin and the  $\epsilon$ -side chain of a lysine residue within a substrate protein. Ubiquitination can produce either monoubiquitinated or polyubiquitinated proteins. The latter is formed when one of the seven lysine residues of ubiquitin is linked to the C-terminal glycine of another ubiquitin. The way the ubiquitin molecules are linked plays an important role in the function of the resulting modified protein.<sup>366-367</sup> The most well documented function of ubiquitination is its role in mediating protein degradation. Moreover, ubiquitin and many other ubiquitin-like proteins (UBLs) are involved in various biological processes, including but not limited to cell meiosis, autophagy, DNA repair, immune response and apoptosis.<sup>368-370</sup>

The traditional ubiquitination assay, which is normally based on affinity interaction, is incapable of detecting individual ubiquitinated proteins and sites in a high throughput mode. Current mass spectrometry-based proteomic technologies have greatly improved detection and characterization of ubiquitination due to their high sensitivity and mass accuracy. Like many other PTMs detected by MS, ubiquitination is characterized by a signature mass shift. The carboxy-terminal of a ubiquitin has an amino-acid sequence of  $-R_{72}L_{73}R_{74}G_{75}G_{76}$ , which is normally cleaved after the  $R_{74}$  residue during trypsin digestion. The carboxy end ( $G_{76}$ ) of ubiquitin is covalently attached to a lysine on the substrate and leads to a missed cleavage on the substrate during trypsin digest. As a result, the signature ubiquitinated peptide contains additional  $G_{75}G_{76}$  residues from ubiquitin, which results in a mass shift of 114.043 Da (G-G) as well as a missed K cleavage at the ubiquitination site.

Detection of ubiquitination by mass spectrometry may generate false positives, which should be carefully examined. Although the mass shift of 114.043 Da can be measured precisely and accurately in mass spec analysis, there are other events which cause a mass increase of 114 Da, (some with exactly the same chemical elemental composition as diglycine), making them indistinguishable from ubiquitination. For instance, the residue asparagine (Asn) has a monoisotopic mass of 114.043. When a peptide contains a K-N motif, a missed cleavage between these two residues will result in a ubiquitination-like spectrum, and the database search engine may not be able to distinguish between di-glycine and asparagine. Another potential issue comes from the iodoacetamide (IAM) used in cysteine (Cys) alkylation prior to protein digestion. Carbamidomethylation leads to a mass increase of 57.021 Da on Cys and sometimes lysine.<sup>371</sup> Further, it has also been reported that IAM can introduce a di-acetamidoacetamide covalent adduct to lysine, resulting in the identical chemical composition as diglycine. The identical masses introduce artifacts into the database search algorithms, resulting in false positive ubiquitination site.<sup>372</sup> To avoid artifacts, an alternative alkylating reagent<sup>372</sup> or adjusted incubation temperature<sup>373</sup> can be used to reduce over-alkylation. False ubiquitination identification can also be caused by the isotopologue peaks of several amino acids with mass close to 114 Da. For instance, leucine

and isoleucine both have a monoisotopic mass of 113.084 Da and their m+1 heavy isotopic mass is 114 Da. If the 114 isotope containing peak is considered as the peptide precursor, it may result in a false ubiquitination identification. In this case, a high mass accuracy threshold and either manual or software validation of the monoisotopic mass can be employed to ensure correct assignment.

An enrichment or purification step is normally required prior to mass spectrometry analysis in order to get deeper ubiquitination coverage, since its overall abundance is low in biological samples. Several affinity approaches are widely used to enrich for ubiquitinated proteins. Ubiquitin-substrate conjugates can be purified by either anti-ubiquitin antibodies<sup>374</sup> or His-tagged ubiquitin.<sup>375</sup> High sensitivity and specificity of antibodies are critical for the success of an antibody approach. The His-tagged ubiquitin method is thought to be of higher specificity, because the procedure can be done under protein denaturing conditions. However, a His-tagged approach becomes challenging for some biological samples, such as animal tissues and clinical samples. Because ubiquitination occurs at different sites and at different levels among individual molecules, using an antibody to pull down whole proteins may generate a significant number of unmodified peptides after protein digestion. Recently, a new antibody has been designed to address this issue targeting the di-glycine motif after protein digestion by trypsin.<sup>376</sup> Polyubiquitinated proteins can also be enriched by Ub binding domains (UBDs), which have high affinity for poly Ub chains.<sup>377</sup>

There are many other UBLs which covalently conjugate to protein substrates and generate diverse modifications. These modifications can be determined by detection of a UBL remnant-containing peptides after trypsin digestion, but there are some challenges. First of all, the mature form of some ubiquitin-like modifiers (e.g. RUB1, NEDD8 and ISG15) have the same C-terminal amino acid sequence as ubiquitin (-RGG), which results in an indistinguishable di-glycine signature adduct. To determine the different modifiers, a pre-fractionation step such as immunoaffinity purification or epitope tagged Ub or UBLs can be used.<sup>378–379</sup> However, many other UBLs (SUMO, URM, and HUB1) result in a relatively long amino acid sequence remnant after trypsin cleavage. Such branched peptides generate much more complicated MS/MS spectra after CID fragmentation, and interpretation is difficult. To address this limitation, C-terminal SUMO mutants with different protease digestion specificities have been employed, enabling a rapid and efficient identification of SUMO sites.<sup>380</sup> The potential of this mutation approach is that it can be easily adapted to determining the sites of conjugation for other UBL proteins (not only SUMO) from a diverse range of organisms.

Mass spectrometry can also be used to determine the linkage sites in polyubiquitin chains. Ubiquitin has seven lysine residues, each of which can be linked to the C-terminal glycine of another ubiquitin, forming polyubiquitin. The biological consequence of polyubiquitin can be highly dependent on which lysine is modified by another ubiquitin molecule. For instance, K48-linkage primarily targets proteins for proteasomal degradation, whereas K63-linkage appears to play a role in DNA repair, subcellular localization, and protein-protein interactions. Antibodies against these linkage structures have been developed.<sup>367</sup> Alternatively, the tryptic ubiquitin peptides derived from different linkage types have unique structures and masses, and therefore can be determined by mass spectrometry.<sup>381–382</sup>

### 3.3 Glycosylation

Glycosylation is thought to be one of the most common protein PTMs. A protein is glycosylated by a covalent link of the glycan to either an amide or alcohol in the protein. An N-linked glycan is attached to the amide group of asparagine residue in an amino acid sequence motif of Asn-X-Ser/Thr, where X represents any amino acid other than proline.<sup>383</sup> An O-linked glycosylation links the glycan to the hydroxyl group of serine or threonine.

Many studies have revealed that glycosylation, especially on membrane, secreted, and body fluid proteins, greatly influences many biological functions, including protein folding, protein turnover, and immunity.

Glycan analysis, or glycomics, has traditionally aimed to define only the glycan modification in a glycoprotein. In contrast, glycoproteomics refers to the full characterization of glyco-proteins and glyco-peptides. Glycomics studies screen for glycan structures, but lose the information on localization sites and from what proteins they are derived. This information is lost because analysis of glycans normally requires the release of glycan moieties from the glycosylated proteins. N-linked glycosylation is generally cleaved by an amidase, such as peptide N-glycosidase F (PNGase F), which releases broad spectrum sugars. However, such a universal enzyme is not available for O-linked glycans because they are very heterogeneous, and have many forms of linkage core structures. Therefore O-glycosylation is typically cleaved by chemical methods, such as hydrazinolysis<sup>384</sup> and alkaline  $\beta$ -elimination.<sup>385</sup> The released glycan can be subsequently analyzed by different methods. Typically, the glycan, either derivatized or underivatized, is determined by mass spectrometry or nuclear magnetic resonance (NMR).

Glycosylated peptides and proteins are typically present in low abundance in complex biological samples, and an enrichment step is normally performed in order to improve sensitivity of glycosylation detection. For this purpose, a number of enrichment approaches have been developed, including lectin affinity,<sup>386–387</sup> hydrazide coupling,<sup>388</sup> HILIC,<sup>389–390</sup> and boronic acid affinity.<sup>391–392</sup> Among these strategies, lectin-affinity is the most widely used due to its high specificity. Lectins are carbohydrate binding proteins, which can specifically attach to sugar moieties. There are a number of well characterized lectins that may be used for enrichment. For instance, Concanavalin A (ConA) binds to mannose glycan whereas wheat germ agglutinin (WGA) preferentially recognizes N-acetylglucosamine. Ulex europaeus agglutinin (UEA) and aleuria aurantia lectin (AAL) are specific fucose binding lectins. N-acetylneuraminic acid can be enriched by maackia amurensis leucoagglutinin (MAL) and hemoagglutinin (MAH). Combinations of multiple lectin forms have been used to achieve higher glycosylation coverage. Another frequently used enrichment method uses hydrazide chemistry, in which the carbohydrate is oxidized to a di-aldehyde, which is then covalently coupled to hydrazide bead.

The obvious drawback of glycan analysis is the loss of the information about which protein and residue site glycosylation occurs. Mass spectrometry-based glycoproteomic strategies are able to address this issue, and tend to be high throughput and accurate methods in this application. There are two different ways to analyze glycosylated peptides using mass spectrometry. The first, like most other PTM detection methods using MS, is to measure the intact glyco-peptides with glycan; the amino acid sequence and glycosylation site can be determined by the spectrum interpretation. Obviously, this method is ideal, since both the glycan and peptides are measured simultaneously. However, in practice this method has had limited success on large scale analysis, due to the fact that glycan heterogeneity makes spectral interpretation difficult for glycosylated peptides, and the sugar group makes peptide fragmentation challenging. In addition to conventional CID method for peptide fragmentation, some new ion dissociation methods, such as infrared multiphoton dissociation (IRMPD),<sup>393</sup> electron-capture dissociation (ECD)<sup>394</sup> and electron-transfer disassociation (ETD),<sup>395</sup> have been introduced to generate more fragment ions derived from peptide backbone with no loss of the glycan moiety. A full characterization of both glycopeptide and glycan can be accomplished by using two complementary dissociation strategies. First, CID fragmentation usually cleaves the glycopeptide at glycosidic bonds, providing information predominantly on the glycan structures. Next, ETD fragmentation can be applied, which generates more peptide backbone fragments for peptide sequence

identification. The other strategy is to remove the glycan first by reacting with amidase. Deglycosylated peptides are then subjected to a typical shotgun proteomics analysis. In contrast, this method only identifies which peptides were glycosylated. For this strategy, PNGase F cleaves N-linked glycans from glycoproteins, deaminating asparagine to aspartic acid, resulting in a 1 Da mass increase. If this deamination happens in an  $^{18}\text{O}$  water environment, the heavy oxygen leads to an additional mass shift of 2 Da. As a result, a 3 Da mass increase on asparagine can be used for glycosylation site determination.<sup>396</sup> As the N-linked glycosylation contains the motif sequence of Asn-X-Ser/Thr, a targeted database which only comprises motif containing sequences can increase the sensitivity of glycopeptide identification and reduce the false discovery rate.<sup>397</sup>

### 3.4 Other PTMs

**3.4.1 Acetylation**—Acetylation is a modification that mainly occurs on protein N-terminal or lysine residue. The N-terminal acetylation is a co-translational modification, and the later one is a post-translational modification. The N-terminal residue of a nascent protein, resulting from genetic initiation codon of 'AUG', is often methionine, which can be acetylated. Frequently, the N-terminal methionine is also be removed by methionine amino peptidases, thereafter the remaining N-terminal residue is revealed and acetylated by N-acetyl transferases (NATs). N-terminal acetylation is a very common modification in eukaryotes, and up to 80% of human proteins have been reported to be acetylated.<sup>398–399</sup> The second type of acetylation is introduced as an acetyl group to the 3-amino group of lysine residues. The lysine acetylation is a reversible process which is catalyzed by acetyl-transferases and deacetylases, respectively.

Acetylation tightly regulates diverse protein functions. The earliest and most well-known study of acetylation was focused on histones, which demonstrated a significant correlation between histone acetylation and gene expression.<sup>400</sup> Thereafter many other biological significances have been discovered to be related to acetylation, which include apoptosis,<sup>401</sup> cellular metabolism,<sup>402</sup> protein stability,<sup>398</sup> and neurodegenerative disorders,<sup>403</sup> among others. Moreover, it has been implied greatly that acetylation is cooperative with many other important PTMs, such as phosphorylation, methylation, ubiquitination, sumoylation.<sup>404</sup>

The traditional detection of acetylation involves radioactive labeling and acetyl-lysine antibody. Both techniques are able to measure the presence of acetylation, but incapable of localizing the acetylation site. With MS, the acetylation sites can be captured by its signature mass shift (42 Da), and the modified site can be interpreted from MS/MS as well. MS-based proteomics has largely extended acetylation from histone to non-histone targets. Choudhary *et al.*<sup>405</sup> has identified 3,600 lysine acetylation sites on 1,750 protein, implying a global occurrence of lysine acetylation. Lundby *et al.*<sup>406</sup> further enlarged the acetylation scale to 15,474 modification sites on 4,541 proteins from 16 rat tissues, and discovered that the sequence motifs for lysine acetylation varied between different cellular components.

Similarly to ubiquitination, the acetylation on lysine residues normally results in a missed cleavage site after trypsin digestion, which should be considered carefully in protein database search. Acetylation should be also distinguished from trimethylation cautiously, as both modification lead to quite closed mass shifts (42.0106 vs. 42.0469 Da). High mass accuracy data enables us to address this issue. Compared to lysine acetylation, N-terminal acetylation has not been investigated globally, mainly due to its high diversity, and the lack of specific enrichment method. Furthermore, the mature proteins are usually N-terminally processed, and the initial methionine residue can be removed. Therefore consideration of non-tryptic peptides would be essential to finding N-terminal peptides in a database search.

**3.4.2 Methylation**—Protein methylation is another common modification, in which a methyltransferases catalyze the addition of methyl groups to carbon, nitrogen, sulfur, and oxygen atoms of several amino acids, of which the arginine and lysine methylations are most widely studied. Similarly to histone acetylation, histone methylation is another most important PTM which plays a crucial role in regulation of chromatin structure and transcription.<sup>407</sup> Methylation occurs to varied extents, which result in mono-, di-, and trimethylations. Therefore, the proteomic database search should take all the possibilities into consideration.

With mass spectrometry, Ong *et al.* employed a modified SILAC method, in which the  $^{13}\text{CD}_4$ -methionine was metabolically converted to  $^{13}\text{CD}_4$ -S-adenosyl methionine, and the later one was then served as the solo methyl donor. The obvious advantage of this method is its increased confidence of methylation identification because the heavy methylated peptide and its light counterpart peptide have a signature mass difference.<sup>408</sup> In *Saccharomyces cerevisiae*, 83 lysine and arginine methylation sites and their specific motifs were identified.<sup>409</sup> In another study, Vermeulen *et al.* identified the specific protein binders to trimethyl-lysine modifications on histones.<sup>410</sup> A less frequently studied methylation on isoaspartyl residues was recently found on p53 using mass spectrometry and was correlated to p53 degradation.<sup>411</sup>

**3.4.3 Cysteine redox modification**—The proteins can be modified through redox reactions, and proteomic studies focusing on these modifications are termed as redox proteomics. Redox proteomics has been recently comprehensively reviewed,<sup>412–413</sup> therefore this review will only briefly cover this field. Among all amino acids, the most oxidation susceptible residue is cysteine as it contains active thiol. The thiol of cysteine can be oxidized to a variety of forms, including disulfide bond between two cysteine molecules from either different proteins or within the same protein, protein sulphenic acid (PSOH), sulphinic acid (PSO<sub>2</sub>H), sulphonic acid (PSO<sub>3</sub>H), S-glutathionylation, and S-nitrosylation.

Like many other PTM analyses, the low abundance of cysteine modifications makes the detection challenging for complex samples. In addition, some of the redox modifications are quite labile, therefore are not directly compatible with sample preparation and MS analysis. Some cysteine modifications caused by small thiol molecules, such as glutathione and cysteine, are more readily to be detected, which have mass shifts of 305 Da and 119 Da compared to their unmodified counterparts, respectively. These mass signatures can be measured and the modification site can be localized as well in MS analysis. However, for disulfide bond linked peptides, the detection can be more complicated. The first method is to measure disulfide bond linked peptides by successive usages of two different alkylating reagents with different molecular mass. For instant, iodoacetamide can be used to initially block the free thiols, which lead to a mass shift of 57 Da. Subsequently the disulfide bonds are reduced, and the newly opened thiols are further alkylated with N-ethylmaleimide, which results in a mass increase of 125 Da. The different mass shifts can be used to determine which cysteine residues were previously free or blocked. The limits of this method are that it is not able to distinguish between different disulfide types, and there is no ways to know which two cysteine residues are linked. Alternatively, the disulfide bond linked peptides can be analyzed directly by MS. In this case, more effort is required for data interpretation as the fragment ions are derived from multiple peptides, which can be either inter-protein or intra-protein.

Protein S-nitrosylation is a labile reversible modification, in which the thiols of cysteine react with nitric oxide, forming the S-nitrosothiol. The S-nitrosylation has been reported to regulate several diseases, including neurodegenerative diseases,<sup>414</sup> diabetes,<sup>415</sup> cancer,<sup>416</sup> and cardiovascular diseases.<sup>417</sup> Because of the labile nature of the S-nitrosylation, the direct

analysis of it with MS is normally less successful. Alternatively, a biotin switch strategy<sup>418</sup> was introduced to measure the nitrosothiols indirectly. In a biotin switch experiment, the free thiols of cysteine residues are first specifically blocked by methyl methanethiosulfonate (MMTS). Next the S-nitrosothiols are reduced by ascorbate, forming thiol, which reacts with a sulfhydryl-specific biotinylating reagent, N-[6-(biotinamido)hexyl]-3'-(2'-pyridyldithio)propionamide (biotin-HPDP). The S-biotinylated proteins or peptide can be further enriched, and subjected to shotgun proteomic analysis. Similarly to biotin method, enrichment of S-nitrosothiols (SNOs), using resin-assisted capture (SNO-RAC) was introduced,<sup>419</sup> providing a more efficient solution.

## 4 Quantitative proteomics

With advances in sample preparation, protein/peptide fractionation, sensitivity and accuracy in modern mass spectrometry, the proteome map of a certain organisms can now be routinely obtained in reasonable depth using shotgun methods.<sup>176–178,420</sup> Two recent studies have been able to characterize more than 10,000 proteins in samples from cultured cell lines.<sup>421–422</sup> Although proteomics was largely aimed at qualitative analysis early on (i.e. generating a protein list from a given organism), it has become apparent that the protein list itself is insufficient for addressing many biological questions. Abundance changes of proteins are principally a reflection of biological processes or disease states. Altered protein levels can be clues for possible drug targets and also potential clinical biomarkers for disease diagnosis, even at an early clinical phase. As a result, new methods for quantitative measurements on both protein expression and PTMs have been developed and have become an integral part of current proteomic studies.

The classic proteomic platform, 2-DE, was the first tool used for proteome quantification. Quantification is performed by comparing the staining densities of proteins on two or more gels, providing a measure of relative quantification. However, 2-DE has its limitations. First of all, the 2-DE gel has a limited resolution for large number of proteins which comprise the whole proteome. The co-appearance of protein spots makes accurate quantification impossible. The incompatibility of 2-DE with hydrophobic proteins means it cannot be easily used for membrane proteins. Additionally, 2-DE is incapable of analyzing low abundant proteins in a high dynamic range sample such as plasma, whose dynamic range of protein expression can vary by up to 12 orders of magnitude.<sup>423</sup> 2-DE-based quantification detects only extreme differences and inaccurately estimates quantity changes. Furthermore, 2-DE is a labor and time consuming strategy that allows only individual gel spots to be identified one at a time.

Given the limitations of 2-DE-based methods for quantification, several strategies have emerged to quantify the proteome based on the mass spectrometry data. The current MS-based quantitative approaches are shown in Figure 6. Since 2-DE quantification has largely been replaced by MS-based methods, the following discussion will cover only the LC-MS-based quantitative strategies. MS-based quantitative data are obtained by either stable isotope labeling or label-free approaches. The label-free strategy measures samples individually, comparing the MS ion intensity of peptides<sup>424–428</sup> or using the number of acquired spectra<sup>175,429–430</sup> matching a peptide/protein as an indicator for their respective amounts in a given sample. The isotope labeling approach allows the mixing of multiple samples at different experimental stages. The absolute or relative protein abundance can be obtained by measuring the intensities of different isotope coded peaks which are distinguished by mass spectrometry (Table 4). Multiplexing quantification not only makes comparison more straightforward, but also reduces valuable instrument time. There are two methods that are used to derive quantitative information from mass spectrometry data: one is to calculate peptide precursor ion abundance in the MS1 scan; alternatively, the isotope

coded reporter ions (which reflect peptide quantity) can be detached from the peptide by fragmentation and measured in the MS2 scan.

## 4.1 Relative quantification

### 4.1.1 MS1-based quantification by isotope labeling

**4.1.1.1 Chemical reaction labeling:** Initial MS-based quantification methods relied on chemical labeling to add isotope-coded reagents to reactive groups on the side chains of amino acids or to the peptide termini. The first chemical labeling methods used for mass spectrometry-based quantification made use of the reaction between the thiol side chain of cysteine and isotope-coded tags in a method called ICAT (isotope-coded affinity tag).<sup>431–432</sup> The ICAT reagent consists of three different elements: a biotin affinity tag, a thiol specific reactive group, and a linker which contains either light or heavy isotopes. In an ICAT experiment, cysteine residues are selectively derivatized with an ICAT reagent containing either eight  $^1\text{H}$  or eight  $^2\text{H}$  atoms, and the protein mixture is then subjected to avidin affinity chromatography to purify ICAT coded proteins. Thereafter, proteins are digested with trypsin and the peptide mixture is analyzed by LC-MS. Because cysteine is not found in every peptide, the ICAT technique can significantly reduce the complexity of the peptide mixture, thus simplifying analysis of complex samples. On the other hand, ICAT obviously eliminates all non-cysteine-containing peptides and therefore is not suitable for comprehensive large-scale quantification. Furthermore, the deuterium tag results in a retention time shift between light and heavy peptides in reversed-phase chromatography, which complicates subsequent data analysis.<sup>433</sup> The second generation of ICAT has been modified to include a  $^{13}\text{C}$ -labeled linker instead of deuterium.<sup>434–436</sup> A similar isotope-coded protein label (ICPL)<sup>437</sup> method uses either six light  $^{12}\text{C}$  or heavy  $^{13}\text{C}$  in isotope tags attached to free amino groups in proteins.

Another chemical labeling strategy using formaldehyde to introduce dimethyl labels, has been used to label the N-terminus and amino group of Lys residues (Figure 7a).<sup>438–439</sup> In this method, the light form of formaldehyde is reacted with cyanoborohydride to introduce double chemical groups of  $\text{CH}_3$  to each primary amine in a peptide (Figure 7a). The medium and heavy labeled reagents generate two sets of  $\text{CHD}_2$  and  $^{13}\text{CD}_3$ , respectively. As a result, 4 Da of mass difference can be detected between three groups by mass spectrometry (Figure 8a–c). Although dimethyl labeling has been shown as an easy, efficient and inexpensive approach, there are still some restrictions. Like in the original ICAT method, there is a chromatographic retention time shift caused by deuterium labeling. Additionally, in the case where the mass difference is only 4 Da, observed in 3-plex labeling (Figure 7a), the small mass difference may lead to isotopic peak overlapping between two peptides (Figure 8b), especially for larger peptides.

**4.1.1.2  $^{18}\text{O}$  labeling:**  $^{18}\text{O}$ -labeling is an MS1-based stable isotope labeling method in which the isotope is introduced during enzymatic reaction.<sup>440–442</sup> In this method, the C-terminal carboxyl group of proteolytic peptide is labeled with two atoms of  $^{18}\text{O}$  by amide bond cleavage and oxygen exchange in the presence of  $^{18}\text{O}$  water. The labeled and unlabeled peptides are distinguished by a 4 Da mass difference in mass spectrometry.  $^{18}\text{O}$ -labeling is an easy and inexpensive isotope labeling method. However,  $^{18}\text{O}$ -labeling is less widely used than other quantitative methods for a number of reasons. The variable incorporation of  $^{18}\text{O}$  atoms into peptides<sup>443–445</sup> is the primary limitation of the method. As oxygen exchange is not always complete, one labeled peptide may contain both singly labeled and doubly labeled subgroups. Together with unlabeled peptides, there will be a total of three groups of peptides with only a 2 Da mass shift separating them, leading to complicated data analysis. A modified  $^{18}\text{O}$ -labeling approach decoupled from protein digestion has been developed which increases the efficiency of producing doubly labeled peptides.<sup>446</sup> However, even with

all peptides doubly labeled, a mass difference of 4 Da is still relatively small, especially for large peptides with broad isotope peak distributions. Quantitative algorithms for  $^{18}\text{O}$ -labeled data must be mindful of isotopic peak overlapping and variable labeling efficiencies in these samples.<sup>445,447</sup>

**4.1.1.3 Metabolic labeling:** Metabolic labeling of proteins for quantification employs a stable isotope-enriched medium or diet to culture or feed living systems. The isotopic atoms are incorporated into the whole proteome through protein synthesis during protein turnover and cell multiplication. The labeled peptides have a mass increase that can be detected by a mass spectrometer. When labeled and unlabeled samples are combined prior to introduction into the mass spectrometer, the ratio of peak intensities in the mass spectrum reflects the relative abundances of the peptides. Metabolic labeling is considered to have higher quantitative accuracy than other labeling methods, since it allows samples to be mixed at the protein level prior to any sample preparation, avoiding the introduction of sample handling errors and systematic variance. However, unlike chemical or enzymatic isotope labeling, metabolic labeling cannot be applied universally to all samples. Stable isotope metabolic labeling first was introduced by Langen *et al.* in 1998, who used an  $^{15}\text{N}$ - and  $^{13}\text{C}$ -labeling approach to compare protein quantities with 2-DE.<sup>448</sup> Other research groups soon reported the successful use of  $^{15}\text{N}$  metabolic labeling in both yeast<sup>449</sup> and a mammalian cell line.<sup>450</sup> Nitrogen, rather than carbon, is typically chosen as the isotopic marker because the isotopic nitrogen reagent is easier to obtain. Moreover, there are on average four times as many carbon atoms in a protein as nitrogen atoms. As a result, a  $^{13}\text{C}$ -labeled peptide usually results in a broad distribution of the peptide isotopic peak, making data analysis more challenging.  $^{15}\text{N}$ -labeling technology has been applied successfully to cells in culture,<sup>450–451</sup> plants,<sup>452–455</sup> *Drosophila melanogaster*,<sup>456</sup> *Caenorhabditis elegans*,<sup>456–457</sup> and mammals.<sup>458–460</sup>  $^{15}\text{N}$  labeling of mammal is referred to as “stable isotope labeling of amino acids in mammals” (SILAM).<sup>52,461</sup> Another metabolic labeling method, known as “stable isotope labeling with amino acids in cell culture” (SILAC), was introduced in 2002.<sup>462</sup> The major difference between SILAM and SILAC is that SILAM results in the labeling of one atom in all amino acids, whereas SILAC labeling is limited to select amino acids (usually Lys and/or Arg). In a typical SILAC experiment, cells are differentially labeled by growing them in light medium with normal Arg or Lys (e.g. Arg-0 or Lys-0) or labeled medium with heavy Arg or Lys (e.g. Arg-6 or Lys-6). The subsequent trypsin digest cleaves the proteins at arginine and lysine residues. Therefore, every tryptic peptide except for the C-terminal contains one labeled amino acid, which makes the mass increase of the labeled peptide predictable. Originally, SILAC was used only for cell culture. However, the SILAC approach also has been recently applied to mouse labeling<sup>463</sup> by feeding the mice a  $^{13}\text{C}_6$ -lysine-labeled diet for four generations, and *Drosophila melanogaster*<sup>464</sup> by feeding with labeled yeast.

The two main forms of metabolic labeling, SILAM and SILAC, each have strengths and shortcomings. SILAC and SILAM display different isotope patterns for labeled peptides in mass spectra (Figure 9). Since only one (or limited) labeled amino acid can be incorporated in any given tryptic SILAC peptide, the mass difference between the unlabeled and labeled peptide can be predicted, which facilitates data analysis. By contrast, the mass increase of a  $^{15}\text{N}$  labeled peptide depends on its chemical composition as well as on the labeling incorporation rate. This variable mass gain usually makes subsequent data analysis more challenging. However, the variable mass shift caused by the  $^{15}\text{N}$  atom number can also be a potential benefit for larger peptides. Thus the fixed mass difference derived from SILAC, e.g. 6 Da, may not be large enough for complete separation of the isotope patterns of each peptide (Figure 9).

The requirement for labeling efficiency varies between  $^{15}\text{N}$ -SILAM and SILAC. Generally, complete or nearly complete labeling is necessary for a straightforward data processing and reliable quantification. However, the relative isotope abundance (RIA) ordinarily does not reach absolute 100% in either  $^{15}\text{N}$  or SILAC, especially in mammal labeling, because of residual unlabeled atoms in the nutritional source and metabolic amino acid recycling. SILAC peptides appear only in two forms in the MS: labeled and unlabeled. If a protein is only partially labeled, the MS of unlabeled peptide is a mixture composed of two parts, the original unlabeled peptide and the partially labeled peptide (Figure 9d). To avoid this, the labeling incorporation rate of SILAC should be close to 100%. By contrast, as the incorporation rate increases in the SILAM method, the labeled MS peaks of  $^{15}\text{N}$  peptide gradually moves towards higher mass. A relatively low labeling incorporation rate is sufficient to separate the labeled and unlabeled MS peak envelopes of  $^{15}\text{N}$  peptides (Figure 9c). Consequently, a high labeling efficiency for the  $^{15}\text{N}$  quantitative strategy is not as stringently required as for SILAC (Figure 9c and d), which reduces the cost and time required for labeling. Complete labeling is easily achieved for actively dividing cells in cell-based experiments because the medium that contains the labeled amino acids can serve as the whole nutritional source for protein synthesis. Moreover, the metabolic rate of cultured cells is normally very high. Mammals such as mouse and rat are not easily fully labeled because of the relatively slow turnover of both proteins and its precursor in the amino acid pool. A labeling approach starting *in utero* has been used to achieve successful labeling.<sup>52,461</sup>

Although there is no clear biological bias caused by incorporation of heavy isotopes, a label swapping strategy is recommended to avoid such a risk, and other systematic technical bias. In a label swapping experiment, the isotope types for two samples are switched in two replicates.<sup>465</sup> Thus any potential effects caused by incomplete SILAC labeling efficiency, protein contaminations, or isotope itself can be removed. An ideal correlation between two swapping replicates should result in two ratios (e.g. Light/Heavy) which are reciprocal to each other.

Although metabolic labeling may provide the best quantification precision and accuracy, it is still used infrequently in proteomic studies, especially those involving animals. One reason for its limited use is the high cost due to the difficulties in manufacturing isotope enriched reagents and the requirement of large quantities of reagents to label animals. Also, not every organism is amendable to metabolic labeling, particularly human samples, although heavy leucine infusion has been used to track human surfactant protein in newborn infants.<sup>466</sup> An indirect 'spike-in' strategy was introduced to address these issues.<sup>459</sup> Briefly, labeled proteins are used as an internal standard, which is simultaneously mixed with every sample to be compared. The final quantification between samples can be derived from the individual ratios between each sample and the uniform standard. The benefits of using 'spike-in' are obvious. First, although the heavy stable isotopes or isotope enriched protein sources are thought to be a safe replacement for light ones, there may be some unknown biological effects caused by changed isotopes or diets. A labeled internal standard is able to eliminate such potential effects. Secondly, once an organism is labeled, it can be used for many diverse experiments. A labeled mouse can provide researchers with organs (brain, heart, liver etc.) with adequate protein content for many studies and so the cost per experiment for metabolic labeling is not as high as it might appear. For organisms not suitable for metabolic labeling, a labeled internal standard which shares the same genome background and has a considerable proteome overlap with the target sample, can be used for quantification.<sup>467-468</sup> Liao *et al.* used  $^{15}\text{N}$  labeled rat brain as 'spike-in' to quantify the proteome and phosphoproteome in cultured primary neurons.<sup>467</sup> The 'spike-in' method allows comparison of unlimited number of samples in parallel but has some inherent restrictions. The overlapping proteins, especially in 'tissue/cell' experiments, represent a

reduced portion of the whole proteome since the samples originate from different cell types. Because a ratio of ratios (ratio/ratio) is used for quantification, only the proteins found in every experiment can be used to draw the final quantitative map. Furthermore, the statistical deviations in a 'spike-in' experiment are typically higher than those in a direct comparison.

**4.1.2 MS2-based quantification by isotope labeling**—Isobaric tags for relative and absolute quantification (iTRAQ<sup>®</sup>)<sup>469</sup> and tandem mass tags (TMT<sup>®</sup>)<sup>470</sup> consist of several functional elements: a unique mass reporter, a cleavable linker as a mass balancer, and an amine-reactive group (Figure 7b). All labeling subdivisions in a set of multiplex tags have the same total mass weight (Figure 7b). Importantly, these labeling elements facilitate co-elution of peptides during LC and co-isolation for fragmentation during MS/MS. Peptide mixtures representing different biological samples are labeled with different forms of the tag by reacting the amine-reactive group of the tag with the peptide N-termini and lysine residues of the peptide. During peptide fragmentation in the mass spectrometer the linker is fragmented, releasing the mass reporter, where the intensity represents the relative abundance of the original peptide (Figure 8d–g). There are several features of isobaric labeling which have led to its widespread use in proteomics. For instance, an isobaric labeling strategy is capable of analyzing multiple labeled pools of peptides, up to 6-plex for TMT<sup>471</sup> and 8-plex<sup>472</sup> for iTRAQ, in a single analysis, reducing analytical time significantly. In contrast, all other isotope labeling approaches are limited to 2 or 3 isotope comparisons. Another benefit of using an isobaric label is that it does not increase the complexity of the MS1 scan (Figure 8e) and does not decrease the precursor signal sensitivity (as in SILAC and SILAM), since all the tags lead to the same mass increase for each labeled peptide. However, isobaric reagents are undetectable in conventional CID in ion trap instruments because these low-mass reporters are ejected from the ion trap during activation.<sup>473</sup> To circumvent this issue, a modified activation method, pulsed q dissociation (PQD)<sup>474</sup> can be used which enables MS/MS reporter ions to be detected in an ion trap mass analyzer. Higher-energy collisional dissociation (HCD)<sup>273</sup> has also been routinely used in iTRAQ and TMT analysis.<sup>275,475</sup> HCD provides better MS2 data quality, because all the MS2 data are acquired with high mass accuracy, but it requires more ions and time to generate a MS2 spectrum compared with that from normal CID. As a result, the HCD-based quantification normally leads to lower proteome coverage. Although isobaric labeling has been shown to be a powerful tool for quantitative proteomics, it suffers from a lack of quantification accuracy,<sup>476–477</sup> which is largely influenced by co-fragmentation of co-eluting peptides with similar m/z. When the MS selects a peptide precursor for subsequent fragmentation, it isolates the precursor in a wide m/z window, which may cover both target peptide and the co-eluting peptides. In this case, the reporter ions do not originate from a pure peptide population, but a mixture. Thus if the co-fragmented peptides do not have the same quantitative ratios, it will result in inaccurate quantifications. Recently, two independent studies have sought to overcome this limitation.<sup>478–479</sup> Ting *et al.*<sup>478</sup> employed an additional stage of mass spectrometry to improve the ion selection specificity. In this case, the most intense MS2 fragment ion was selected for a further fragmentation, and the reporter ions released during the second fragmentation were used for quantification. To avoid selection of non-labeled fragments in MS2, the researchers used protease Lys-C, not trypsin, to digest the proteins. The resulting peptides all contained Lys and were labeled with isobaric tags at both termini. Thus all the fragment ions in MS2 were tagged and able to release reporter ions in the MS3 analysis. Wenger *et al.*<sup>479</sup> reduced the charge of peptides by using proton-transfer ion-ion reactions, generating more purified precursors for further fragmentation. For instance, if a doubly charged peptide and a triply charged peptide have the same m/z, using the Wenger method both charges can be reduced by one, generating singly charged and doubly charged peptides respectively. These two newly charged peptides will now have different m/z, avoiding precursor interference. Both strategies show promise

as good solutions for isobaric labeling, but an increase in the number of steps in the procedures may lead to fewer peptide measurements and identifications over the same period of analysis time.

Similarly, another potential weakness of the isobaric strategy, which has not been widely recognized, is the bias caused by using a single intensity measurement. In a shotgun proteomic experiment, a peptide is usually triggered for fragmentation only once (or limited times) due to the 'dynamic exclusion' function. Therefore the reporter ions used for quantification only reflect the peptide abundances at that given time. In contrast, MS1-based quantification normally measures the whole peptide elution profile, which reduces relative variability, especially for low abundance peptides.

**4.1.3 Label free quantification**—As has been described, use of a labeling method that employs stable isotopes allows for accurate quantification of protein expression. However, certain limitations, including additional sample processing steps, cost of labeling reagents, inefficient labeling, difficulty in analysis of low abundance peptides, and limitation of sample number, are associated with use of these labeling techniques. As an alternative, label-free methods can be used to avoid the drawbacks of labeling methods. Quantification of protein expression using a label-free methods can be achieved by two methods: 1) spectral counting uses the frequency of peptide identification of a particular protein as a measure of relative abundance and 2) ion intensity uses the mass spectrometric chromatographic signal intensity of peptide peaks belonging to a particular protein.<sup>480</sup>

In spectral counting<sup>175, 429, 430</sup> the frequency of peptide spectral matches of a specific protein can be correlated with the amount of a protein. Use of the spectral counting method can be applied to low to moderate mass resolution (0.1–1 Da) LC-MS data. Spectral counting has shown a linear correlation and good reproducibility over a large dynamic range for determination of relative protein abundance.<sup>481,482</sup> In addition, the spectral count approach makes use of simpler normalization and statistical analysis than does the ion intensity method. A protein abundance index (PAI), defined as the number of identified peptides divided by the number of theoretically observable peptides for each protein, is used in label-free quantification to calculate the abundance of a protein.<sup>483</sup> This has been further converted to exponentially modified PAI (emPAI), which has greater proportionality to the protein abundance in a sample.<sup>484</sup> The emPAI calculation can be used for database search platforms such as SEQUEST and Mascot. One of the drawbacks of spectral counting is the inherency of peptides with different physicochemical properties to produce bias in MS result. To improve this disadvantage, a modified spectral counting method, called the absolute protein expression (APEX) was introduced for measurement of protein concentration per cell, based on the number of observed peptides for a protein and the probability of detection of peptides by MS.<sup>485</sup> APEX is a useful tool for large-scale analysis of proteomic data, as shown by estimation of cellular protein concentrations of human pathogen *Leptospira interrogans*.<sup>486</sup>

The intensity based approach for quantitative label free MS involves integration of all chromatographic peak areas of a given protein; the area under the curve (AUC) directly correlates with the concentration of peptides in the range of 10 fmol – 100 pmol.<sup>424,425</sup> Thus, the process for quantification of proteins based on AUC by measurement of ion abundance at specific retention times for given ionized peptides, called the ion count, is useful for quantification within the given detection limits of the instrument.<sup>487</sup> Although the basic idea behind this method is rather straightforward, various factors to ensure reproducibility and accuracy must be taken into account. Factors that need be taken into consideration include co-eluting peptides, particularly when peptide signals are spread over a large retention time, multiple signals for the same peptide due to technical or biological

variation in retention time, mass spectrometer speed and sensitivity, and background noise due to chemical interference. Computational methods that consider mass accuracy and alignment of retention times of peptides across various sets, background noise, and peak abundance normalization have been used to address these issues.<sup>488</sup> Finding the balance between MS and MS/MS frequency for quantification of proteins using AUC methods, while at the same time maximizing protein identifications, is imperative. With the advent of high mass accuracy mass spectrometers, this can be accomplished by performing multiple analyses of the sample separately in MS and MS/MS modes,<sup>427</sup> matching accurate mass and retention times for identification of peptides, and integrating the ion intensities of corresponding peptides to calculate the protein concentration.<sup>489</sup> A combination of both spectral counting- and ion intensity-based label-free quantification further improved accuracy.<sup>490</sup>

As described in the “Mass spectrometers and peptide MS analysis” section, data-dependent analysis (DDA), where the mass spectrometer scans parent ions and selects the most abundant ions for fragmentation, is used for processing of most MS analysis. Using this method, low abundance peptides are often omitted, resulting in a low dynamic range of detection. Data-independent analysis (DIA) acquires data based on the sequential isolation and fragmentation of specific precursor windows. Using this method, signal to noise ratio showed 3–5 fold improvement, and peptides unidentified by DDA methods were detected.<sup>233</sup> Use of an advanced method, called XDIA, with increased spectra and peptide identification using ETD and CID, was reported to improve confidence in quantification statistics and protein coverage.<sup>491</sup> Recently, a DIA method called SWATH was developed for quantitative and qualitative detection of all proteins and peptides in a single analysis using fast, high resolution Q-ToF.<sup>236</sup>

Despite the fact that there are always new developments in software programs and technical aspects of analytical instruments, high quality results that are both sensitive and specific can be generated only if the software is compatible with the specified instrument. Therefore, one of the obvious challenges for quantitative proteomics is choosing the most appropriate analysis software, considering file compatibility, user friendliness, computational requirements, data visualization, and automation. Some software tools for label-free quantification have been described below. SIEVE, developed by ThermoFisher Scientific, makes use of the base peak chromatogram to extract data and align peptide signals across LC-MS runs with statistical validation for determination of changes in peptide abundance. Other software including QuanLynx (Waters), Elucidator (Rosetta Biosoftware), and Expressionist (Genedata) are also commercially available.<sup>492</sup> ProteinQuant detects the first occurrence of a peptide MS/MS scan and notes its corresponding retention time in the master file.<sup>493</sup> Within the retention time window associated with the precursor MS/MS scan time, it generates a base peak chromatogram for precursor m/z and selects the maximum intensity ions. IDEAL-Q is another automated tool for use in label free quantification.<sup>494</sup> This tool accepts raw data in mzXML format and peptide and protein identifications from search engines such as SEQUEST,<sup>495</sup> Mascot,<sup>496</sup> or X!Tandem.<sup>497</sup> IDEAL-Q uses a fragmental regression algorithm to find unidentified peptides in a specific LC-MS/MS run through a predicted retention time from the same identified peptide in other runs. The predicted elution time is used to detect the precursor ion isotope cluster of the assigned peptide for quantification after noise filtering. SuperHirn<sup>498</sup> and PEPPEr<sup>499</sup> software make use of MS/MS scan in addition to retention time and m/z, which significantly improves alignment accuracy, but is still limited to high resolution data. SuperHirn uses an alternative alignment strategy and intensity correlation to detect outliers with significantly lower similarity values between LC-MS runs; therefore, low quality runs for the alignment process are excluded.<sup>492,498</sup> Census, which is compatible with label and label-free analyses, reads the spectral file and constructs the chromatogram for each peptide generating a XML file,

which is further loaded into a Census graphic user interface application. Results are exported after consideration of various peptide parameters, including regression ratio, regression score, determinant factor, data points, and outlier peptides.<sup>500</sup> A new developed software called IdentiQuantXL was compared with various software packages, including SIEVE, Msight, PEPPER, ProteinQuant, and IDEAL-Q, based on key features, such as elution patterns, global or individual alignment, pattern- or identity-based, and multiple filters using various different samples, i.e., standard peptides, kidney tissue lysates, HT29-MTX cell lysates, and serum.<sup>501</sup> Recently, Skyline, open source software tool for quantitative data processing from targeted proteomics experiments performed using SRM,<sup>502</sup> expanded its capabilities to process label-free quantification of proteomics data using Skyline MS1 filtering.<sup>503</sup> This software provides key features including visual inspection of peak picking and both automated and manual integration, and can be used with mass spectrometers from major companies.

Methods for label-free quantification in high throughput shotgun proteomics have come a long way in recent years. Each method has advantages and disadvantages and it is difficult to say that one method is more accurate than the others. With the advent of a large number of fast, accurate, and sensitive instruments and software program for validation, label-free quantification is becoming a popular alternative to the use of labeling methods.

#### 4.2 Absolute quantification

The relative quantification methods described above are currently the standard methods used in the quantitative proteomics field, providing valuable information about altered protein expression. However, absolute quantification is still a much sought after goal in proteomics, as it allows us to have a more precise and comparable definition of the proteins and their PTMs within a biological system, which could include detailed information on cellular organization, protein quantities and dynamics, and protein stoichiometries within complexes, among other biological phenomena.

The basic principle of current absolute quantification methods is to introduce stable isotope labeled standards whose abundances are predefined. This method is equivalent to 'isotope dilution analysis', which was initially used in small molecule determination.<sup>504</sup> As discussed previously, metabolically labeled proteins can also be 'spiked-in' to serve as internal standards in proteomics experiments. For relative quantification, metabolically labeled standards usually contain the whole proteome, but protein abundance for individual proteins is unknown. The 'spike-in' methods used for absolute quantification employ standards with limited number of peptides and known concentration. Labeled standards can be generated in three ways: chemically synthesized stable isotope labeled peptides (AQUA),<sup>505</sup> biologically synthesized artificial quantification concatemer (QconCAT) peptides<sup>506</sup> and intact isotope labeled protein standard absolute quantification (PSAQ) or absolute SILAC.<sup>507-509</sup>

The AQUA method relies on the relative comparison between an internal isotope-labeled standard peptide and unlabeled sample digest, in which a chemically identical peptide is expected. The labeled peptide can be mixed with the sample during protein digestion or right before the LC-MS analysis. Analysis of paired peptides in selective mode (SRM or MRM)<sup>510</sup> enables the MS to monitor both the intact peptide mass and one or more specific fragment/transition ions of that peptide. In combination with the retention time, the AQUA platform can eliminate ambiguities in peptide assignments and extend the quantification dynamic range. The preferred instrument for SRM/MRM analysis is the triple quadrupole mass spectrometer, since the three space tandem quadrupole analyzers can provide the best performance in precursor selection, fragmentation, and transition ion scan. In addition, the AQUA method is suitable for PTM determination.<sup>505,511-512</sup> A synthesized peptide with modified residue and heavy isotope label is used as the internal standard. An unmodified

heavy peptide standard can also be added to extract quantitative information on both the total protein and modified protein levels. A disadvantage of AQUA is that this method is not fully compatible with protein fractionation. A target protein subjected to sample fractionation normally has incomplete recovery, leading to an increased error in the quantification.<sup>513</sup> The selection of the optimal peptide standard and the amount of the standard to be added to the sample are very important criteria for the AQUA method. A survey run is always recommended to determine which peptides are well ionized and fragmented and if there are any interfering ions. The selected peptides are typically fully tryptic, without missed cleavage, and not excessively long. Moreover, peptides with reactive amino acids and peptides shared by more than one protein are avoided. Unlike the other relatively global quantitative approaches, AQUA does not measure protein quantity in a high throughput manner but determines one or a few target peptides of interest.

QconCAT is an AQUA-like strategy which enables absolute quantification on a larger scale. Briefly, an artificial protein formed from concatenation of tryptic peptides can be produced by either *in vivo* or *in vitro* methods from a single artificial gene, and then purified by its His6x-tag. The QconCAT polypeptide is mixed with the biological samples prior to proteolysis. After trypsin digestion, all released peptides can act as internal standards for multiple proteins. QconCAT can now include roughly 50 peptide standards,<sup>514</sup> which may correspond to 20–30 proteins. Similar to AQUA, a well-designed QconCAT construct is necessary for the final success of quantification. The design of QconCAT involves many aspects, which have been reviewed by Brownridge *et al.*<sup>514</sup> The major advantage of QconCAT is its ability to monitor several proteins simultaneously, which is very useful in protein complex and pathway studies. However, like AQUA, QconCAT is not compatible with protein fractionation, because the spiked-in standard is an artificial protein different than any of the target proteins.

The PSAQ method possesses some features of both AQUA and QconCAT.<sup>507</sup> Like AQUA, a PSAQ standard targets one protein. However, PSAQ employs an intact stable isotope labeled protein (which can be expressed *in vivo* or *in vitro*) rather than one peptide for use as the internal standard, therefore it provides multiple peptide standards for a given protein, with higher sequence coverage and better statistical reliability. Another strength of PSAQ is that the standard and sample can be mixed at the beginning stage of the sample preparation, making it compatible with prefractionation and resulting in enhanced sensitivity and the release of peptides during proteolysis will mimic the rate of release for the target protein.

Both QconCAT and PSAQ can provide higher quantitative efficiency than AQUA, as the former two allow quantifying multiple peptides by spiking in one standard. For QconCAT and PSAQ strategies, an accurate measurement of the standard quantity is important for subsequent quantification of peptides. Cysteine or amino acid analysis have been used for this purpose,<sup>506–507</sup> where the protein is first hydrolyzed and then quantified by comparison with amino acid standards. A drawback of QconCAT and PSAQ is they are not easily used to measure modified peptides, whereas AQUA is more adaptable to synthesis of peptides with specific modifications.

Besides the isotope labeled standard methods discussed above, there is another type of absolute quantification which is less used, but potentially useful. This strategy employs synthetic unlabeled peptides which may be labeled with isotope reagent, such as iTRAQ<sup>54</sup> or <sup>18</sup>O water.<sup>515</sup> The synthesized peptide and the endogenous peptide sample, including its counterpart, are labeled with different isotopes by chemical or enzymatic reactions. The advantage of this method is the use of unlabeled synthesized peptides and readily available isotopic labeling reagents.

Absolute quantification has shown promise in targeted proteomics. However, currently it can only be carried out on a limited number of target proteins and peptides, mainly due to the need for a synthetic internal peptide standard. Moreover, the degree to which the measured quantity accurately reflects the real sample quantity, especially after the time-consuming sample preparation procedures and physical/chemical treatment of the sample, is still under debate. The reliability of the quantification can be greatly affected by protein degradation and the completeness of protein digestion. Therefore it is possible that the absolute protein concentration derived from these strategies does not accurately reflect the real protein level in the biological samples. Measurement of larger peptides with middle-down proteomics or of intact proteins with top-down may help to alleviate these challenges.

### 4.3 Multiple comparisons

From ground-breaking development and automation of instrumentation, analytical methodologies, and bioinformatics pipelines, the most common proteomics experiment has quickly shifted from methodological to biological. Global quantitative biological comparisons are now possible, particularly with increasingly parallelized isotopic peptide labeling strategies,<sup>516</sup> which has increased the need for appropriate statistical consideration of protein expression changes with multiple testing correction.<sup>517</sup> The comparisons made with proteomics are similar to genomics and transcriptomics studies,<sup>518</sup> so many of the same principles are beginning to be used in data analysis. These include the ANOVA t-test<sup>519–520</sup> and multiple test corrections of Bonferroni,<sup>521</sup> Benjamini-Hochberg,<sup>522</sup> Storey-Tibshirani,<sup>523</sup> and Bayes.<sup>524</sup> Application and description of these tests in a proteomic context have been demonstrated on label-free spectral counting data<sup>525–527</sup> and metabolic-labeling data.<sup>270,524</sup> These comparison allow for extraction of statistically-significant changed proteins for biological validation. Further development and application of these methodologies to shotgun proteomic data sets will be essential to finding interesting biological changes.

## 5 Bioinformatics

Bioinformatics is an essential aspect of shotgun proteomics, especially with the production of increasingly complex data sets from increasingly comprehensive MS-based proteomic analyses. A schematic of a common shotgun proteomics bioinformatics pipeline is illustrated in Figure 10b. The integration of bioinformatic tools into proteomics pipelines and bundled software packages has begun to help standardize and simplify proteomic analysis for transfer of complicated, powerful methodologies from the experts to the users. Examples of these are the Integrated Proteomics Pipeline (IP2, <http://integratedproteomics.com/>), pFind Studio (<http://pfind.ict.ac.cn/>),<sup>528</sup> ProteoIQ (<http://www.bioinquire.com/>), ProteomeDiscoverer ([www.thermoscientific.com](http://www.thermoscientific.com)), Scaffold (<http://www.proteomesoftware.com/>), MaxQuant (<http://maxquant.org/>),<sup>529</sup> and Transproteomics Pipeline (<http://tools.proteomecenter.org>).<sup>530</sup>

Shotgun proteomics has developed into an automated pipeline of powerful analytical techniques. The automation has the advantage of wide-spread use – and unfortunately also misuse. A systematic study by the Human Proteome Organization (HUPO) of multiple proteomics labs revealed dramatic variability in results for a 20 protein test sample.<sup>531</sup> This highlighted the need for standardization of proteomic methodologies, which has followed,<sup>532</sup> and appropriate operator training, understanding, and usage of database search methods. As complex proteomes are commonly analyzed with shotgun proteomics, more complex standards are being developed and characterized for evaluation of system performance.<sup>533–536</sup> Other recent efforts for addressing this issue include development of quality control software to monitor key proteomic metrics both during and after LC-MS runs

and bioinformatics analysis.<sup>537–539</sup> Proper use<sup>540</sup> and comparison of proteomic algorithms and their performance<sup>541</sup> has also been highlighted.

### 5.1 Automated computational peptide identification methods

After analytical methodologies were worked out for the identification of peptides with tandem mass spectrometry,<sup>542</sup> one of the main limitations was manual interpretation of fragmentation spectra. Interpretation generally required an expert hand and, even then, was time consuming. For a LC-MS/MS run of a complex protein/peptide mixture, manual interpretation of all spectra was unrealistic. Thus, the crux of and one of the most influential contributions to the field of mass spectrometry-based proteomics was automation of identification of candidate peptide sequences from MS/MS spectra with software. An excellent conceptual review on this topic provides an overview of the different methodologies for matching peptide spectra to peptide sequences.<sup>543</sup> The first software developed for this purpose, SEQUEST, leveraged two timely technological developments, sequencing of genomes and computer software algorithms to match MS/MS peptide spectra to peptide sequences, and is still routinely used today.<sup>495</sup> Manually interpreting peptide MS/MS spectra is done by finding mechanistically-expected amino acid-specific fragment ions or mass intervals to build confidence in a sequence. The SEQUEST algorithm uses a similar strategy by simultaneously comparing all experimental fragment ions to theoretically generated fragment ions at the spectrum level. The SEQUEST software makes two key calculations which define whether a peptide sequence is a confident match for a fragmentation spectrum: XCorr and  $\Delta$ CN. XCorr is a statistical calculation of the correlation of the theoretical and experimental spectra.  $\Delta$ CN, the difference between the top peptide spectrum match (PSM) and the second best PSM, is then calculated. Initially,  $\Delta$ CN was used as the best measure of whether a PSM was correct,<sup>543</sup> but subsequent filtering methods and software have been introduced to statistically assess the confidence of a large set of identified PSMs. The theoretical peptide spectra are generated using a database of proteins, from as complex as the entire translated genome to as simple as a list of high confidence, validated translated protein candidates. For a tryptic digestion, only theoretical tryptic peptides of the appropriate length (typically 6 – 50 amino acids long) are considered as potential PSMs. Mascot is another commonly used search engine which instead relies on the probabilistic matching of fragment ions.<sup>496</sup> The Mascot search methodology originated from MALDI- and ESI-MS experiments on individually digested proteins where peptide mass fingerprints are used as the unique identifier of proteins. A probability score is calculated to determine how likely a calculated peptide mass fingerprint spectra matches an experimental one by chance. A similar probability scoring strategy of phosphorylation site informative fragment ions has proven useful for the confident localization of phosphorylation sites in global phosphoproteomics data.<sup>337</sup> Mascot has since been modified and expanded to calculate probability scores for peptide MS/MS data for complex mixtures and to include decoy matches within database searches.<sup>544–545</sup> X!Tandem followed SEQUEST and Mascot, combining three important attributes from each through the calculation of an expectation value.<sup>546–547</sup> The expectation value is a probabilistic assessment (similar to the Mascot probability score) of the correlation of an experimental and theoretical spectrum (similar to XCorr in SEQUEST) and how much better a peptide match is than a stochastic match (similar to  $\Delta$ CN in SEQUEST). The combination of database search engine results using various methods has improved overall identification numbers and confidence.<sup>548–549</sup> New software for identifying peptides from mass spectrometry data are continually being developed to improve upon widely-used methodologies, newly-implemented fragmentation methods, and for niche applications. A number of software tools have been developed to improve upon the SEQUEST-like search methodology to improve speed, sensitivity, and accuracy through consideration of fragment ion intensity,<sup>550–552</sup> algorithm architecture,<sup>553–556</sup> and higher quality data.<sup>557–559</sup> Similar to trends with SEQUEST, an

improved Mascot-like search engine (Andromeda),<sup>560</sup> mass accuracy-calibrated filtering,<sup>561</sup> and advanced machine learning filtering with Percolator<sup>562</sup> capitalize on higher quality data from newer mass spectrometers. On-the-fly database searching has allowed for intelligent, directed data acquisition.<sup>563–564</sup> Fragmentation spectra of metabolically labeled peptides were used to validate peptide spectrum matches from unlabeled peptides.<sup>565</sup> Probabilistic matching of fragment ions has also proven useful for identifying sequence tags – abundant peptide fragment ion patterns.<sup>566–569</sup> In particular, sequence tags facilitate the identification of protein site mutations and unexpected PTMs that would otherwise be missed by a traditional SEQUEST- or Mascot-like database search.<sup>570–572</sup> Database search engines have been either modified or newly developed to handle spectra from new fragmentation methods, such as ECD<sup>573–574</sup> and ETD,<sup>575–577</sup> and variant peptides, such as non-ribosomal peptides.<sup>578–579</sup> Other metrics have been introduced to improve PSM confidence and database search speed such as peptide precursor mass accuracy<sup>249,558–559,580–584</sup> and charge state predication,<sup>585–586</sup> empirical statistical models,<sup>482,587</sup> and amino acid composition.<sup>588</sup> Statistical analysis of large data sets from CID,<sup>589–590</sup> ETD,<sup>591–592</sup> and HCD<sup>593</sup> have contributed to a better understanding of peptide fragmentation mechanisms for spectral prediction and peptide identification algorithms. In an effort to combine both the computational power that has proven useful with database searches and the versatility of manual spectra interpretation, *de novo* sequencing methods are becoming a popular area of development. Although database searching is extremely powerful and versatile, *de novo* sequencing has provided a complementary and alternative unbiased means for identifying peptide sequences, particularly for unknown peptides and proteins, in an automated fashion.<sup>594–596</sup> Combining *de novo* sequencing with database search methods<sup>597</sup> have improved the speed<sup>598</sup> and sensitivity<sup>599</sup> of database searches and the confidence of *de novo* sequenced peptides.<sup>600</sup> Application of *de novo* sequencing to HCD fragmented<sup>601</sup> and Lys-N digested, ETD fragmented<sup>602</sup> peptides have exploited less common ion types for sequence determination.

Peptide search results are a mix of correct and false PSMs. The use of a decoy protein database was cleverly introduced as a way to filter for confident PSMs and is still most commonly used,<sup>603</sup> although some argue that a peptide probability is adequate and more resistant to false discoveries.<sup>604</sup> A decoy protein database is created by reversing or scrambling protein sequences from a normal database which is then appended to the normal protein database. These decoy peptides and proteins are then identified, with lower frequency, during the database search and used as a measure of random, incorrect PSMs. The random and incorrect decoy PSMs can then be used to establish a cutoff for the high confidence PSMs and to estimate the false discovery rate (FDR) of PSMs, peptides, and proteins. Generally, two peptides have been required for identification of a protein, but there is also evidence that a single high confidence peptide identification can be a better indicator than two lower confidence peptides.<sup>603,605</sup> Semi-supervised machine learning of peptide filtering improved peptide identifications 17% from tryptic digests and 77% from non-tryptic digests.<sup>606–607</sup> Application of statistical methods in combination with the target decoy approach provided greater significance analysis of peptide matches.<sup>608</sup> An algorithm to optimize the protein filtering improved protein identifications by ~25% over common methods.<sup>609</sup> Considering both PSM and protein false discovery simultaneously in a two-dimensional target decoy strategy yielded a more versatile and accurate representation of peptides and proteins within a sample.<sup>610</sup> A strategy has been presented to infer the proteins present within a complex sample from hundreds of LC-MS runs where protein false discovery rates are difficult to control using conventional PSM false discovery rates.<sup>611</sup> Similarly, protein database size and quality directly influence the size and quality of the identified protein and peptide candidates. Essentially, with a larger, less curated database the chances of matching decoy PSMs increases, subsequently decreasing the sensitivity of correct PSMs. Thus, database curation is an essential and continuing aspect of proteomics

research, recently illustrated with a customized protein database derived from RNA-Seq data.<sup>612</sup> For newly sequenced organisms yet to be annotated, a proteogenomic approach can be used to search against the 6-frame translated genome.<sup>613–617</sup> For organisms which have yet to be sequenced or are extinct,<sup>618–622</sup> a combination of known databases can be used to identify peptides and proteins. These methodologies have even been used to find protein sequences and correct gene annotations for highly studied and characterized organisms such as *S. cerevisiae*,<sup>623</sup> *C. elegans*,<sup>624</sup> and *Mus musculus*,<sup>625</sup> and *Arabidopsis*.<sup>626</sup> Leveraging known protein interaction networks, discarded PSMs were matched to proteins expected to be expressed and present as part of functional protein complexes.<sup>627</sup> Although software has eased the burden of manual interpretation of tens of thousands of tandem mass spectra, manually validating peptide spectra for proteins and post-translational modifications is still an essential principle of proteomic analysis.

## 5.2 Protein inference from peptide identifications

Peptides identified with tandem mass spectrometry are used as surrogate representations of intact proteins and their post-translational modifications. We have already described many of the experimental advantages of analyzing proteins as peptides. However, an ongoing challenge of shotgun proteomics is correctly mapping identified peptide sequences to protein sequences, particularly for redundant proteins and isoforms. Further, as proteomics methods continue to become more sensitive and comprehensive and protein databases increase in size, the ability to properly assign peptides to proteins has become even more challenging. At the same time, shotgun proteomics experiments have become more complicated and biology-driven, so the importance of correctly identifying and quantifying proteins from peptides has continued to increase. These challenges include the stochastic sampling and identification of unique peptides from protein isoforms (homologous proteins) or unrelated proteins which share non-unique tryptic peptide sequences (redundant proteins). When peptide sequences are shared among proteins, they are by definition more prevalent and abundant than their unique peptide counterparts, and thus also more easily identified. Conversely, unique peptides are less abundant and harder to identify. Unfortunately, the hard to identify unique peptides carry the most experimental evidence for unambiguous, confident identification of both homologous and redundant proteins. Thus, many bioinformatics strategies have been proposed and implemented to directly address this inherent challenge. Due to the complexity of this issue, a number of guidelines have been established by journal editors and conference organizers which aid in simplifying publication of proteomics data.<sup>628–632</sup>

Initially, when MS data sets and protein databases were small, peptide database search engines<sup>482,587</sup> and filtering software<sup>633</sup> assembled and listed all proteins for which there was peptide evidence. In particular, DTASelect and Contrast software packages grouped redundant proteins and isoforms within a single proteomic analysis and allowed for comparison between experiments based on spectral counts and sequence coverage. Generation of a minimal protein list which best described the identified peptide spectrum matches with the Isoform Resolver algorithm yielded a more conservative representation of proteins present in a sample.<sup>634</sup> A similar strategy using bipartite graphical analysis in a software package called IDPicker further improved the accuracy and transparency of protein identifications within a single proteomic analysis.<sup>635–636</sup> Further classifying peptides as either ambiguous or unambiguous for not only protein sequences and protein isoforms, but also genes using the ProteinClassifier software added further confidence and clarity to protein inferences.<sup>637</sup> The majority of these strategies rely solely on the prevalence of peptide identifications, but a few methods have incorporated protein probabilities. The software tool PANORAMICS facilitated calculation of peptide-to-protein assignment probabilities by combining Mascot PSM probabilities and the probabilities of peptides and

proteins generating spectra.<sup>638</sup> Other distinctive strategies seek to bypass assembling protein identifications from peptide identifications and directly perform protein sequencing from tandem mass spectra of peptides. An “overlap → layout → consensus” strategy for assembling proteins was borrowed from DNA fragment sequencing for “sequencing” proteins from non-specific protein digestions.<sup>639</sup> Essentially, repeat peptide identifications were clustered and signal averaged to generate a consensus spectra. Then the consensus spectra were all aligned until the overlaps created a full protein sequence for  $\alpha$ -synuclein. A similar strategy with peptide alignment using spectral network analysis has facilitated the unbiased identification of post-translational modifications on proteins.<sup>640–641</sup> Although these sequencing strategies have yet to be applied to highly complex mixtures, it has aided in clustering millions of tandem mass spectra from large scale experiments for faster and more sensitive database searches.<sup>642</sup> An excellent tutorial on the protein inference problem and implementation of many of these methodologies has been described elsewhere.<sup>643</sup>

The identification of a single high confidence unique peptide spectrum match can provide high confidence in the presence of a protein, particularly when other high confidence non-unique peptides are also present to support that protein inference. However, quantification by both label-free and isotopic labeling methods can be more complicated. Generally, strategies that remedy the protein inference problem also improve protein quantification.<sup>637</sup> In the case of label-free detection using spectral counting, non-unique peptides must be appropriately assigned to redundant proteins to estimate their relative abundances. Similar strategies to address protein identification inference are regularly applied to spectral counting-based quantification, such as with Abacus,<sup>644</sup> but caveats exist. A systematic study using six homologous albumins spiked into a complex mixture illustrated that distributing shared spectral counts based on the number of unique spectral counts led to the most accurate and reproducible quantification results.<sup>645</sup> For relative ratio measurements using either label-free peptide ion currents or isotopically-labeled peptide ratios, the most accurate method for protein quantification is the use of only unique peptides sequences.<sup>646–648</sup> This method is the most stringent as it completely ignores the problem of shared peptides, and unfortunately a wealth of quantitative measurements, making it the least sensitive and comprehensive. Additionally, this method is highly dependent on multiple proteomics parameters; high complexity samples, poor LC-MS sensitivity, and large database size can all limit the number of unique peptides identified and thus quantified from a sample. Further, a small number of quantitative measurements from unique peptides can limit statistical analysis. An advantage of label-free ion current and metabolically-labeled peptide quantification is that multiple sequential measurements are made on peptide chromatographic peaks from MS1 precursor scans, accounting for measurement variability and noise within MS1 scans over the chromatographic peak timescale. Isobaric peptide labeling is used to perform quantification in the MS2 scan and is thus only measured when a peptide precursor ion is sampled for fragmentation. The estimation of measurement variability and noise for a peptide can only be done on peptides that are resampled and thus are highly dependent on the abundance of a protein and its corresponding peptides. Therefore, the ability to use all peptide quantification information, unique or not, would benefit isobaric labeling methods for protein quantification the most. One proposed solution for the use of shared peptides in protein quantification using ratio measurements was demonstrated with peptide ion current measurements.<sup>649</sup> The methodology combines the concepts described above used for protein identification inference and for distributing spectral counts of shared peptides among proteins with unique peptide counts. With peptide ratio measurements, peptides shared among proteins can be grouped and assigned to proteins with unique peptide ratio measurements which are statistically similar. The study showed that peptide measurements for protein groups from transferrin and histone 1 isoforms could be appropriately grouped to quantitate individual isoforms of each. Similarly, the differentiation of peptide measurements among the shared peptides from the redundant

proteins desmin, vimentin, and keratin was possible. A similar strategy for inclusion of shared peptides in protein quantification was recently demonstrated using a combinatorial optimization to reduce relative abundance measurement error.<sup>650</sup> The effect of shared peptides on protein identification and quantification accuracy remains a common challenge in shotgun proteomics. Further software development in the bioinformatics field should continue to address these challenges.<sup>651</sup> The continued development of middle-down proteomics, which identifies larger peptide fragments with higher probability of being unique, should also help to address this wide-spread challenge.<sup>14,297,652–654</sup>

## 6 Applications

### 6.1 Protein-protein interactions and protein complexes

Typically, biological processes are carried out by interactions between many biomolecules. There are diverse types of interactions, such as protein-DNA or RNA, DNA-DNA, DNA-RNA and protein-protein interactions. Of these biomolecule interactions, protein-protein interaction are challenging to study due to their high diversity.

The classic approach to studying protein-protein interactions is yeast two-hybrid (Y2H) system (Figure 11a) which was introduced more than 20 years ago.<sup>655</sup> In a Y2H experiment, a transcript factor is split into two subunits, one is the binding domain (BD) and the other one is the activating domain (AD). The engineered bait protein is fused to the BD and the second protein (prey) is fused to the AD. If the bait and prey proteins interact with each other, the transcript factor can be activated, starting the transcription of reporter gene. Otherwise, there will be no transcription of the reporter gene (Figure 11a). Y2H was designed to investigate direct binary interactions. The initial Y2H experiments focused on the interactions between a limited number of proteins. However, after genome scaled resources of open reading frames (ORFeomes) became available, comprehensive network maps have been drawn for various model organisms by large scale Y2H, including *Saccharomyces cerevisiae*,<sup>656–659</sup> *Drosophila melanogaster*,<sup>660</sup> *Caenorhabditis elegans*,<sup>661</sup> and more recently even human.<sup>662–665</sup>

While Y2H has been widely used in high throughput protein interactome studies, there are limitations. The primary limitation is that the protein interactions do not occur in their physiological cellular conditions, and it is only possible to detect direct interaction between two proteins; any information on indirect interactions will be ignored.

An approach complementary to Y2H based on affinity purification and mass spectrometry (AP-MS) was developed to explore the protein-protein interaction for both targeted and large scale research. In an AP-MS experiment the target protein of interest, together with its interacting partners, is purified from a protein mixture. The purified protein complex is then subjected to shotgun proteomic identification and quantification (Figure 11b).<sup>10,614</sup> PTM information can also be collected if desired. Because it is carried out under endogenous conditions, the AP-MS strategy enables identification of both direct and indirect interactors. Generally, the topology of the protein interactome is not found from the AP-MS strategy because it is incapable of distinguishing direct and indirect interactors (Figure 11b). However, chemical cross-linking of protein complexes and the analysis of their cross-linked peptides can provide this information.<sup>666</sup> Alternatively, the reverse purification of protein complex components can be used to infer topology, as in global AP-MS studies. Methods which analyze intact protein complexes are also being developed to address these issues.<sup>667–668</sup> Ideally, interactome experiments based on AP-MS should employ high quality monoclonal antibodies against the bait proteins. However, this is sometimes difficult due to a lack of good antibodies. Commonly, an engineered protein with an affinity has been used in AP-MS based protein interactome studies. Instead of using a specific antibody against

target protein, the affinity tag system employs a uniform tag specific purification which can be used for many bait proteins. A widely used tag system is tandem affinity purification (TAP),<sup>669</sup> which consists of calmodulin-binding peptide (CBP) and protein A of *Staphylococcus aureus* (ProtA), linked by a tobacco etch virus (TEV) cleavage site (Figure 12). The TAP fused protein and its interactors are first pulled down by the distal ProtA affinity tag, and then released by cutting at the TEV cleavage site. The bait protein complex is subsequently subjected to the second purification step, which binds the CBP (Figure 12). The advantage of TAP compared with the normal single-step procedure is reduced background protein levels. Ideally, all the proteins non-specifically binding to the affinity beads are excluded after the second purification. Many other variations of TAP have been developed with other features.<sup>670</sup>

Large scale studies on protein interactomes have also been achieved using AP-MS approach. The pioneering large scale work by AP-MS focused on the yeast system,<sup>671–674</sup> but other organisms have also been investigated, including *Escherichia coli*,<sup>562,675–676</sup> and human.<sup>677</sup>

Although AP-MS has potential for drawing the whole interactome map, results can contain both false positive and false negative interactors. False positives are mainly caused by nonspecific interactions, such as proteins binding to affinity matrices while false negatives are generally from weak interactions. If experimental conditions do not mimic the physiological conditions, weakened binding between interactors can be observed. Because removal of non-specific binding protein typically requires intensive wash steps, weak interacting proteins can be lost.

Using AP-MS, either in target or large scale studies, researchers have obtained valuable evidence on how proteins interact. This information can help us gain understanding of disease and biological process mechanisms, and the influenced pathways can provide promising targets for pharmaceutical discovery. Furthermore, the multiple interactors in a network may work together as diagnostic or prognostic biomarkers.

## 6.2 Comparative systems analysis

Many comparative global proteomic studies seek to find the most relevant proteins in a biological process through expression changes upon perturbation. With enough relative protein change measurements, comparative analysis of entire biological systems can be performed. Proteomic systems analysis has been primarily driven by improvements in the comprehensiveness of quantitative analysis and the bioinformatics tools which facilitate these categorizations and comparisons. Quantification of protein expression and post-translational modification changes on thousands of proteins consecutively allows for analysis of the entire system under either normal or perturbed conditions. In particular, pathway analysis tools such as Ingenuity have simplified these comparisons. A PubMed search with key words “comparative”, “proteomics”, and “mass spectrometry” yields almost 2,500 citations. This large number of publications is a representation of how prevalent and important these types of studies have become. Since this number of publications is surely outside the scope of this review we have attempted to list and briefly describe notable studies in a chronological fashion below. Phosphoproteomic-based systems based studies are described in the ‘Investigation of signaling networks through phosphoproteomics’ section. These references should provide examples of the capabilities of shotgun proteomics in systems biology.

Integration of quantitative proteomics data with DNA microarray and protein interaction data from yeast initially demonstrated the ability to perform quantitative pathway analysis.<sup>678</sup> Nutrient limitation of yeast and comparison of transcript and protein expression data revealed alternative carbon source pathways are transcriptionally upregulated with

reduced glucose, while nitrogen scavenging pathways are upregulated post-transcriptionally through more efficient translation and/or reduced protein degradation with reduced ammonia.<sup>679</sup> Application of absolute protein quantification measurements revealed protein abundance within prokaryotes (*E. coli*) is equally regulated between transcriptional and translational mechanisms while in eukaryotes (*S. cerevisiae*) > 70% is controlled by post-transcriptional, mRNA-directed mechanisms.<sup>680</sup> Subtractive proteomics on nuclear envelopes discovered new integral membrane proteins linked to dystrophies.<sup>681</sup> Comparative proteomic analysis of long-lived *C. elegans* mutants identified insulin signaling targets that modulate lifespan.<sup>457</sup> Comprehensive quantitative proteomics of haploid versus diploid yeast suggested efficient control of the pheromone pathway and mating response.<sup>178</sup> Proteomic analyses revealed evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans.<sup>682</sup> Lysine acetylation was found to target protein complexes and co-regulate major cellular functions.<sup>405</sup> New roles for protein glycosylation were discovered in eukaryotes, namely mitochondrial protein function and localization, from global analysis of the yeast glycoproteome.<sup>683</sup> Three functionally different human cell lines were compared; large abundance differences from various functional protein classes were observed along with low abundance cell-type specific proteins and highly enriched cell-surface proteins.<sup>684</sup> Comparison of transcript and protein levels in yeast upon osmotic stress revealed transcript increases correlated to protein increases, but transcript decreases didn't affect protein abundance.<sup>685</sup> A quantitative comparison between human induced-pluripotent and embryonic stem cell lines found the majority of proteins (97.8%) unchanged, indicating differences are from experimental conditions and not molecular signatures; the 58 changed proteins were involved in metabolism, antigen processing, and cell adhesion.<sup>686</sup> Through systematic deletion of the two kinases, one phosphatase, and two N-acetyltransferases within *Mycoplasma pneumonia*, followed by proteomic analysis, the cross-talk between phosphorylation and lysine acetylation modifications and their roles in regulating protein abundance were demonstrated.<sup>687</sup> From these system-wide discoveries, there is no doubt that shotgun proteomics has become an essential tool in biological studies.

### 6.3 Systems analysis of PTMs

Signal networks induced by various extracellular and intracellular signals are essential for regulation of cells and organisms; many diseases can occur due to malfunctions in signal pathways. In order to understand biology and pathology in cell systems, study of the mechanisms of signal networks is necessary. Regulation of post-translational modifications (PTMs), including phosphorylation, glycosylation, acetylation, ubiquitination, and methylation, is one of the major mechanisms in signal networks. For example, once binding of endogenous or exogenous ligands to membrane receptors has occurred, a signal can usually be amplified through a phosphorylation cascade, affecting the level of expression of various genes. In addition, a signaling network may be involved in crosstalk with other signaling networks.

Western blotting using specific antibodies has traditionally been performed for detection of these PTMs, with signal intensities from Western blotting as an estimation of changes to PTMs in a specific state. However, limitations to use of antibodies include nonspecificity of antibodies, variance in quantification, and absence of specific antibodies to specific proteins or PTM motifs. The advantages of methods recently introduced for mapping PTM with MS have been demonstrated, even though there are challenges to identification of modified peptides of low abundance and analysis of MS/MS spectra from complex data sets have been reported.<sup>688,689</sup> Below, we have described how various applications of MS-based shotgun proteomics are utilized for characterization of signal networks through PTMs.

**6.3.1 Investigation of signal networks through phosphoproteomics**—The first study of the global signaling network of epidermal growth factor (EGF) based on MS was reported in 2004.<sup>690</sup> In this study, the time course of EGF signals through their activation upon stimulation of epidermal growth factor was determined by identifying 81 signaling proteins and 31 novel effectors. In addition, 6,600 phosphorylation sites from 2,244 proteins showing dynamic regulation by EGF stimulation were found.<sup>690</sup> Several studies focusing on Tyr phosphorylation signaling have been reported for insulin stimulation, which plays an important role in control of glucose, and many new insulin-regulated substrates have been identified.<sup>691</sup> Similarly, Tyr phosphorylated proteins have been quantified in order to reveal signal networks of platelet-derived growth factor (PDGF) for proliferation of mesenchymal stem cells.<sup>692</sup> In addition, a SILAC-based study of the differentiation of human embryonic stem cells identified 3,067 phosphorylation sites.<sup>693</sup> A label-free quantification study of the effect of multiple tyrosine receptor kinases in growth of undifferentiated embryonic stem cells identified 2,546 phosphorylation sites.<sup>694</sup> Phosphoproteomic analysis of toll-like receptor (TLR)-activated macrophages confirmed canonical targets within the TLR pathway and identified the cytoskeleton and new signaling molecules as targets of phosphorylation.<sup>695</sup> Based on two different fragmentation methods, CID and ETD, more than 10,000 phosphorylation sites have been detected in embryonic stem cells.<sup>696</sup>

In yeast, quantitative phosphoproteomics has been applied to the study of the pheromone signaling pathway and 139 of the more than 700 identified phosphopeptides were found to be differentially regulated.<sup>697</sup> In addition, findings from a study of MAPK pathways required for mating versus filamentous growth determined that Fus3 (mating-specific MAPK) can prevent crosstalk between the two pathways.<sup>698</sup> Recent work on the phosphoproteome of the yeast centrosome identified 297 phosphorylation sites from different cell cycle stages, and mutation studies of phosphorylated residues, identified by MS in Spc42 and  $\gamma$ -tubulin, demonstrated the important roles of phosphorylation in the biology and function of centrosomes.<sup>699</sup>

Through phosphorylation, protein kinases can affect the activity, cellular location, protein binding, and stability of proteins. The catalytic domains of most kinases are highly conserved and are involved in ATP binding and phosphorylation of target proteins; allosteric modulation of this catalytic domain is exerted through a regulatory domain. Based on substrate recognition sites, kinases can be categorized into two classes: Ser/Thr kinases and tyrosine kinases. Many global approaches to phosphorylation have been reported. However, due to low abundance of protein kinases, compared with their substrate proteins, phosphorylation studies of protein kinases have been underrepresented.<sup>314,315,350,700</sup> Antibody-mediated purification of tyrosine phosphorylated proteins<sup>692</sup> and immobilized kinase inhibitors has been performed for concurrent enrichment of protein kinases.<sup>701,702</sup>

Use of kinase-selective affinity purification for the kinome approach across the cell cycle resulted in quantification of 219 protein kinases from S and M phase-arrested human cancer cells and also identified more than 1,000 phosphorylation sites on protein kinases.<sup>703</sup> Development of an integrated method termed KAYAK (Kinase Activity Assay for Kinome profiling) for large-scale kinase activity profiling using stable-isotope dilution and high resolution MS has been reported.<sup>704</sup> In this study, 90 site-specific peptide phosphorylations were measured, and the source of activity for a peptide derived from a PI3K regulatory subunit was identified, which was confirmed to be a novel Src family kinase site *in vivo*.<sup>704</sup> Throughput and multiplicity in this method was improved by assessing rates of phosphorylation for all 90 of the same peptides used previously in a single reaction.<sup>705</sup> The modified KAYAK strategy reports activation of cellular signaling pathways, including mitogen stimulation of HEK-293 and HeLa cells. In addition, combining single-reaction KAYAK with kinase inhibition enables resolution of pathway specificity.<sup>705</sup>

**6.3.2 Investigation of signal networks through proteomics of other PTMs—O-GlcNAc** ( $\beta$ -*O*-linked 2-acetamido-2-deoxy- $\beta$ -*D*-glucopyranose) modification plays an important role in signaling networks, including nuclear transport of cytosolic proteins, protein degradation, and tolerance of cellular stress. This modification is reversible and is controlled by two enzymes, O-GlcNAc transferase (OGT) and O-GlcNAcase (OGA).<sup>706</sup> O-GlcNAcylation competes with phosphorylation at the same serine/threonine sites or at proximal sites on some proteins. O-GlcNAc modification is a low abundance and labile modification, therefore, detection of the structure of glycans and of modification sites on proteins using traditional mass spectrometric methods is difficult. For identification of O-GlcNAcylated proteins or modification sites, a number of enrichment strategies have been employed including: O-GlcNAc specific lectin succinyl wheat germ agglutinin (sWGA),<sup>707</sup> immunoprecipitation using O-GlcNAc specific antibody,<sup>708</sup> peptide tagging with BEMED ( $\beta$ -eliminated followed by Michael addition with dithiothreitol),<sup>709</sup> chemoenzymatic tagging-enrichment strategy,<sup>710</sup> and a photocleavable biotin-alkyne reagent (PCbiotin-alkyne) tag.<sup>711</sup> In studies to determine a dynamic relationship between O-GlcNAcylation and GSK-3 dependent phosphorylation, at least 10 proteins showed increased O-GlcNAcylation upon GSK-3 inhibition by lithium, whereas 19 other proteins showed decreased O-GlcNAcylation.<sup>712</sup> To examine the effect of site-specific phosphorylation dynamics on globally elevated O-GlcNAcylation, sequential phosphopeptide enrichment was combined with iTRAQ labeling which resulted in the quantification of 711 phosphopeptides in NIH3T3 cells.<sup>713</sup> According to finding from these studies, cross-talk between two major PTMs is extensive; this result may be due to both steric competition of occupancy at the same or proximal sites and regulation of the other's enzymatic activity. In addition, regulation of activities of Calcium/calmodulin-dependent kinase IV (CaMKIV) and IKK $\alpha$  kinase (IKK) by O-GlcNAc modification has been reported.<sup>714,715</sup> Pharmacologically or genetically induced increases in O-GlcNAc resulted in severe defects in mitotic progression and cytokinesis.<sup>716,717</sup> These data clearly exemplify the importance of the interplay between O-GlcNAcylation and phosphorylation in cellular signaling pathways.

Ubiquitination (Ub), conjugation of Ub to the 3-amino group of a lysine in a substrate protein, is involved in regulation of various cellular pathways, including signal network, protein degradation, and cell division. While mono-Ub is related to ligand-mediated endocytosis, poly-Ub is related to the proteasome pathway for protein degradation. Using quantitative MS, identification and quantification of various forms of Ub bound to EGFR have been achieved using the Ub-AQUA method. This result demonstrated that more than 50% of all EGFR bound Ub was a poly-Ub form, which was targeted to the lysosome.<sup>718</sup> Cyclin B1 is also modified by Ub chains of complex topology, rather than by homogeneous Lys 48-linked chains, and is degraded by the 26S proteasome.<sup>719</sup>

Methylation and acetylation of histones is known to be part of an epigenetic marking system using different combinations of histone modification for control of gene expression.<sup>720,721</sup> Findings from SILAC-based proteomic analysis in breast cancer cells treated with a histone deacetylase (HDAC) inhibitor showed induction of lysine acetylation in chromatin-associated proteins as well as non-histone proteins, including transcriptional factors, chaperones, and cell structure proteins.<sup>722</sup> Regulation of EGFR trafficking by lysine deacetylase HDAC6 through control of acetylation of alpha-tubulin has been reported.<sup>723</sup>

S-nitrosylation, modification of cysteine thiol by a nitric oxide (NO), is emerging as an important PTM involved in regulation of cell signaling.<sup>724</sup> Due to the labile nature of the S-NO bond, detection of S-nitrosylated proteins is difficult, because it is frequently lost during sample preparation. Several methods, including the biotin switch assay and the chemical reduction/chemiluminescence assay, have been developed.<sup>725,726</sup> The biotin switch method

converts S-nitrosylated Cys to biotinylated Cys. This method has been used for identification and characterization of S-nitrosylated proteins in *Arabidopsis* plants, resulting in identification of 63 proteins from cell cultures and 52 proteins from leaves as candidates for S-nitrosylation, including signaling, redox-related, and stress-related proteins.<sup>727</sup>

Classical biochemical studies using antibody-based approaches to conduct unbiased large-scale studies of various extracellular and intracellular signals are limited. Because regulation of many proteins occurs after transcription of genes, global MS-based PTM proteomics enables system-wide characterization of signaling networks and complements gene expression analysis.

#### 6.4 Protein turnover

With modern mass spectrometers and quantitative methods, large scale and targeted protein information is obtainable, including measurement of quantitative protein abundance levels and protein modifications. However, current methods only provide a ‘snapshot picture’ of a cellular state at the time that sampling occurs. Protein abundance and modification are both highly dynamic and tightly regulated. Measurement of the dynamics of a proteome may provide critical answers for many biological questions. It has been shown that there is a poor correlation between mRNA and protein levels,<sup>678,728–729</sup> which may be partly explained by the protein synthesis/degradation rate. Moreover, protein turnover information can also be valuable as a biomarker. Since protein quantity is determined by synthesis and degradation rates, a change in protein turnover rates may precede a change in protein expression level. Several studies have observed significant changes in protein turnover but smaller changes on protein expression.<sup>730–731</sup>

Protein turnover analysis is not a new idea, and can be dated back to 1930’s.<sup>732</sup> A typical turnover study involves administering either radioactive or stable isotope labeled tracers, which are incorporated into the protein synthesis. The resulting labeled proteins can be captured and determined by either scintillation counting or mass spectrometry. The classic ‘pulse-chase’ strategy uses radioactive isotopes to partially label proteins in a ‘pulse’ period, and measures their decay during the ‘chase’ period, when the non-radioactive isotope is applied. The advantages of this platform are that it only requires labeling part of a protein, and only the labeled fraction is measured along the time course. However, this method is more suitable for assessment of the average decay rate for all proteins in a biological system. Individual protein turnover information can only be obtained by protein purification, which limits the throughput of this method.

Besides isotope tracers, there are other methods developed for protein turnover or stability studies. One strategy is to inhibit the protein biosynthesis using cycloheximide.<sup>733</sup> The protein degradation rates can be assessed by protein band intensities during the time course of cycloheximide treatment because there is no newly synthesized protein during this period. This method is usually used for the measurement of one or a limited number of proteins, since western blot is a low-efficiency detection assay. Recently, this strategy was scaled up to nearly 4,000 proteins by using a collection of yeast strains expressing tandem affinity purification (TAP) tag<sup>733</sup> (Figure 12), which was universally used for target in immunodetection. One drawback of this method is that the inhibition of protein synthesis may greatly influence the biological nature of cells, and further change protein stability. Other newly developed methods rely on the libraries of clones expressing different proteins tagged by fluorescence.<sup>734–735</sup> These methods indeed make protein turnover analysis possible on a large scale. However, the establishment and utilization of such engineered libraries are time and labor intensive processes and it is still quite challenging to apply this method to all organisms, especially mammals.

As discussed above, radioactive labeling, protein synthesis inhibition, and fusion protein libraries all have benefits and shortcomings for protein dynamics studies. Alternatively, several MS and stable isotope based approaches have emerged to explore protein turnover on a large scale for both single cell systems and animals. Compared with the radioactive isotope 'pulse-chase' method, a stable isotope labeling turnover experiment typically only includes a 'pulse' phase. This means that upon switching the medium/diet, the organism is only grown in an isotope-enriched environment until the sampling time points. The ratios between unlabeled and labeled peptides are used to interpret the protein turnover rates.

The most common type of protein turnover study is carried out on a single cell system, such as a cultured cell line. In this case, the design of the experiment is relatively straightforward: the protein synthesis precursor pool is principally composed of the amino acids from the medium. A switch of medium from an unlabeled to a labeled state or *vice versa* results in a rapid incorporation of isotopes in the newly synthesized proteins. In other words, almost all the newly synthesized proteins generated after the medium change will be labeled with the same relative isotope abundance (RIA) as is present in the medium. Data analysis is straightforward because the two groups of proteins, pre-existing and newly synthesized, both have defined RIA and predictable mass spectrometry distribution.<sup>736</sup> An example of this method used the pulsed SILAC (pSILAC) approach to show the influence of microRNAs on protein synthesis.<sup>737</sup> miRNA transfected and control HeLa cells were subjected to either medium-heavy or heavy SILAC medium in the pulse. The mass spectrometry peak abundances of medium-heavy or heavy groups, representing the newly synthesized proteins, were used to compare the protein synthesis. More than 3000 proteins were quantified on their synthesis level, and hundreds of genes were observed to be directly repressed by microRNA. Another study employed a triple-SILAC method to measure protein synthesis and degradation on 8,041 HeLa cell proteins.<sup>738</sup>

However, this strategy cannot be easily adapted to whole animal studies because the change of dietary isotope does not lead to a sudden and complete labeling of the precursor pool. Incomplete labeling results from the breakdown and re-use of unlabeled amino acids in protein recycling. For animals, such a rapid precursor pool change can only occur via intravenous injection of the isotope tracer by either continuous infusion<sup>739–740</sup> or a flooding dose.<sup>741</sup> Alternatively, oral administration of stable isotope labeled water,  $^2\text{H}_2\text{O}$  or  $\text{H}_2^{18}\text{O}$ ,<sup>742</sup> also achieves fast equilibration between body water and free amino acid. The limitation of these methods is that measuring protein turnover of individual proteins in a high throughput manner is not always possible due to its low overall RIA. Two recent studies have used this water strategy, measuring the protein turnover on a global scale.<sup>743–744</sup> Next, the sudden intake of the isotope tracer by intravenous injection may alter protein synthesis and lead to artificial protein turnover results,<sup>745</sup> and the hydrogen/deuterium exchange is reversible and can result in a back-exchange from deuterium to hydrogen during sample preparation in aqueous solvents.<sup>746–747</sup> Furthermore, the water method relies on the analysis of a minor isotopic peak pattern change, which could be easily influenced by error in the ion intensity measurement. To avoid these complications the isotope tracer can be administered more naturally via the diet<sup>748–749</sup> with fewer side effects. The major challenge here is that the protein synthesized during pulse phase will be labeled to varied isotope extents since the precursor pool will saturate gradually. In one study, chicken was chosen because of its continuous diet intake which leads to a quick increase and plateauing of precursor RIA that subsequently remains constant.<sup>748</sup> However, this method is not applicable to other mammals, including mice.

Two recent independent studies have employed similar  $^{15}\text{N}$  labeled diets to determine the protein turnover in mice at the proteome scale.<sup>749–750</sup> Both groups switched the normal diets of the mice to a  $^{15}\text{N}$  labeled diet at a given time point and analyzed the protein samples after

different feeding times, ranging from a half day to several weeks. They then measured the ratio between unlabeled and labeled peptides derived from the brain and plasma proteins. The ratio of the unlabeled/labeled peptides reflects the abundances of pre-existing and newly synthesized proteins (Figure 13). Both studies revealed the mouse protein turnover information on a large scale. For quantification purposes, special algorithms were developed for data interpretation, as the partial  $^{15}\text{N}$  labeled peptides generated complicated isotopic peak distribution. Price *et al* tried to define the  $^{15}\text{N}$  incorporation rates of all newly synthesized peptides.<sup>750</sup> The observed  $^{15}\text{N}$  peptides were a heterogeneous group synthesized from a time window with varied incorporation rates. Furthermore, the  $^{14}\text{N}$  and  $^{15}\text{N}$  peptides are most likely to overlap with each other when the precursor RIA is low (Figure 13b–c). Zhang *et al.* avoided the requirement for the knowledge of labeling incorporation rates and complete separation of the tryptic peptides, providing a straightforward data processing platform.<sup>749</sup> In both studies, an  $^{15}\text{N}$  diet was used instead of a SILAC diet to determine protein turnover in mammals because if a SILAC diet was used, part of the newly synthesized SILAC peptide would appear in the unlabeled isotope pattern unless the precursor is fully labeled, making it impossible to distinguish between the two with MS. Although the newly synthesized SILAM peptides may also appear at a low incorporation rate shortly after the labeling, it is still possible to separate it from the pre-existing unlabeled peptide. The  $^{15}\text{N}$ -labeled tracer, present in all amino acids, significantly amplifies protein labeling efficiency (Figure 13). For instance, if a protein is synthesized in a 20% labeled precursor pool, SILAC will have 80% of total newly synthesized peptides appearing in unlabeled form, whereas SILAM will have little chance of generating totally unlabeled peptides, because all the 20 amino acids are labeled.

The order of isotopic treatment is not critical, since newly synthesized peptides can be detected as either heavy or light. Both will lead to equivalent results. However, since fully labeling an organism consumes a great deal of time and money, it is generally preferable to start with peptides in an initial condition of natural isotope, followed by partial heavy isotope labeling. Recently, a reverse strategy was conducted in rat to study extremely long-lived proteins which may provide insight into cell aging and neurodegenerative diseases.<sup>751</sup> The rats were fed a  $^{15}\text{N}$  diet starting *in utero*. After the labeling efficiency reached 100%, the food supply was switched to standard  $^{14}\text{N}$ . After 12 months, almost all the proteins had been resynthesized with natural isotope composition and any remaining  $^{15}\text{N}$  peptides represented the initially synthesized proteins. These long-lived proteins were found to be related to the nuclear pore complex, the central transport channels that mediate all molecular trafficking in and out of the nucleus. Clearly, the reverse-labeling method was more suitable for this study, since it did not require long term (12 months) isotope feeding and avoided any confusion caused by  $^{14}\text{N}$  peptide contamination which comes from other sources, such as keratin. In another study, researchers used multi-isotope imaging mass spectrometry, together with  $^{15}\text{N}$ -labeling, to reveal that the stereocilia were generally remarkably slowly turned over, but a rapid turnover was found only at the tips of stereocilia.<sup>752</sup>

With current MS-based strategies, decoupled protein synthesis and degradation rates cannot be easily measured simultaneously, because the quantitative measurements only represent the relative abundance, not the absolute abundance. The ratio between unlabeled and labeled isotope peaks can be affected by any change in pre-existing and newly synthesized proteins and the turnover rate measured describes the relative change between protein synthesis and degradation. Currently, the fractional turnover rate calculation is based on the assumption that the protein is in steady-state (i.e. the protein has equal rates for synthesis and degradation), is not always true.

## 6.5 Clinical application

Recent advances in MS-based proteomics have made it an indispensable tool for use in the clinic. The goal of clinical proteomics is the identification of disease-specific biomarkers using clinical samples, such as plasma and tissues from patients. Identification of useful biomarkers of disease should provide a better chance for early diagnosis, optimization of treatment, and means for monitoring progress during treatment. To increase the chance for identification of biomarkers, MS-based proteomics technologies that enable identification of a large number of proteins with high throughput using small amounts of materials have been developed. Clinical proteomics using the latest proteomic technologies in conjunction with advanced bioinformatics is currently being used for identification of molecular signatures of diseases based on protein pathways and signaling cascades, and holds great promise for the future of medicine.

**6.5.1 Biomarkers**—A biomarker is a substance used as an indicator of normal biological processes, pathogenic processes, or pharmacological responses to a therapeutic invention.<sup>753,754</sup> A biomarker can be a traceable substance used for examination of change in expression of a protein that reflects the risk or progression of a disease, or the susceptibility of the disease to a treatment. Serum markers, such as prostate-specific antigen (PSA) for prostate cancer, C-reactive protein (CRP) for heart disease, CA-125 for ovarian cancer, and CA 19-9 for colorectal and pancreatic cancer, are currently being used; however, they do not show sufficient sensitivity or specificity for early detection. Ideally, a unique biomarker for each disease should have the capacity for diagnosis of patients who do not have specific symptoms. Development of biomarker panels comprised of several proteins could result in improved detection and clinical management of various patients and provide higher sensitivities and specificities than a single protein biomarker.

### 6.5.2 Biomarker discovery based on clinical proteomics

**6.5.2.1 Sources for biomarkers:** Blood is in contact with virtually all cells of the organism; therefore, it is considered the optimal source for discovery of biomarkers. As a source of biomarkers, blood is easily accessible, and its collection is low risk and relatively non-invasive. Release of tissue-related proteins into the blood stream occurs through a specific secretory mechanisms, either shedding from the cell surface or nonspecific leakage.<sup>755</sup>

As an alternative to blood, other biofluids, including urine, cerebrospinal fluid, pancreatic juice, nipple aspirate fluid, and bile can be used. Urine is a common source for discovery of biomarkers (for urogenital disease as well as non-urogenital) due to its simple and non-invasive collection.<sup>756</sup> Urinary exosomes, endogenous microparticles identified in urine, have been shown to be a source for urine biomarker discovery using LC-based proteomics.<sup>757,758</sup> Cerebrospinal fluid can be used for discovery of biomarkers for disease of the central nervous system.<sup>759</sup> Pancreatic juice is used for non-invasive diagnosis of pancreatic cancer. Nipple aspirate fluid, found within the ductal and lobular system in non-pregnant and non-lactating women<sup>760</sup> has been used in proteomic analysis of breast cancer.<sup>761</sup> Proximity of bile to the biliary tract makes it important for study of cholangiocarcinoma, a primary cancer that arises from biliary epithelial cells; secreted or shed proteins in bile can be relevant for its study.<sup>762</sup>

Use of tumor tissues as a source of biomarkers is crucial for cancer biology. While the use of tumor tissue is limited by the small size of sample collected, the advantage of using clinical samples for biomarker discovery is that normal tissue from the surrounding area of the tumor can be used for paired analysis. Laser Capture Microdissection (LCM) has been employed to reduce the heterogeneity in human tumor tissues used in oncoproteomics studies.<sup>763</sup> Recent work to extract intact proteins from formalin fixed and paraffin-

embedded (FFPE) tissues has provided a huge potential for conduct of biomarker studies using millions of stored specimens.<sup>764</sup>

**6.5.2.2 Prefractionation and pre-treatment of clinical samples:** The complexity of the serum proteome presents a broad range of technical limitations for its use in clinical proteome analysis. Plasma proteins show a dynamic range of 10 orders of magnitude in concentration from albumin to the least abundant proteins.<sup>55</sup> Fortunately, clinical proteomics research has benefitted from development of various fractionation techniques and sensing approaches for management of the complexity of proteins and dynamic range in serum samples,<sup>765,766</sup> further described in the “Protein depletion and equalization methods” section. One strategy designed to overcome the issue of dynamic range coupled protein prefractionation with depletion methods for removal of abundant proteins in plasma proteomes. Several approaches can be used for removal of abundant proteins from plasma. One approach involves the use of immunoaffinity columns such as top-12 and top-14 IgY depletion columns.<sup>767</sup>

Most protein biomarkers used clinically, such as PSA, CA125, and CEA, are glycosylated.<sup>768,769,770,771</sup> Alteration of cellular glycosylation is known to occur in association with different disease states.<sup>772</sup> There has been a great deal of work in discovering new potential markers using the shotgun method based on glycopeptides/glycoproteins or glycans structures, especially in the cancer area. In the study of glycoproteins using proteomic approaches, various types of lectin chromatography are known to be capable of enriching glycosylated proteins from complex matrices<sup>387</sup> and a covalent method for capture using hydrazide is used for capture of glycoproteins/glycopeptides.<sup>773</sup> In addition, a combination of titanium dioxide chromatography with HILIC showed selective enrichment of sialic acid-containing glycopeptides.<sup>774</sup> Recently, MS-based analysis of glycopeptides generated from immobilized-Pronase digestion has been enhanced by sensitive and reproducible chip-based porous graphitized carbon (PGC) nano-LC.<sup>775</sup> Simultaneously, the shotgun method based on glycans structures has been developed to discover new potential markers, especially in the cancer area. Two methods, compositional mass profiling and structure-specific chromatographic profiling, are currently used for comprehensive analysis of glycans. For compositional mass profiling method, Lebrilla’s group has reported extensively on the application of a MALDI-MS based glycomics platform to discover glycans biomarkers of various cancers.<sup>776–778</sup> To differentiate or distinguish between glycans isomers, a method for structure-specific chromatographic profiling of glycans has been introduced<sup>779–780</sup> These studies utilize structure-sensitive chromatography, such as PGC and HILIC, to separate glycans online prior to MS analysis. Recently, a microfluidic chip packed with graphitized carbon was used to chromatographically separate the glycans and 300 N-glycan species were identified from serum samples of prostate cancer patients.<sup>781</sup> Meanwhile, O-glycan structures of MUC1, a cell-surface glycoprotein, in breast, prostate, and gastric cancer samples have been determined by LC-MS/MS.<sup>782</sup> Software to analyze tandem spectra of oligosaccharides has been developed. For N-linked glycans analysis, GlycosidIQ<sup>783</sup> including only known glycans in its library, and GlycoFragment/GlycoSearchMS<sup>784</sup> including hypothetical glycans structures have been developed. Recently, GlycoWorkbench<sup>785</sup> was designed specifically to assist those with greater knowledge of glycans MS with annotation of glycans fragment spectra. Moreover, Cartoonist Two<sup>786</sup> was developed for interpreting tandem spectra of O-linked glycans.

Different sample processing techniques, such as size fractionation, protein A/G depletion, and magnetic bead separation (C3, C8 and WCX), are known to be effective for reduction of the dynamic range and complexity of serum samples.<sup>787,788</sup> Evolution of a combination of separation, detection and labeling strategies has enhanced possibility for sensitive,

multiplexed detection of low-abundance disease-specific protein biomarkers in serum samples.

**6.5.2.3 Example of biomarkers identified for various diseases:** Proteomics has been used widely for identification of biomarkers of various diseases, including cancer, autoimmune disease, and infectious disease. In a gel-based proteomic study, increased expression of epidermal fatty acid binding protein 5, methylcrotonoyl-CoA carboxylase 2, palmitated protein A2, ezrin, and stomatin-like protein 2, and decreased expression of smooth muscle 22 were identified as diagnostic markers associated with lymph node metastasis in prostate cancer.<sup>789</sup> 2-D DIGE and MALDI-ToF MS were used for detection of potential serological biomarkers, S100A8 and S100A9 (Calgranulin A and B), in colorectal cancer.<sup>790</sup> In an experiment for discovery of biomarkers for head-and-neck squamous cell carcinoma, 4-plex iTRAQ was used for comparison of paired and non-paired non-cancerous samples with cancerous tissues.<sup>791</sup> iTRAQ has also been used to study of cancer angiogenesis, metastasis, epithelial mesenchymal transition, cancer therapy resistance, and cancer secretome. In another study, sera of patients with colorectal cancer were investigated using LC-MS/MS; growth/differentiation factor 15, divergent member of the transforming growth factor-beta superfamily, and trefoil factor 3, which is secreted from goblet cells of the gastrointestinal tract with mitogenic and antiapoptotic activity, were identified as potential biomarkers for early diagnosis of colorectal cancer.<sup>792</sup> Performance of serum profiling using proteomics has resulted in identification of biomarkers for many autoimmune diseases, including systemic lupus erythematosus (SLE),<sup>793</sup> rheumatoid arthritis (RA),<sup>794</sup> multiple sclerosis,<sup>795</sup> and Crohn's disease (CD).<sup>796</sup> In addition, use of serum profiling has also resulted in identification of biomarkers for many other human diseases, including stroke,<sup>797</sup> non-alcoholic fatty liver disease (NAFLD),<sup>798</sup> diabetic nephropathy,<sup>799</sup> Down syndrome (DS),<sup>800</sup> sarcoidosis,<sup>801</sup> and graft-versus-host disease.<sup>802</sup> The Van Eyk group has performed extensive work in the area of proteomics for predicting heart attacks.<sup>803-804</sup> In addition to cancers and autoimmune disorders, serum proteome analysis has also provided biomarkers for many infectious diseases, including tuberculosis,<sup>805</sup> severe acute respiratory syndrome (SARS),<sup>806</sup> and hepatitis.<sup>807</sup> Most of proteins listed in this section are putative biomarkers and have not been rigorously verified.

#### 6.5.2.4 Verification and validation of biomarkers<sup>3</sup>

**6.5.3**—MS based proteomics holds special promise for identification of biomarkers that can be used for blood tests. However, as shown in Figure 14, only a small number of blood-based biomarkers have been qualified or verified,<sup>808</sup> and even fewer have been validated (which requires larger groups of patients for testing of the proposed biomarker)<sup>809</sup> because current proteomics methods present several bottlenecks for analysis of large numbers of samples. Verification of biomarkers can require 100 – 1,000 samples, whereas validation of biomarkers requires analysis of even larger numbers (thousands to tens of thousands).<sup>808</sup> The same assays are used for verification and clinical validation of biomarkers.<sup>810</sup> Due to high throughput and high sensitivity for quantification of proteins, ELISA is regarded as a general assay method. However, development of an ELISA method for one biomarker candidate is expensive and requires more than one year, and specific antibodies are sometimes unavailable.<sup>811</sup> Due to its multiplexability, specificity, and sensitivity, MRM can also be used for validation, large-scale validation, or validation for subsequent analysis.<sup>812</sup> In addition, using MRM, reproducibility can be maintained across multiple laboratories. Monitoring initiated detection and sequencing-multiple reaction monitoring (MIDAS-MRM) allows for rapid and cost-effective absolute quantification and validation of biomarkers without the need for antibodies.<sup>813</sup> As examples of high-throughput MRM analysis using clinical samples, MRM-MS was applied to verify 12 target proteins for diabetic retinopathy using 3 groups of vitreous and plasma samples from 3 different stages of patients.<sup>814</sup> Kuzyk

*et al.* demonstrated the rigorous MRM quantification of the 45 endogenous plasma proteins with 31 of these being putative biomarkers of cardiovascular disease.<sup>815</sup> This study was done without sample enrichment or fractionation in a single 45 min MRM analysis, in combination with stable-isotope labeled peptide standards. To improve the limit of quantification, MS3 fragment ions were monitored instead of MS2 fragment ions for SRM transitions, termed MRM<sup>3</sup>.<sup>816</sup> It was demonstrated that this method could show 3- to 5-fold improvements in limit of detection using 5 model proteins and detect protein biomarkers at the low ng/mL level in nondepleted human serum.<sup>816</sup> In addition, coupling a nanospray ionization multicapillary inlet/dual electrodynamic ion funnel interface to a commercial triple quadrupole mass spectrometer improved ion transmission efficiency that increased by ~70-fold of average MRM peak intensities.<sup>817</sup> Various software related to MRMs has been developed to facilitate surrogate peptide determination and data processing. The commercial software include Pinpoint from Thermo Scientific and MultiQuant from AB Sciex for their TSQ and QTRAP mass spectrometers, and the open source software includes Skyline<sup>502</sup> and MRMer.<sup>818</sup> Methods such as SISCAPA (Stable Isotope Standards and Capture by Anti-Peptide Antibodies) and PRISM (high-pressure, high-resolution separations coupled with intelligent selection and multiplexing) have been introduced as alternative versions of MRMs. SISCAPA<sup>819–820</sup> uses a combination of bead-based affinity co-capture of native and SIS peptides, and their relative amounts are quantified by MRM-MS. SISCAPA has been shown to increase the sensitivity of MRM-MS by 1,800 and 18,000-fold for selected tryptic peptides of alpha(1)-antichymotrypsin and lipopolysaccharide-binding protein, respectively.<sup>821</sup> Recently, a group developed an antibody-free strategy that involves PRISM for sensitive SRM-based targeted protein quantification. This approach demonstrated accurate and reproducible quantification of proteins at concentrations in the 50–100 pg/mL range in human plasma/serum.<sup>822</sup> Prospects

Despite the discovery of many new potential markers using proteomics, only a few biomarkers have been developed into clinical applications for disease screening and patient monitoring and approved by the FDA.<sup>823</sup> OVA1, the first biomarker identified through proteomics, has recently been approved by the US FDA. OVA1 uses the *in vitro* diagnostic multivariate index assay (IVDMIA), which includes five biomarkers for assessment of ovarian cancer risk in women diagnosed with ovarian tumors prior to a planned surgery. The experience with development of the first FDA-cleared biomarker has demonstrated that the path from discovery, validation, to clinical use of a biomarker is long and arduous and that proper study design and development of assays that employ robust analytical methods suitable for large-scale validation are needed. Recent advances in MS-based proteomics technology with sensitive and accurate MS technology, innovative bioinformatics, and MRM methods will support each step in biomarker development, from discovery to clinical application.

## 7 Perspectives

Mass spectrometry-based methods to analyze and identify proteins are clearly a powerful approach to study biology. Initial uses of the technology simply involved common experiments such as analyzing proteins separated by gel electrophoresis. These experiments quickly evolved into approaches to directly identify proteins in complexes and then new types of experiments were developed such as methods to identify the topology of proteins embedded in lipid bilayers. As a powerful ‘hypothesis-generating engine’,<sup>19</sup> shotgun proteomics has shed light on many different biological processes. Large-scale studies, including both protein expression and PTM characterizations enable us to have a systematic picture of all the expressed proteins in biological networks. Additionally, targeted studies are gaining use as a way to answer specific biological questions using MS-based technology. Mass spectrometry complements traditional biochemistry and molecular biology techniques

for both biological discovery and validation and its role in biology has dramatically expanded.

Beyond the initial biological experiment the shotgun proteomic process consists of 4 steps and it is at each step where improvements in technology or methodology can be realized to improve the overall process. As shotgun proteomics requires the digestion of complex mixtures of proteins, the proteolysis of proteins is a crucial first step of the process. The success of a shotgun analysis is thus dependent on efficient and complete digestion of proteins as this can lead to high sequence coverage. Not surprisingly there have been several studies to optimize the process and to improve the analysis of recalcitrant proteins such as membrane proteins.<sup>824</sup> Membrane proteins are a class of proteins that can be difficult to digest because most of the protein is embedded in a lipid bilayer. This attribute of membrane proteins has been exploited by Blackler *et al* to better identify membrane proteins.<sup>825</sup> Proteinase K is used to digest exposed protein regions in the lipid bilayer and peripheral membrane proteins. Once the digestion is complete the solution is centrifuged to remove peptides from the solution. The transmembrane regions of the membrane proteins precipitate with the lipid bilayer. These polypeptides are then digested with cyanogen bromide since many of these regions have methionine located in these segments. As these peptides are often hydrophobic, Blackler and Wu showed much improved separation and recovery of these peptides by using liquid chromatography at elevated temperature.<sup>184</sup> The interesting aspect of this method was the improved specificity for detection of membrane proteins, which can be difficult to achieve as proteins can stick to the lipid bilayer or get denatured during the cell lysis procedure. Methods have also appeared to improve the digestion of proteins or to preferentially digest and deplete proteins based on abundance.<sup>65,81</sup> The state of sample preparation is good, but there can always be improvements especially on methods to improve digestion of low abundance proteins.

A critical function of peptide separations is to reduce ion suppression, increase dynamic range, and minimize isobaric interferences. The digestion of whole cell lysates can potentially produce 100,000's of peptides over a wide range of abundances, isoelectric points, and hydrophobicities. These peptide mixtures need to be separated to avoid the simultaneous introduction of large numbers of peptides into the mass spectrometer, which would compromise dynamic range and recovery by suppressing the ionization of peptides. To this end, much emphasis has been placed on creating separation strategies to fractionate complex mixtures of peptides. Most strategies combine different separation techniques together to create multidimensional methods although reversed-phase HPLC is generally the method used to introduce peptides into the mass spectrometer. Offline strategies of gel electrophoresis, isoelectric focusing, chromatofocusing and HPLC have been used at the protein level to reduce the complexity of protein mixtures. Isoelectric focusing and ion exchange methods have been employed to prefractionate peptides in an offline manner. Ion exchange has been combined with reversed-phase HPLC to effect multidimensional separations both online and offline. While efforts to push single dimension separations to achieve sufficient resolution and peak capacity to resolve 100,000's of peptides are underway, they are unlikely to achieve the peak capacity of multidimensional separations.

Despite the maturity of mass analyzers the last decade has seen significant innovation. Time-of-Flight mass analyzers have continued to evolve in terms of resolution and scan speed. MALDI-ToF methods have seen a drop in use in proteomics in favor of tandem MS methods in which the ToF is often combined with quadrupole mass filters. The last decade saw significant innovations in the development of the linear ion trap, which increased the ion capacity of ion traps and improved performance. The big innovation for proteomics was the optimization and development of the Orbitrap mass spectrometer.<sup>241,826</sup> While this device has been around for many years, it took the invention of methods to inject ions into

the device and the development of Fourier transform methods to record ion currents to create a powerful high resolution, high mass accuracy mass analyzer. The other older mass analyzer experiencing a renaissance was the ion mobility device.<sup>827</sup> Ion mobility separates by both  $m/z$  and molecular shape and thus can be useful to provide a separation method prior to mass analysis. It is potentially a good strategy to fractionate ion populations using a gas phase method prior to tandem mass spectrometry of individual ions. A clear advantage to doing this by ion mobility is the much faster time frames for ion separation versus chromatographic methods, but fractionation of the ion population has to occur with reasonable resolution to minimize ion dilution. Ultimately, tandem mass spectrometers should be able to completely sample the population of peptide ions in a reasonable amount of time.

An essential component of shotgun proteomics is software to interpret the data. The basic framework for large-scale analysis of tandem mass spectrometry data is completed. Software tools are available, efficient and reasonably accurate for the analysis of shotgun proteomic data.<sup>543</sup> There is concern that some tools over-fit data or use filtering criteria that don't meet community standards, but in general results are excellent when community standards are followed.<sup>44,828</sup> Additional capabilities will derive from software tools that aid in the identification of mutations, *de novo* analysis, and the identification of heavily modified peptides. As dissociation methods are better able to fragment larger polypeptides, tools to dissect out the combinations and patterns of modifications will improve. Software tools for the analysis of quantitative data are improving and many tools are able to analyze most of the different quantitative strategies that have been developed. An area that will see more activity is data-independent acquisition (DIA).<sup>233,829</sup> As detection limits improve in mass spectrometers it is increasingly difficult to cleanly isolate and fragment a unique peptide ion (isobaric interferences) and consequently methods have been developed to acquire multiple precursor ions simultaneously for fragmentation. Software tools to deconvolute and analyze the data have been developed, but more tools are needed especially as people innovate with new methods and experiments for DIA.

As mass spectrometers have improved for the analysis of intact proteins, increasingly shotgun proteomics will be combined with "top down" analysis to develop a more complete picture of the proteoforms for a protein. The dissociation methods such as ETD and ECD will allow the fragmentation of larger polypeptides, which can fill the middle space between bottom up and top down. In the past decade, the development of mass spectrometers designed for peptide/protein complex analysis has advanced rapidly, mostly in mass accuracy and sensitivity. However, there is still room for further improvement. Deep proteome discovery will require enhanced signal sensitivity for low abundant peptides in samples with large dynamic ranges. Direct detection of large or modified peptides/proteins will require innovations in fragmentation and ionization methods to improve spectra quality. In addition to improvements to the mass spectrometer, advances in the techniques involved in sample preparation, fractionation, or enrichment prior to MS and data processing after MS will also enhance the role of proteomics in biology.

## Acknowledgments

We gratefully thank Claire Delahunty and Xuemei Han for their critical review of this manuscript and constructive comments and suggestions. Funding was provided by the following NIH grants P41 GM103533/P41 RR011823, R01 DK074798, R01 HL079442, and R01 MH067880.

## Biographies

Author Photos and Biographies.



Yaoyang Zhang was born in Zhengzhou, a city in central China, in 1981. Yaoyang completed his bachelor and master degrees from Beijing Institute of Technology in 2006, under supervision of Professor Yulin Deng. He then spent four years at Max-Planck Institute of Psychiatry in Munich with Professor Christoph W. Turck, and obtained his Ph.D. in 2010. Yaoyang joined the Yates lab as a research associate at The Scripps Research Institute in 2011. His research interests include using mass spectrometry in quantitative proteomics, protein-protein interaction studies, detection of protein post-translational modifications, and protein turnover analysis.



Bryan R. Fonslow studied chemistry (B.S.) at Michigan Technological University. He then received his Ph.D. in analytical chemistry from the University of Minnesota where he worked with Professor Michael T. Bowser developing micro free-flow electrophoresis for bioanalytical separations and assays. His interest in applying mass spectrometry to bioanalytical separations led him to a postdoctoral position in the Yates lab at The Scripps Research Institute. His current work focuses on applying capillary separations and mass spectrometry to the analysis of protein complexes and post-translational modifications involved in the biology of aging.



Bing Shan was born in China in 1981. She obtained her Ph.D. in biology from University of Munich in 2010, and then joined in the Yates lab as a postdoctoral fellow at the Scripps Research Institute in 2011. Currently she focuses on proteomic studies in heart and neurodegenerative diseases.



Moon-Chang Baek received a Ph.D. in Pharmacy at Seoul National University in 1997 and has held postdoctoral fellow positions in the Department of Biological Chemistry and Molecular Pharmacology at Harvard Medical School (1998–2002) as well as in the Department of Life Science at POSTECH (2002–2005). During the periods, he has studied on traditional proteomics using various mass spectrometries and also developed a fusion technology bridging between cell biology and proteomics. He is currently an associate professor at School of Medicine of Kyungpook National University, Korea and has focused on the identification of biomarker candidates for kidney diseases, drug-induced liver injury, and cancer diseases by developing various proteomics technologies including glycoproteomics, membrane proteomics as well as enrichment technologies for clinic specimen (especially microvesicles) analysis.



Dr. Yates is a Professor in the Department of Chemical Physiology at The Scripps Research Institute. His research interests include development of integrated methods for tandem mass spectrometry analysis of protein mixtures, bioinformatics using mass spectrometry data, and biological studies involving proteomics. He is the lead inventor of the SEQUEST software for correlating tandem mass spectrometry data to sequences in the database and principle developer of the shotgun proteomics technique for the analysis of protein mixtures. His laboratory has developed the use of proteomic techniques to analyze protein complexes, posttranslational modifications, organelles and quantitative analysis of protein expression for the discovery of new biology. Many proteomic approaches developed by Yates have become a national and international resource to many investigators in the scientific community. Dr. Yates led an NIDCR funded Center to characterize the Saliva proteome and has been involved in an NCRF funded Research Resource Center for the last 15 years and was involved in an NSF funded Science and Technology Center for 10 years. He has received the American Society for Mass Spectrometry research award, the Pehr Edman Award in Protein Chemistry, the American Society for Mass Spectrometry Biemann Medal, the HUPO Distinguished Achievement Award in Proteomics, Herbert Sober Award from the ASBMB, and the Christian Anfinsen Award from The Protein Society. He has published 580 scientific articles.

## 9 Abbreviations

<b>2D-PAGE</b>	two dimensional polyacrylamide gel electrophoresis
<b>APEX</b>	absolute protein expression

<b>AP-MS</b>	affinity purification and mass spectrometry
<b>AQUA</b>	absolute quantification
<b>BD</b>	binding domain
<b>CID/CAD</b>	collision-induced/assisted dissociation
<b>CID</b>	collision-induced dissociation
<b>DDA</b>	data-dependent analysis
<b>DIA</b>	data-independent analysis
<b>DIGE</b>	difference gel electrophoresis
<b>ECD</b>	electron capture dissociation
<b>ESI</b>	electrospray ionization
<b>ETcaD</b>	ETD with supplementary CID fragmentation
<b>ETD</b>	electron transfer dissociation
<b>ETnoD</b>	non-dissociative electron transfer
<b>FDR</b>	false discovery rate
<b>FT-ICR</b>	fourier transform ion cyclotron resonance
<b>HAP</b>	hydroxyapatite
<b>HCD</b>	higher-energy collisional dissociation
<b>HILIC</b>	hydrophilic interaction liquid chromatography
<b>IAA</b>	iodoacetamide
<b>ICAT</b>	isotope-coded affinity tag
<b>ICPL</b>	isotope-coded protein label
<b>IEF</b>	isoelectric focusing
<b>IMAC</b>	ion affinity chromatography
<b>IPG</b>	immobilized pH gradient
<b>iTRAQ</b>	isobaric tags for relative and absolute quantitation
<b>LC-MS</b>	liquid chromatography-mass spectrometry
<b>MRM</b>	multiple reaction monitoring
<b>MudPIT</b>	Multidimensional protein identification technology
<b>nLC</b>	nanoflow liquid chromatography
<b>NMR</b>	nuclear magnetic resonance
<b>PQD</b>	pulsed Q dissociation
<b>PSAQ</b>	intact isotope labeled protein standard absolute quantification
<b>PSM</b>	peptide spectrum match
<b>PTMs</b>	post-translational modifications
<b>Q</b>	quadrupole
<b>QQQ</b>	triple quadrupole

<b>RIA</b>	relative isotope abundance
<b>SCX</b>	strong-cation exchange
<b>SILAC</b>	stable isotope labeling with amino acids in cell culture
<b>SILAM</b>	stable isotope labeling of amino acids in mammals
<b>SRM</b>	selected reaction monitoring
<b>TAP</b>	tandem affinity purification
<b>TEV</b>	tobacco etch virus
<b>TF</b>	transcription factors
<b>TMT</b>	tandem mass tag
<b>ToF</b>	time of flight
<b>Ub</b>	ubiquitin
<b>UBDs</b>	Ub binding domains
<b>UBLs</b>	ubiquitin-like proteins
<b>UPLC</b>	ultrahigh pressure liquid chromatography
<b>WAX</b>	weak anion exchange
<b>WGA</b>	wheat germ agglutinin
<b>Y2H</b>	yeast two-hybrid

## References

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczy J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissole SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M. *Nature*. 2001; 409:860. [PubMed: 11237011]
2. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C. *Science*. 2001; 291:1304. [PubMed: 11181995]

3. Wilkins MR, Pasquali C, Appel RD, Ou K, Golaz O, Sanchez JC, Yan JX, Gooley AA, Hughes G, Humphery-Smith I, Williams KL, Hochstrasser DF. *Biotechnology (N Y)*. 1996; 14:61. [PubMed: 9636313]
4. James P. *Q Rev Biophys*. 1997; 30:279. [PubMed: 9634650]
5. Gstaiger M, Aebersold R. *Nat Rev Genet*. 2009; 10:617. [PubMed: 19687803]
6. Makarov A, Denisov E, Kholomeev A, Balschun W, Lange O, Strupat K, Horning S. *Anal Chem*. 2006; 78:2113. [PubMed: 16579588]
7. Makarov A. *Anal Chem*. 2000; 72:1156. [PubMed: 10740853]
8. Wolters DA, Washburn MP, Yates JR 3rd. *Anal Chem*. 2001; 73:5683. [PubMed: 11774908]
9. Yates JR 3rd. *Annu Rev Biophys Biomol Struct*. 2004; 33:297. [PubMed: 15139815]
10. Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, Yates JR 3rd. *Nat Biotechnol*. 1999; 17:676. [PubMed: 10404161]
11. Yates JR 3rd. *J Mass Spectrom*. 1998; 33:1. [PubMed: 9449829]
12. Han X, Jin M, Breuker K, McLafferty FW. *Science*. 2006; 314:109. [PubMed: 17023655]
13. Tran JC, Zamdborg L, Ahlf DR, Lee JE, Catherman AD, Durbin KR, Tipton JD, Vellaichamy A, Kellie JF, Li M, Wu C, Sweet SM, Early BP, Siuti N, LeDuc RD, Compton PD, Thomas PM, Kelleher NL. *Nature*. 2011; 480:254. [PubMed: 22037311]
14. Wu C, Tran JC, Zamdborg L, Durbin KR, Li M, Ahlf DR, Early BP, Thomas PM, Sweedler JV, Kelleher NL. *Nat Methods*. 2012; 9:822. [PubMed: 22706673]
15. Yates JR. *Nat Methods*. 2011; 8:633.
16. Ong SE, Mann M. *Nat Chem Biol*. 2005; 1:252. [PubMed: 16408053]
17. Aebersold R, Mann M. *Nature*. 2003; 422:198. [PubMed: 12634793]
18. McDonald WH, Yates JR 3rd. *Curr Opin Mol Ther*. 2003; 5:302. [PubMed: 12870441]
19. Cravatt BF, Simon GM, Yates JR 3rd. *Nature*. 2007; 450:991. [PubMed: 18075578]
20. Han X, Aslanian A, Yates JR 3rd. *Curr Opin Chem Biol*. 2008; 12:483. [PubMed: 18718552]
21. Yates JR, Ruse CI, Nakorchevsky A. *Annu Rev Biomed Eng*. 2009; 11:49. [PubMed: 19400705]
22. Ahrens CH, Brunner E, Qeli E, Basler K, Aebersold R. *Nat Rev Mol Cell Biol*. 2010; 11:789. [PubMed: 20944666]
23. Choudhary C, Mann M. *Nat Rev Mol Cell Biol*. 2010; 11:427. [PubMed: 20461098]
24. Britton LM, Gonzales-Cope M, Zee BM, Garcia BA. *Expert Rev Proteomics*. 2011; 8:631. [PubMed: 21999833]
25. Bensimon A, Heck AJ, Aebersold R. *Annu Rev Biochem*. 2012; 81:379. [PubMed: 22439968]
26. Gingras AC, Raught B. *FEBS Lett*. 2012; 586:2723. [PubMed: 22710165]
27. Mann M, Kelleher NL. *Proc Natl Acad Sci U S A*. 2008; 105:18132. [PubMed: 18818311]
28. Boersema PJ, Mohammed S, Heck AJ. *J Mass Spectrom*. 2009; 44:861. [PubMed: 19504542]
29. Cox J, Mann M. *Annu Rev Biochem*. 2011; 80:273. [PubMed: 21548781]
30. Angel TE, Aryal UK, Hengel SM, Baker ES, Kelly RT, Robinson EW, Smith RD. *Chem Soc Rev*. 2012; 41:3912. [PubMed: 22498958]
31. Di Palma S, Hennrich ML, Heck AJ, Mohammed S. *J Proteomics*. 2012; 75:3791. [PubMed: 22561838]
32. Ramautar R, Heemskerk AA, Hensbergen PJ, Deelder AM, Busnel JM, Mayboroda OA. *J Proteomics*. 2012; 75:3814. [PubMed: 22609513]
33. Stergachis AB, MacLean B, Lee K, Stamatoyannopoulos JA, MacCoss MJ. *Nat Methods*. 2011; 8:1041. [PubMed: 22056677]
34. Jiang X, Ye M, Liu G, Feng S, Cui L, Zou H. *J Proteome Res*. 2007; 6:2287. [PubMed: 17488005]
35. Mosley AL, Florens L, Wen Z, Washburn MP. *J Proteomics*. 2009; 72:110. [PubMed: 19038371]
36. Gauci S, Veenhoff LM, Heck AJ, Krijgsveld J. *J Proteome Res*. 2009; 8:3451. [PubMed: 19453177]
37. Hwang SI, Lundgren DH, Mayya V, Rezaul K, Cowan AE, Eng JK, Han DK. *Mol Cell Proteomics*. 2006; 5:1131. [PubMed: 16540461]

38. Emmott E, Rodgers MA, Macdonald A, McCrory S, Ajuh P, Hiscox JA. *Mol Cell Proteomics*. 2010; 9:1920. [PubMed: 20467043]
39. Moulder R, Lonnberg T, Elo LL, Filen JJ, Rainio E, Corthals G, Oresic M, Nyman TA, Aittokallio T, Lahesmaa R. *Mol Cell Proteomics*. 2010; 9:1937. [PubMed: 20467038]
40. Foster LJ, de Hoog CL, Zhang Y, Xie X, Mootha VK, Mann M. *Cell*. 2006; 125:187. [PubMed: 16615899]
41. Emmott E, Wise H, Loucaides EM, Matthews DA, Digard P, Hiscox JA. *J Proteome Res*. 2010; 9:5335. [PubMed: 20701360]
42. Li J, Cai T, Wu P, Cui Z, Chen X, Hou J, Xie Z, Xue P, Shi L, Liu P, Yates JR 3rd, Yang F. *Proteomics*. 2009; 9:4539. [PubMed: 19670372]
43. Jiang XS, Zhou H, Zhang L, Sheng QH, Li SJ, Li L, Hao P, Li YX, Xia QC, Wu JR, Zeng R. *Mol Cell Proteomics*. 2004; 3:441. [PubMed: 14739257]
44. Rezaul K, Wu L, Mayya V, Hwang SI, Han D. *Mol Cell Proteomics*. 2005; 4:169. [PubMed: 15598749]
45. Deng N, Zhang J, Zong C, Wang Y, Lu H, Yang P, Wang W, Young GW, Korge P, Lotz C, Doran P, Liem DA, Apweiler R, Weiss JN, Duan H, Ping P. *Mol Cell Proteomics*. 2011; 10:M110000117. [PubMed: 20495213]
46. Meany DL, Xie H, Thompson LV, Arriaga EA, Griffin TJ. *Proteomics*. 2007; 7:1150. [PubMed: 17390297]
47. Forner F, Foster LJ, Campanaro S, Valle G, Mann M. *Mol Cell Proteomics*. 2006; 5:608. [PubMed: 16415296]
48. Warren CM, Geenen DL, Helseth DL Jr, Xu H, Solaro RJ. *J Proteomics*. 2010; 73:1551. [PubMed: 20394843]
49. Li K, Hornshaw MP, van Minnen J, Smalla KH, Gundelfinger ED, Smit AB. *J Proteome Res*. 2005; 4:725. [PubMed: 15952719]
50. Cheng D, Hoogenraad CC, Rush J, Ramm E, Schlager MA, Duong DM, Xu P, Wijayawardana SR, Hanfelt J, Nakagawa T, Sheng M, Peng J. *Mol Cell Proteomics*. 2006; 5:1158. [PubMed: 16507876]
51. Li KW, Miller S, Klychnikov O, Loos M, Stahl-Zeng J, Spijker S, Mayford M, Smit AB. *J Proteome Res*. 2007; 6:3127. [PubMed: 17625814]
52. McClatchy DB, Liao L, Park SK, Venable JD, Yates JR. *Genome Res*. 2007; 17:1378. [PubMed: 17675365]
53. McClatchy DB, Liao L, Lee JH, Park SK, Yates JR 3rd. *J Proteome Res*. 2012; 11:2467. [PubMed: 22397461]
54. Muntou RP, Tweedie-Cullen R, Livingstone-Zatchej M, Weinandy F, Waidelich M, Longo D, Gehrig P, Potthast F, Rutishauser D, Gerrits B, Panse C, Schlapbach R, Mansuy IM. *Mol Cell Proteomics*. 2007; 6:283. [PubMed: 17114649]
55. Anderson NL, Anderson NG. *Mol Cell Proteomics*. 2002; 1:845. [PubMed: 12488461]
56. Colantonio DA, Dunkinson C, Bovenkamp DE, Van Eyk JE. *Proteomics*. 2005; 5:3831. [PubMed: 16145708]
57. Kay R, Barton C, Ratcliffe L, Matharoo-Ball B, Brown P, Roberts J, Teale P, Creaser C. *Rapid Commun Mass Spectrom*. 2008; 22:3255. [PubMed: 18803344]
58. Warder SE, Tucker LA, Strelitzer TJ, McKeegan EM, Meuth JL, Jung PM, Saraf A, Singh B, Lai-Zhang J, Gagne G, Rogers JC. *Anal Biochem*. 2009; 387:184. [PubMed: 19454248]
59. Mahn A, Ismail M. *J Chromatogr B Analyt Technol Biomed Life Sci*. 2011; 879:3645.
60. Pieper R, Su Q, Gatlin CL, Huang ST, Anderson NL, Steiner S. *Proteomics*. 2003; 3:422. [PubMed: 12687610]
61. Tu C, Rudnick PA, Martinez MY, Cheek KL, Stein SE, Slebos RJ, Liebler DC. *J Proteome Res*. 2010; 9:4982. [PubMed: 20677825]
62. Thulasiraman V, Lin S, Gheorghiu L, Lathrop J, Lomas L, Hammond D, Boschetti E. *Electrophoresis*. 2005; 26:3561. [PubMed: 16167368]
63. Li L, Sun C, Freeby S, Yee D, Kieffer-Jaquinod S, Guerrier L, Boschetti E, Lomas L. *J Proteomics Bioinform*. 2009; 2:485.

64. Fonslow BR, Carvalho PC, Academia K, Freeby S, Xu T, Nakorchevsky A, Paulus A, Yates JR 3rd. *J Proteome Res.* 2011; 10:3690. [PubMed: 21702434]
65. Fonslow BR, Stein BD, Webb KJ, Xu T, Choi J, Park SK, Yates JR 3rd. *Nat Methods.* 2013; 10:54. [PubMed: 23160281]
66. Keil, B. *Specificity of proteolysis.* Springer-Verlag; Berlin ; New York: 1992.
67. Siepen JA, Keevil EJ, Knight D, Hubbard SJ. *J Proteome Res.* 2007; 6:399. [PubMed: 17203985]
68. Thiede B, Lamer S, Mattow J, Siejak F, Dimmler C, Rudel T, Jungblut PR. *Rapid Commun Mass Spectrom.* 2000; 14:496. [PubMed: 10717661]
69. MacCoss MJ, McDonald WH, Saraf A, Sadygov R, Clark JM, Tasto JJ, Gould KL, Wolters D, Washburn M, Weiss A, Clark JI, Yates JR 3rd. *Proc Natl Acad Sci U S A.* 2002; 99:7900. [PubMed: 12060738]
70. Xu T, Wong CC, Kashina A, Yates JR 3rd. *Nat Protoc.* 2009; 4:325. [PubMed: 19229197]
71. Choudhary G, Wu SL, Shieh P, Hancock WS. *J Proteome Res.* 2003; 2:59. [PubMed: 12643544]
72. Swaney DL, Wenger CD, Coon JJ. *J Proteome Res.* 2010; 9:1323. [PubMed: 20113005]
73. Klammer AA, MacCoss MJ. *J Proteome Res.* 2006; 5:695. [PubMed: 16512685]
74. Tran BQ, Hernandez C, Waridel P, Potts A, Barblan J, Lisacek F, Quadroni M. *J Proteome Res.* 2011; 10:800. [PubMed: 21166477]
75. Bian Y, Ye M, Song C, Cheng K, Wang C, Wei X, Zhu J, Chen R, Wang F, Zou H. *J Proteome Res.* 2012; 11:2828. [PubMed: 22468782]
76. Gilmore JM, Kettenbach AN, Gerber SA. *Anal Bioanal Chem.* 2012; 402:711. [PubMed: 22002561]
77. Gauci S, Helbig AO, Slijper M, Krijgsveld J, Heck AJ, Mohammed S. *Anal Chem.* 2009; 81:4493. [PubMed: 19413330]
78. Strader MB, Tabb DL, Hervey WJ, Pan C, Hurst GB. *Anal Chem.* 2006; 78:125. [PubMed: 16383319]
79. Hervey, WJt; Strader, MB.; Hurst, GB. *J Proteome Res.* 2007; 6:3054. [PubMed: 17616116]
80. Chen EI, Cociorva D, Norris JL, Yates JR 3rd. *J Proteome Res.* 2007; 6:2529. [PubMed: 17530876]
81. Manza LL, Stamer SL, Ham AJ, Codreanu SG, Liebler DC. *Proteomics.* 2005; 5:1742. [PubMed: 15761957]
82. Wisniewski JR, Zougman A, Nagaraj N, Mann M. *Nat Methods.* 2009; 6:359. [PubMed: 19377485]
83. Zhou JY, Dann GP, Shi T, Wang L, Gao X, Su D, Nicora CD, Shukla AK, Moore RJ, Liu T, Camp DG 2nd, Smith RD, Qian WJ. *Anal Chem.* 2012; 84:2862. [PubMed: 22339560]
84. Chen EI, McClatchy D, Park SK, Yates JR 3rd. *Anal Chem.* 2008; 80:8694. [PubMed: 18937422]
85. Kadiyala CS, Tomechko SE, Miyagi M. *PLoS One.* 2010; 5:e15332. [PubMed: 21209883]
86. Sun L, Tao D, Han B, Ma J, Zhu G, Liang Z, Shan Y, Zhang L, Zhang Y. *Anal Bioanal Chem.* 2011; 399:3387. [PubMed: 21079928]
87. Wang H, Qian WJ, Mottaz HM, Clauss TR, Anderson DJ, Moore RJ, Camp DG 2nd, Khan AH, Sforza DM, Pallavicini M, Smith DJ, Smith RD. *J Proteome Res.* 2005; 4:2397. [PubMed: 16335993]
88. Wang N, Xu M, Wang P, Li L. *Anal Chem.* 2010; 82:2262. [PubMed: 20092350]
89. Chatterjee D, Ytterberg AJ, Son SU, Loo JA, Garrell RL. *Anal Chem.* 2010; 82:2095. [PubMed: 20146460]
90. Moon H, Wheeler AR, Garrell RL, Loo JA, Kim CJ. *Lab Chip.* 2006; 6:1213. [PubMed: 16929401]
91. Nelson WC, Peng I, Lee GA, Loo JA, Garrell RL, Kim CJ. *Anal Chem.* 2010; 82:9932. [PubMed: 21058643]
92. Wheeler AR, Moon H, Bird CA, Loo RR, Kim CJ, Loo JA, Garrell RL. *Anal Chem.* 2005; 77:534. [PubMed: 15649050]
93. Wheeler AR, Moon H, Kim CJ, Loo JA, Garrell RL. *Anal Chem.* 2004; 76:4833. [PubMed: 15307795]

94. Freije JR, Mulder PP, Werkman W, Rieux L, Niederlander HA, Verpoorte E, Bischoff R. *J Proteome Res.* 2005; 4:1805. [PubMed: 16212436]
95. Ethier M, Hou W, Duewel HS, Figeys D. *J Proteome Res.* 2006; 5:2754. [PubMed: 17022646]
96. Guo W, Bi H, Qiao L, Wan J, Qian K, Girault HH, Liu B. *Mol Biosyst.* 2011; 7:2890. [PubMed: 21804973]
97. Qin W, Song Z, Fan C, Zhang W, Cai Y, Zhang Y, Qian X. *Anal Chem.* 2012; 84:3138. [PubMed: 22413971]
98. Lopez-Ferrer D, Capelo JL, Vazquez J. *J Proteome Res.* 2005; 4:1569. [PubMed: 16212408]
99. Lopez-Ferrer D, Heibeck TH, Petritis K, Hixson KK, Qian W, Monroe ME, Mayampurath A, Moore RJ, Belov ME, Camp DG 2nd, Smith RD. *J Proteome Res.* 2008; 7:3860. [PubMed: 18686986]
100. Sun W, Gao S, Wang L, Chen Y, Wu S, Wang X, Zheng D, Gao Y. *Mol Cell Proteomics.* 2006; 5:769. [PubMed: 16339992]
101. Segu ZM, Hammad LA, Mechref Y. *Rapid Commun Mass Spectrom.* 2010; 24:3461. [PubMed: 21072803]
102. Ye X, Li L. *Anal Chem.* 2012; 84:6181. [PubMed: 22708679]
103. Ye X, Li L. *Anal Chem.* 2012; 84:6181. [PubMed: 22708679]
104. Lin S, Yao G, Qi D, Li Y, Deng C, Yang P, Zhang X. *Anal Chem.* 2008; 80:3655. [PubMed: 18407620]
105. Chen WY, Chen YC. *Anal Chem.* 2007; 79:2394. [PubMed: 17284012]
106. Swatkoski S, Gutierrez P, Ginter J, Petrov A, Dinman JD, Edwards N, Fenselau C. *J Proteome Res.* 2007; 6:4525. [PubMed: 17902642]
107. Basile F, Hauser N. *Anal Chem.* 2011; 83:359. [PubMed: 21138252]
108. Lee AM, Sevinsky JR, Bundy JL, Grunden AM, Stephenson JL Jr. *J Proteome Res.* 2009; 8:3844. [PubMed: 19425607]
109. Lopez-Ferrer D, Petritis K, Hixson KK, Heibeck TH, Moore RJ, Belov ME, Camp DG 2nd, Smith RD. *J Proteome Res.* 2008; 7:3276. [PubMed: 18605748]
110. Lopez-Ferrer D, Petritis K, Lourette NM, Clowers B, Hixson KK, Heibeck T, Prior DC, Pasatolic L, Camp DG 2nd, Belov ME, Smith RD. *Anal Chem.* 2008; 80:8930. [PubMed: 19551971]
111. Klose J. *Humangenetik.* 1975; 26:231. [PubMed: 1093965]
112. O'Farrell PH. *J Biol Chem.* 1975; 250:4007. [PubMed: 236308]
113. Banks RE, Dunn MJ, Hochstrasser DF, Sanchez JC, Blackstock W, Pappin DJ, Selby PJ. *Lancet.* 2000; 356:1749. [PubMed: 11095271]
114. Perdew GH, Schaup HW, Selivonchick DP. *Anal Biochem.* 1983; 135:453. [PubMed: 6660519]
115. Rabilloud T, Adessi C, Giraudel A, Lunardi J. *Electrophoresis.* 1997; 18:307. [PubMed: 9150907]
116. Wildgruber R, Reil G, Drews O, Parlar H, Gorg A. *Proteomics.* 2002; 2:727. [PubMed: 12112855]
117. Ghiggeri GM, Candiano G, Delfino G, Queirolo C, Vecchio G, Gianazza E, Righetti PG. *Carbohydr Res.* 1985; 145:113. [PubMed: 4092214]
118. Wildgruber R, Harder A, Obermaier C, Boguth G, Weiss W, Fey SJ, Larsen PM, Gorg A. *Electrophoresis.* 2000; 21:2610. [PubMed: 10949137]
119. Chevalier F, Centeno D, Rofidal V, Tauzin M, Martin O, Sommerer N, Rossignol M. *J Proteome Res.* 2006; 5:512. [PubMed: 16512665]
120. Chevallet M, Luche S, Diemer H, Strub JM, Van Dorsselaer A, Rabilloud T. *Proteomics.* 2008; 8:4853. [PubMed: 19003863]
121. Agrawal GK, Thelen JJ. *Methods Mol Biol.* 2009; 527:3. [PubMed: 19241001]
122. Sondej M, Denny PA, Xie Y, Ramachandran P, Si Y, Takashima J, Shi W, Wong DT, Loo JA, Denny PC. *Clin Proteomics.* 2009; 5:52. [PubMed: 20161393]
123. Lopez JL. *J Chromatogr B Analyt Technol Biomed Life Sci.* 2007; 849:190.
124. Unlu M, Morgan ME, Minden JS. *Electrophoresis.* 1997; 18:2071. [PubMed: 9420172]
125. Bredemeyer AJ, Lewis RM, Malone JP, Davis AE, Gross J, Townsend RR, Ley TJ. *Proc Natl Acad Sci U S A.* 2004; 101:11785. [PubMed: 15280543]

126. Lee AY, Park BC, Jang M, Cho S, Lee DH, Lee SC, Myung PK, Park SG. *Proteomics*. 2004; 4:3429. [PubMed: 15468300]
127. Hecker M, Antelmann H, Buttner K, Bernhardt J. *Proteomics*. 2008; 8:4958. [PubMed: 19003856]
128. Mahairas GG, Sabo PJ, Hickey MJ, Singh DC, Stover CK. *J Bacteriol*. 1996; 178:1274. [PubMed: 8631702]
129. Afjehi-Sadat L, Lubec G. *Nat Protoc*. 2007; 2:2318. [PubMed: 17947972]
130. Laemmli UK. *Nature*. 1970; 227:680. [PubMed: 5432063]
131. Shevchenko A, Wilm M, Vorm O, Mann M. *Anal Chem*. 1996; 68:850. [PubMed: 8779443]
132. Lundby A, Olsen JV. *Methods Mol Biol*. 2011; 753:143. [PubMed: 21604121]
133. Michel PE, Reymond F, Arnaud IL, Jossierand J, Girault HH, Rossier JS. *Electrophoresis*. 2003; 24:3. [PubMed: 12652567]
134. Heller M, Michel PE, Morier P, Crettaz D, Wenz C, Tissot JD, Reymond F, Rossier JS. *Electrophoresis*. 2005; 26:1174. [PubMed: 15704250]
135. Wang H, Chang-Wong T, Tang HY, Speicher DW. *J Proteome Res*. 2010; 9:1032. [PubMed: 20014860]
136. Moritz RL, Ji H, Schutz F, Connolly LM, Kapp EA, Speed TP, Simpson RJ. *Anal Chem*. 2004; 76:4811. [PubMed: 15307793]
137. Moritz RL, Clippingdale AB, Kapp EA, Eddes JS, Ji H, Gilbert S, Connolly LM, Simpson RJ. *Proteomics*. 2005; 5:3402. [PubMed: 16052629]
138. Chen EI, Hewel J, Felding-Habermann B, Yates JR 3rd. *Mol Cell Proteomics*. 2006; 5:53. [PubMed: 16272560]
139. Garbis SD, Roumeliotis TI, Tyritzis SI, Zorpas KM, Pavlakis K, Constantinides CA. *Anal Chem*. 2011; 83:708. [PubMed: 21174401]
140. Wang H, Clouthier SG, Galchev V, Misek DE, Duffner U, Min CK, Zhao R, Tra J, Omenn GS, Ferrara JL, Hanash SM. *Mol Cell Proteomics*. 2005; 4:618. [PubMed: 15703445]
141. Schlautman JD, Rozek W, Stetler R, Mosley RL, Gendelman HE, Ciborowski P. *Proteome Sci*. 2008; 6:26. [PubMed: 18789151]
142. Wilm M, Mann M. *Anal Chem*. 1996; 68:1. [PubMed: 8779426]
143. Hillenkamp F, Karas M. *Methods Enzymol*. 1990; 193:280. [PubMed: 1963669]
144. Cramer R, Gobom J, Nordhoff E. *Expert Rev Proteomics*. 2005; 2:407. [PubMed: 16000086]
145. Ogorzalek Loo RR, Stevenson TI, Mitchell C, Loo JA, Andrews PC. *Anal Chem*. 1996; 68:1910. [PubMed: 21619102]
146. Ogorzalek Loo RR, Mitchell C, Stevenson TI, Martin SA, Hines WM, Juhasz P, Patterson DH, Peltier JM, Loo JA, Andrews PC. *Electrophoresis*. 1997; 18:382. [PubMed: 9150916]
147. Miliotis T, Kjellstrom S, Nilsson J, Laurell T, Edholm LE, Marko-Varga G. *J Mass Spectrom*. 2000; 35:369. [PubMed: 10767766]
148. Preisler J, Hu P, Rejtar T, Karger BL. *Anal Chem*. 2000; 72:4785. [PubMed: 11055691]
149. Preisler J, Hu P, Rejtar T, Moskovets E, Karger BL. *Anal Chem*. 2002; 74:17. [PubMed: 11795786]
150. Rejtar T, Hu P, Juhasz P, Campbell JM, Vestal ML, Preisler J, Karger BL. *J Proteome Res*. 2002; 1:171. [PubMed: 12643537]
151. Wall DB, Berger SJ, Finch JW, Cohen SA, Richardson K, Chapman R, Drabble D, Brown J, Gostick D. *Electrophoresis*. 2002; 23:3193. [PubMed: 12298091]
152. Chen HS, Rejtar T, Andreev V, Moskovets E, Karger BL. *Anal Chem*. 2005; 77:7816. [PubMed: 16316193]
153. Zhang N, Li N, Li L. *J Proteome Res*. 2004; 3:719. [PubMed: 15359724]
154. Takats Z, Wiseman JM, Ifa DR, Cooks RG. *CSH Protoc*. 2008; 2008:pdb top37. [PubMed: 21356818]
155. Takats Z, Wiseman JM, Ifa DR, Cooks RG. *CSH Protoc*. 2008; 2008:pdb prot4993. [PubMed: 21356811]

156. Peng IX, Shiea J, Ogorzalek Loo RR, Loo JA. *Rapid Commun Mass Spectrom.* 2007; 21:2541. [PubMed: 17639579]
157. Nayak R, Liu J, Sen AK, Knapp DR. *Anal Chem.* 2008; 80:8840. [PubMed: 18937429]
158. Peng IX, Ogorzalek Loo RR, Shiea J, Loo JA. *Anal Chem.* 2008; 80:6995. [PubMed: 18683952]
159. Trimpin S, Inutan ED, Herath TN, McEwen CN. *Mol Cell Proteomics.* 2010; 9:362. [PubMed: 19955086]
160. Qiao L, Sartor R, Gasilova N, Lu Y, Tobolkina E, Liu B, Girault HH. *Anal Chem.* 2012; 84:7422. [PubMed: 22876737]
161. Inutan ED, Trimpin S. *Mol Cell Proteomics.* 2012
162. Pagnotti VS, Chubatyi ND, McEwen CN. *Anal Chem.* 2011; 83:3981. [PubMed: 21528896]
163. Pagnotti VS, Inutan ED, Marshall DD, McEwen CN, Trimpin S. *Anal Chem.* 2011; 83:7591. [PubMed: 21899326]
164. Wang B, Inutan ED, Trimpin S. *J Am Soc Mass Spectrom.* 2012; 23:442. [PubMed: 22238127]
165. Peng Y, Zhang S, Gong X, Ma X, Yang C, Zhang X. *Anal Chem.* 2011; 83:8863. [PubMed: 22032471]
166. Jung S, Effelsberg U, Tallarek U. *Anal Chem.* 2011; 83:9167. [PubMed: 22022890]
167. Cech NB, Krone JR, Enke CG. *Anal Chem.* 2001; 73:208. [PubMed: 11199967]
168. Kocher T, Swart R, Mechtler K. *Anal Chem.* 2011; 83:2699. [PubMed: 21388192]
169. Krisp C, McKay MJ, Wolters DA, Molloy MP. *Anal Chem.* 2012; 84:1592. [PubMed: 22224914]
170. Kulevich SE, Frey BL, Kreitinger G, Smith LM. *Anal Chem.* 2010; 82:10135. [PubMed: 21114270]
171. Cagney G, Amiri S, Premawaradena T, Lindo M, Emili A. *Proteome Sci.* 2003; 1:5. [PubMed: 12946274]
172. Consortium CeS. *Science.* 1998; 282:2012. [PubMed: 9851916]
173. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, George RA, Lewis SE, Richards S, Ashburner M, Henderson SN, Sutton GG, Wortman JR, Yandell MD, Zhang Q, Chen LX, Brandon RC, Rogers YH, Blazej RG, Champe M, Pfeiffer BD, Wan KH, Doyle C, Baxter EG, Helt G, Nelson CR, Gabor GL, Abril JF, Agbayani A, An HJ, Andrews-Pfannkoch C, Baldwin D, Ballew RM, Basu A, Baxendale J, Bayraktaroglu L, Beasley EM, Beeson KY, Benos PV, Berman BP, Bhandari D, Bolshakov S, Borkova D, Botchan MR, Bouck J, Brokstein P, Brottier P, Burtis KC, Busam DA, Butler H, Cadieu E, Center A, Chandra I, Cherry JM, Cawley S, Dahlke C, Davenport LB, Davies P, de Pablos B, Delcher A, Deng Z, Mays AD, Dew I, Dietz SM, Dodson K, Doup LE, Downes M, Dugan-Rocha S, Dunkov BC, Dunn P, Durbin KJ, Evangelista CC, Ferraz C, Ferriera S, Fleischmann W, Fosler C, Gabrielian AE, Garg NS, Gelbart WM, Glasser K, Glodek A, Gong F, Gorrell JH, Gu Z, Guan P, Harris M, Harris NL, Harvey D, Heiman TJ, Hernandez JR, Houck J, Hostin D, Houston KA, Howland TJ, Wei MH, Ibegwam C. *Science.* 2000; 287:2185. [PubMed: 10731132]
174. Usaite R, Wohlschlegel J, Venable JD, Park SK, Nielsen J, Olsson L, Yates JR III. *J Proteome Res.* 2008; 7:266. [PubMed: 18173223]
175. Washburn MP, Wolters D, Yates JR 3rd. *Nat Biotechnol.* 2001; 19:242. [PubMed: 11231557]
176. Schrimpf SP, Weiss M, Reiter L, Ahrens CH, Jovanovic M, Malmstrom J, Brunner E, Mohanty S, Lercher MJ, Hunziker PE, Aebersold R, von Mering C, Hengartner MO. *PLoS Biol.* 2009; 7:e48. [PubMed: 19260763]
177. Brunner E, Ahrens CH, Mohanty S, Baetschmann H, Loevenich S, Potthast F, Deutsch EW, Panse C, de Lichtenberg U, Rinner O, Lee H, Pedrioli PG, Malmstrom J, Koehler K, Schrimpf S, Krijgsveld J, Kregenow F, Heck AJ, Hafen E, Schlapbach R, Aebersold R. *Nat Biotechnol.* 2007; 25:576. [PubMed: 17450130]
178. de Godoy LM, Olsen JV, Cox J, Nielsen ML, Hubner NC, Frohlich F, Walther TC, Mann M. *Nature.* 2008; 455:1251. [PubMed: 18820680]
179. Nagaraj N, Kulak NA, Cox J, Neuhauser N, Mayr K, Hoerning O, Vorm O, Mann M. *Mol Cell Proteomics.* 2012; 11:M111 013722. [PubMed: 22021278]

180. Gatlin CL, Eng JK, Cross ST, Detter JC, Yates JR 3rd . *Anal Chem.* 2000; 72:757. [PubMed: 10701260]
181. Gatlin CL, Kleemann GR, Hays LG, Link AJ, Yates JR 3rd . *Anal Biochem.* 1998; 263:93. [PubMed: 9750149]
182. Masuda T, Sugiyama N, Tomita M, Ishihama Y. *Anal Chem.* 2011; 83:7698. [PubMed: 21888424]
183. Zhou F, Lu Y, Ficarro SB, Webber JT, Marto JA. *Anal Chem.* 2012; 84:5133. [PubMed: 22519751]
184. Speers AE, Blackler AR, Wu CC. *Anal Chem.* 2007; 79:4613. [PubMed: 17500534]
185. Venable JD, Okach L, Agarwalla S, Brock A. *Anal Chem.* 2012; 84:9601. [PubMed: 23025328]
186. MacNair JE, Lewis KC, Jorgenson JW. *Anal Chem.* 1997; 69:983. [PubMed: 9075400]
187. Motoyama A, Venable JD, Ruse CI, Yates JR 3rd . *Anal Chem.* 2006; 78:5109. [PubMed: 16841936]
188. Chen CJ, Chen WY, Tseng MC, Chen YR. *Anal Chem.* 2012; 84:297. [PubMed: 22097963]
189. Winter D, Pipkorn R, Lehmann WD. *J Sep Sci.* 2009; 32:1111. [PubMed: 19360781]
190. Cristobal A, Hennrich ML, Giansanti P, Goerdalay SS, Heck AJ, Mohammed S. *Analyst.*
191. Marchetti N, Cavazzini A, Gritti F, Guiochon G. *J Chromatogr A.* 2007; 1163:203. [PubMed: 17632112]
192. Marchetti N, Guiochon G. *J Chromatogr A.* 2007; 1176:206. [PubMed: 18036600]
193. Ruta J, Guillaume D, Rudaz S, Veuthey JL. *J Sep Sci.* 2010; 33:2465. [PubMed: 20658489]
194. Schuster SA, Boyes BE, Wagner BM, Kirkland JJ. *J Chromatogr A.* 2012; 1228:232. [PubMed: 21855080]
195. Schuster SA, Wagner BM, Boyes BE, Kirkland JJ. *J Chromatogr Sci.* 2010; 48:566. [PubMed: 20819282]
196. Iwasaki M, Miwa S, Ikegami T, Tomita M, Tanaka N, Ishihama Y. *Anal Chem.* 2010; 82:2616. [PubMed: 20222674]
197. Iwasaki M, Sugiyama N, Tanaka N, Ishihama Y. *J Chromatogr A.* 2012; 1228:292. [PubMed: 22078304]
198. Yamana R, Iwasaki M, Wakabayashi M, Nakagawa M, Yamanaka S, Ishihama Y. *J Proteome Res.* 2013; 12:214. [PubMed: 23210603]
199. Boersema PJ, Mohammed S, Heck AJ. *Anal Bioanal Chem.* 2008; 391:151. [PubMed: 18264818]
200. Yang Y, Boysen RI, Hearn MT. *J Chromatogr A.* 2009; 1216:5518. [PubMed: 19535084]
201. Chen ML, Li LM, Yuan BF, Ma Q, Feng YQ. *J Chromatogr A.* 2012; 1230:54. [PubMed: 22340893]
202. de Jong EP, Griffin TJ. *J Proteome Res.* 2012; 11:5059. [PubMed: 22950739]
203. Busnel JM, Schoenmaker B, Ramautar R, Carrasco-Pancorbo A, Ratnayake C, Feitelson JS, Chapman JD, Deelder AM, Mayboroda OA. *Anal Chem.* 2010; 82:9476. [PubMed: 21028888]
204. Zhong X, Maxwell EJ, Chen DD. *Anal Chem.* 2011; 83:4916. [PubMed: 21528898]
205. Faserl K, Sarg B, Kremser L, Lindner H. *Anal Chem.* 2011; 83:7297. [PubMed: 21848273]
206. Li Y, Champion MM, Sun L, Champion PA, Wojcik R, Dovichi NJ. *Anal Chem.* 2012; 84:1617. [PubMed: 22182061]
207. Heemskerk AA, Busnel JM, Schoenmaker B, Derks RJ, Klychnikov O, Hensbergen PJ, Deelder AM, Mayboroda OA. *Anal Chem.* 2012; 84:4552. [PubMed: 22494114]
208. Wang Y, Fonslow BR, Wong CC, Nakorchevsky A, Yates JR 3rd . *Anal Chem.* 2012; 84:8505. [PubMed: 23004022]
209. Li Y, Wojcik R, Dovichi NJ, Champion MM. *Anal Chem.* 2012; 84:6116. [PubMed: 22690842]
210. Shen Y, Jacobs JM, Camp DG 2nd, Fang R, Moore RJ, Smith RD, Xiao W, Davis RW, Tompkins RG. *Anal Chem.* 2004; 76:1134. [PubMed: 14961748]
211. Motoyama A, Xu T, Ruse CI, Wohlschlegel JA, Yates JR 3rd . *Anal Chem.* 2007; 79:3623. [PubMed: 17411013]
212. Dai J, Wang LS, Wu YB, Sheng QH, Wu JR, Shieh CH, Zeng R. *J Proteome Res.* 2009; 8:133. [PubMed: 19053533]

213. Gilar M, Olivova P, Daly AE, Gebler JC. *J Sep Sci.* 2005; 28:1694. [PubMed: 16224963]
214. Zhou F, Cardoza JD, Ficarro SB, Adelmant GO, Lazaro JB, Marto JA. *J Proteome Res.* 2010; 9:6242. [PubMed: 20873769]
215. Wang N, Xie C, Young JB, Li L. *Anal Chem.* 2009; 81:1049. [PubMed: 19178338]
216. Malmstrom J, Lee H, Nesvizhskii AI, Shteynberg D, Mohanty S, Brunner E, Ye M, Weber G, Eckerskorn C, Aebersold R. *J Proteome Res.* 2006; 5:2241. [PubMed: 16944936]
217. Essader AS, Cargile BJ, Bundy JL, Stephenson JL Jr. *Proteomics.* 2005; 5:24. [PubMed: 15672457]
218. Hubner NC, Ren S, Mann M. *Proteomics.* 2008; 8:4862. [PubMed: 19003865]
219. Taouatas N, Altelaar AF, Drugan MM, Helbig AO, Mohammed S, Heck AJ. *Mol Cell Proteomics.* 2009; 8:190. [PubMed: 18824475]
220. Boersema PJ, Divecha N, Heck AJ, Mohammed S. *J Proteome Res.* 2007; 6:937. [PubMed: 17256977]
221. Di Palma S, Boersema PJ, Heck AJ, Mohammed S. *Anal Chem.* 2011; 83:3440. [PubMed: 21443167]
222. Hao P, Qian J, Ren Y, Sze SK. *J Proteome Res.* 2011; 10:5568. [PubMed: 22014306]
223. Hao P, Guo T, Li X, Adav SS, Yang J, Wei M, Sze SK. *J Proteome Res.* 2010; 9:3520. [PubMed: 20450224]
224. Hao P, Guo T, Sze SK. *PLoS One.* 2011; 6:e16884. [PubMed: 21373199]
225. Hao P, Qian J, Dutta B, Cheow ES, Sim KH, Meng W, Adav SS, Alpert A, Sze SK. *J Proteome Res.* 2012; 11:1804. [PubMed: 22239700]
226. Ritorto MS, Cook K, Tyagi K, Pedrioli PG, Trost M. *J Proteome Res.* 2013
227. Dai J, Jin WH, Sheng QH, Shieh CH, Wu JR, Zeng R. *J Proteome Res.* 2007; 6:250. [PubMed: 17203969]
228. Hennrich ML, Groenewold V, Kops GJ, Heck AJ, Mohammed S. *Anal Chem.* 2011; 83:7137. [PubMed: 21815630]
229. Ficarro SB, Zhang Y, Carrasco-Alfonso MJ, Garg B, Adelmant G, Webber JT, Luckey CJ, Marto JA. *Mol Cell Proteomics.* 2011; 10:O111 011064. [PubMed: 21788404]
230. Zarei M, Sprenger A, Gretzmeier C, Dengjel J. *J Proteome Res.* 2012; 11:4269. [PubMed: 22768876]
231. Zhou F, Sikorski TW, Ficarro SB, Webber JT, Marto JA. *Anal Chem.* 2011; 83:6996. [PubMed: 21851055]
232. Wong CC, Cociorva D, Venable JD, Xu T, Yates JR 3rd. *J Am Soc Mass Spectrom.* 2009; 20:1405. [PubMed: 19467883]
233. Venable JD, Dong MQ, Wohlschlegel J, Dillin A, Yates JR. *Nat Methods.* 2004; 1:39. [PubMed: 15782151]
234. Panchaud A, Scherl A, Shaffer SA, von Haller PD, Kulasekara HD, Miller SI, Goodlett DR. *Anal Chem.* 2009; 81:6481. [PubMed: 19572557]
235. Ledvina AR, Savitski MM, Zubarev AR, Good DM, Coon JJ, Zubarev RA. *Anal Chem.* 2011; 83:7651. [PubMed: 21913643]
236. Gillet LC, Navarro P, Tate S, Rost H, Selevsek N, Reiter L, Bonner R, Aebersold R. *Mol Cell Proteomics.* 2012; 11:O111 016717. [PubMed: 22261725]
237. Mayya V, Rezaul K, Cong YS, Han D. *Mol Cell Proteomics.* 2005; 4:214. [PubMed: 15608339]
238. Blackler AR, Klammer AA, MacCoss MJ, Wu CC. *Anal Chem.* 2006; 78:1337. [PubMed: 16478131]
239. Second TP, Blethrow JD, Schwartz JC, Merrihew GE, MacCoss MJ, Swaney DL, Russell JD, Coon JJ, Zabrouskov V. *Anal Chem.* 2009; 81:7757. [PubMed: 19689114]
240. Syka JE, Marto JA, Bai DL, Horning S, Senko MW, Schwartz JC, Ueberheide B, Garcia B, Busby S, Muratore T, Shabanowitz J, Hunt DF. *J Proteome Res.* 2004; 3:621. [PubMed: 15253445]
241. Yates JR, Cociorva D, Liao L, Zabrouskov V. *Anal Chem.* 2006; 78:493. [PubMed: 16408932]

242. Michalski A, Damoc E, Lange O, Denisov E, Nolting D, Muller M, Viner R, Schwartz J, Remes P, Belford M, Dunyach JJ, Cox J, Horning S, Mann M, Makarov A. *Mol Cell Proteomics*. 2012; 11:O111 013698. [PubMed: 22159718]
243. Tolmachev AV, Monroe ME, Jaitly N, Petyuk VA, Adkins JN, Smith RD. *Anal Chem*. 2006; 78:8374. [PubMed: 17165830]
244. Becker CH, Kumar P, Jones T, Lin H. *Anal Chem*. 2007; 79:1702. [PubMed: 17297976]
245. Petyuk VA, Jaitly N, Moore RJ, Ding J, Metz TO, Tang K, Monroe ME, Tolmachev AV, Adkins JN, Belov ME, Dabney AR, Qian WJ, Camp DG 2nd, Smith RD. *Anal Chem*. 2008; 80:693. [PubMed: 18163597]
246. Zhang J, Ma J, Dou L, Wu S, Qian X, Xie H, Zhu Y, He F. *J Proteome Res*. 2009; 8:849. [PubMed: 19102693]
247. Cox J, Michalski A, Mann M. *J Am Soc Mass Spectrom*. 2011; 22:1373. [PubMed: 21953191]
248. Lee KA, Farnsworth C, Yu W, Bonilla LE. *J Proteome Res*. 2011; 10:880. [PubMed: 21133379]
249. Haas W, Faherty BK, Gerber SA, Elias JE, Beausoleil SA, Bakalarski CE, Li X, Villen J, Gygi SP. *Mol Cell Proteomics*. 2006; 5:1326. [PubMed: 16635985]
250. Scheltema RA, Kamleh A, Wildridge D, Ebikeme C, Watson DG, Barrett MP, Jansen RC, Breitling R. *Proteomics*. 2008; 8:4647. [PubMed: 18937253]
251. Zhang Y, Wen Z, Washburn MP, Florens L. *Anal Chem*. 2011; 83:9344. [PubMed: 22044264]
252. Wu S, Kaiser NK, Meng D, Anderson GA, Zhang K, Bruce JE. *J Proteome Res*. 2005; 4:1434. [PubMed: 16083297]
253. Olsen JV, de Godoy LM, Li G, Macek B, Mortensen P, Pesch R, Makarov A, Lange O, Horning S, Mann M. *Mol Cell Proteomics*. 2005; 4:2010. [PubMed: 16249172]
254. Cox J, Mann M. *Nat Biotechnol*. 2008; 26:1367. [PubMed: 19029910]
255. Steen H, Kuster B, Mann M. *J Mass Spectrom*. 2001; 36:782. [PubMed: 11473401]
256. Elias JE, Haas W, Faherty BK, Gygi SP. *Nat Methods*. 2005; 2:667. [PubMed: 16118637]
257. Andrews GL, Simons BL, Young JB, Hawkridge AM, Muddiman DC. *Anal Chem*. 2011; 83:5442. [PubMed: 21619048]
258. Michalski A, Damoc E, Hauschild JP, Lange O, Wieghaus A, Makarov A, Nagaraj N, Cox J, Mann M, Horning S. *Mol Cell Proteomics*. 2011; 10:M111 011015. [PubMed: 21642640]
259. Kelstrup CD, Young C, Lavalley R, Nielsen ML, Olsen JV. *J Proteome Res*. 2012
260. Venne K, Bonneil E, Eng K, Thibault P. *Anal Chem*. 2005; 77:2176. [PubMed: 15801752]
261. Saba J, Bonneil E, Pomies C, Eng K, Thibault P. *J Proteome Res*. 2009; 8:3355. [PubMed: 19469569]
262. Canterbury JD, Yi X, Hoopmann MR, MacCoss MJ. *Anal Chem*. 2008; 80:6888. [PubMed: 18693747]
263. Baker ES, Livesay EA, Orton DJ, Moore RJ, Danielson WF 3rd, Prior DC, Ibrahim YM, LaMarche BL, Mayampurath AM, Schepmoes AA, Hopkins DF, Tang K, Smith RD, Belov ME. *J Proteome Res*. 2010; 9:997. [PubMed: 20000344]
264. Swearingen KE, Hoopmann MR, Johnson RS, Saleem RA, Aitchison JD, Moritz RL. *Mol Cell Proteomics*. 2012; 11:M111 014985. [PubMed: 22186714]
265. Shvartsburg AA, Creese AJ, Smith RD, Cooper HJ. *Anal Chem*. 2011; 83:6918. [PubMed: 21863819]
266. Shvartsburg AA, Smith RD. *Anal Chem*. 2011; 83:23. [PubMed: 21117630]
267. Shvartsburg AA, Creese AJ, Smith RD, Cooper HJ. *Anal Chem*. 2010; 82:8327. [PubMed: 20843012]
268. Shvartsburg AA, Singer D, Smith RD, Hoffmann R. *Anal Chem*. 2011; 83:5078. [PubMed: 21667994]
269. Creese AJ, Cooper HJ. *Anal Chem*. 2012; 84:2597. [PubMed: 22280549]
270. Ting L, Cowley MJ, Hoon SL, Guilhaus M, Raftery MJ, Cavicchioli R. *Mol Cell Proteomics*. 2009; 8:2227. [PubMed: 19605365]
271. Bereman MS, Canterbury JD, Egertson JD, Horner J, Remes PM, Schwartz J, Zabrouskov V, MacCoss MJ. *Anal Chem*. 2012; 84:1533. [PubMed: 22192247]

272. McAlister GC, Phanstiel DH, Brumbaugh J, Westphall MS, Coon JJ. *Mol Cell Proteomics*. 2011; 10:O111 009456. [PubMed: 21393638]
273. Olsen JV, Macek B, Lange O, Makarov A, Horning S, Mann M. *Nat Methods*. 2007; 4:709. [PubMed: 17721543]
274. McAlister GC, Phanstiel D, Wenger CD, Lee MV, Coon JJ. *Anal Chem*. 2010; 82:316. [PubMed: 19938823]
275. Pichler P, Kocher T, Holzmann J, Mohring T, Ammerer G, Mechtler K. *Anal Chem*. 2011; 83:1469. [PubMed: 21275378]
276. Griffin TJ, Xie H, Bandhakavi S, Popko J, Mohan A, Carlis JV, Higgins L. *J Proteome Res*. 2007; 6:4200. [PubMed: 17902639]
277. Guo T, Gan CS, Zhang H, Zhu Y, Kon OL, Sze SK. *J Proteome Res*. 2008; 7:4831. [PubMed: 18837533]
278. DeGnore JP, Qin J. *J Am Soc Mass Spectrom*. 1998; 9:1175. [PubMed: 9794085]
279. Olsen JV, Mann M. *Proc Natl Acad Sci U S A*. 2004; 101:13417. [PubMed: 15347803]
280. Schroeder MJ, Shabanowitz J, Schwartz JC, Hunt DF, Coon JJ. *Anal Chem*. 2004; 76:3590. [PubMed: 15228329]
281. Palumbo AM, Tepe JJ, Reid GE. *J Proteome Res*. 2008; 7:771. [PubMed: 18181561]
282. Ulintz PJ, Yocum AK, Bodenmiller B, Aebersold R, Andrews PC, Nesvizhskii AI. *J Proteome Res*. 2009; 8:887. [PubMed: 19072539]
283. Nagaraj N, D'Souza RC, Cox J, Olsen JV, Mann M. *J Proteome Res*. 2010; 9:6786. [PubMed: 20873877]
284. Jedrychowski MP, Huttlin EL, Haas W, Sowa ME, Rad R, Gygi SP. *Mol Cell Proteomics*. 2011; 10:M111 009910. [PubMed: 21917720]
285. Zubarev RA, Kelleher NL, McLafferty FW. *Journal of the American Chemical Society*. 1998; 120:3265.
286. Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. *Proc Natl Acad Sci U S A*. 2004; 101:9528. [PubMed: 15210983]
287. Kelleher NL, Zubarev RA, Bush K, Furie B, Furie BC, McLafferty FW, Walsh CT. *Anal Chem*. 1999; 71:4250. [PubMed: 10517147]
288. Cooper HJ, Akbarzadeh S, Heath JK, Zeller M. *J Proteome Res*. 2005; 4:1538. [PubMed: 16212404]
289. Creese AJ, Cooper HJ. *J Am Soc Mass Spectrom*. 2007; 18:891. [PubMed: 17350280]
290. Kalli A, Hakansson K. *J Proteome Res*. 2008; 7:2834. [PubMed: 18549259]
291. Sweet SM, Bailey CM, Cunningham DL, Heath JK, Cooper HJ. *Mol Cell Proteomics*. 2009; 8:904. [PubMed: 19131326]
292. Satake H, Hasegawa H, Hirabayashi A, Hashimoto Y, Baba T, Masuda K. *Anal Chem*. 2007; 79:8755. [PubMed: 17902701]
293. Compton PD, Strukl JV, Bai DL, Shabanowitz J, Hunt DF. *Anal Chem*. 2012; 84:1781. [PubMed: 22182179]
294. Udeshi ND, Compton PD, Shabanowitz J, Hunt DF, Rose KL. *Nat Protoc*. 2008; 3:1709. [PubMed: 18927556]
295. Aguiar M, Haas W, Beausoleil SA, Rush J, Gygi SP. *J Proteome Res*. 2010; 9:3103. [PubMed: 20377248]
296. Palumbo AM, Reid GE. *Anal Chem*. 2008; 80:9735. [PubMed: 19012417]
297. Taouatas N, Drugan MM, Heck AJ, Mohammed S. *Nat Methods*. 2008; 5:405. [PubMed: 18425140]
298. McAlister GC, Berggren WT, Griep-Raming J, Horning S, Makarov A, Phanstiel D, Stafford G, Swaney DL, Syka JE, Zabrouskov V, Coon JJ. *J Proteome Res*. 2008; 7:3127. [PubMed: 18613715]
299. McAlister GC, Phanstiel D, Good DM, Berggren WT, Coon JJ. *Anal Chem*. 2007; 79:3525. [PubMed: 17441688]

300. Swaney DL, McAlister GC, Wirtala M, Schwartz JC, Syka JE, Coon JJ. *Anal Chem.* 2007; 79:477. [PubMed: 17222010]
301. Wu SL, Huhmer AF, Hao Z, Karger BL. *J Proteome Res.* 2007; 6:4230. [PubMed: 17900180]
302. Swaney DL, McAlister GC, Coon JJ. *Nat Methods.* 2008; 5:959. [PubMed: 18931669]
303. Frese CK, Altelaar AF, Hennrich ML, Nolting D, Zeller M, Griep-Raming J, Heck AJ, Mohammed S. *J Proteome Res.* 2011; 10:2377. [PubMed: 21413819]
304. Mischerikow N, van Nierop P, Li KW, Bernstein HG, Smit AB, Heck AJ, Altelaar AF. *Analyst.* 2010; 135:2643. [PubMed: 20714520]
305. Bridon G, Bonneil E, Muratore-Schroeder T, Caron-Lizotte O, Thibault P. *J Proteome Res.* 2012; 11:927. [PubMed: 22059388]
306. McAlister GC, Russell JD, Rumachik NG, Hebert AS, Syka JE, Geer LY, Westphall MS, Pagliarini DJ, Coon JJ. *Anal Chem.* 2012; 84:2875. [PubMed: 22335612]
307. Ledvina AR, Beauchene NA, McAlister GC, Syka JE, Schwartz JC, Griep-Raming J, Westphall MS, Coon JJ. *Anal Chem.* 2010; 82:10068. [PubMed: 21062032]
308. Vasicek LA, Ledvina AR, Shaw J, Griep-Raming J, Westphall MS, Coon JJ, Brodbelt JS. *J Am Soc Mass Spectrom.* 2011; 22:1105. [PubMed: 21953052]
309. Ledvina AR, Lee MV, McAlister GC, Westphall MS, Coon JJ. *Anal Chem.* 2012; 84:4513. [PubMed: 22480380]
310. Blume-Jensen P, Hunter T. *Nature.* 2001; 411:355. [PubMed: 11357143]
311. Hunter T. *Cell.* 2000; 100:113. [PubMed: 10647936]
312. MacDonald JA, Mackey AJ, Pearson WR, Haystead TA. *Mol Cell Proteomics.* 2002; 1:314. [PubMed: 12096113]
313. Roepstorff P, Kristiansen K. *Biomed Mass Spectrom.* 1974; 1:231. [PubMed: 4441617]
314. Olsen JV, Blagoev B, Gnäd F, Macek B, Kumar C, Mortensen P, Mann M. *Cell.* 2006; 127:635. [PubMed: 17081983]
315. Beausoleil SA, Jedrychowski M, Schwartz D, Elias JE, Villen J, Li J, Cohn MA, Cantley LC, Gygi SP. *Proc Natl Acad Sci U S A.* 2004; 101:12130. [PubMed: 15302935]
316. Pinkse MW, Uitto PM, Hilhorst MJ, Ooms B, Heck AJ. *Anal Chem.* 2004; 76:3935. [PubMed: 15253627]
317. Andersson L, Porath J. *Anal Biochem.* 1986; 154:250. [PubMed: 3085541]
318. Posewitz MC, Tempst P. *Anal Chem.* 1999; 71:2883. [PubMed: 10424175]
319. Michel H, Hunt DF, Shabanowitz J, Bennett J. *J Biol Chem.* 1988; 263:1123. [PubMed: 3121625]
320. Trojer L, Stecher G, Feuerstein I, Lubbad S, Bonn GK. *J Chromatogr A.* 2005; 1079:197. [PubMed: 16038305]
321. Haydon CE, Evers PA, Aveline-Wolf LD, Resing KA, Maller JL, Ahn NG. *Mol Cell Proteomics.* 2003; 2:1055. [PubMed: 12885952]
322. Ficarro SB, McClelland ML, Stukenberg PT, Burke DJ, Ross MM, Shabanowitz J, Hunt DF, White FM. *Nat Biotechnol.* 2002; 20:301. [PubMed: 11875433]
323. Han G, Ye M, Zou H. *Analyst.* 2008; 133:1128. [PubMed: 18709185]
324. Larsen MR, Thingholm TE, Jensen ON, Roepstorff P, Jorgensen TJ. *Mol Cell Proteomics.* 2005; 4:873. [PubMed: 15858219]
325. Thingholm TE, Jorgensen TJ, Jensen ON, Larsen MR. *Nat Protoc.* 2006; 1:1929. [PubMed: 17487178]
326. Kweon HK, Hakansson K. *Anal Chem.* 2006; 78:1743. [PubMed: 16536406]
327. Li Y, Lin H, Deng C, Yang P, Zhang X. *Proteomics.* 2008; 8:238. [PubMed: 18081192]
328. Bodenmiller B, Mueller LN, Mueller M, Domon B, Aebersold R. *Nat Methods.* 2007; 4:231. [PubMed: 17293869]
329. Li QR, Ning ZB, Tang JS, Nie S, Zeng R. *J Proteome Res.* 2009; 8:5375. [PubMed: 19761217]
330. Mazanek M, Mituloviae G, Herzog F, Stingl C, Hutchins JR, Peters JM, Mechtler K. *Nat Protoc.* 2007; 2:1059. [PubMed: 17545998]
331. Simon ES, Young M, Chan A, Bao ZQ, Andrews PC. *Anal Biochem.* 2008; 377:234. [PubMed: 18396144]

332. Thingholm TE, Jensen ON, Robinson PJ, Larsen MR. *Mol Cell Proteomics*. 2008; 7:661. [PubMed: 18039691]
333. Iliuk AB, Martin VA, Alicie BM, Geahlen RL, Tao WA. *Mol Cell Proteomics*. 2010; 9:2162. [PubMed: 20562096]
334. Han G, Ye M, Zhou H, Jiang X, Feng S, Tian R, Wan D, Zou H, Gu J. *Proteomics*. 2008; 8:1346. [PubMed: 18318008]
335. Alpert AJ. *J Chromatogr*. 1990; 499:177. [PubMed: 2324207]
336. Alpert AJ. *Anal Chem*. 2008; 80:62. [PubMed: 18027909]
337. Beausoleil SA, Villen J, Gerber SA, Rush J, Gygi SP. *Nat Biotechnol*. 2006; 24:1285. [PubMed: 16964243]
338. Hennrich ML, van den Toorn HW, Groenewold V, Heck AJ, Mohammed S. *Anal Chem*. 2012; 84:1804. [PubMed: 22303813]
339. Xu Y, Sprung R, Kwon SW, Kim SC, Zhao Y. *J Proteome Res*. 2007; 6:1153. [PubMed: 17330948]
340. Hung CW, Kubler D, Lehmann WD. *Electrophoresis*. 2007; 28:2044. [PubMed: 17523138]
341. McNulty DE, Annan RS. *Mol Cell Proteomics*. 2008; 7:971. [PubMed: 18212344]
342. Dong M, Ye M, Cheng K, Song C, Pan Y, Wang C, Bian Y, Zou H. *J Proteome Res*. 2012; 11:4673. [PubMed: 22871156]
343. Reynolds EC, Riley PF, Adamson NJ. *Anal Biochem*. 1994; 217:277. [PubMed: 8203756]
344. Ruse CI, McClatchy DB, Lu B, Cociorva D, Motoyama A, Park SK, Yates JR 3rd. *J Proteome Res*. 2008; 7:2140. [PubMed: 18452278]
345. Zhang X, Ye J, Jensen ON, Roepstorff P. *Mol Cell Proteomics*. 2007; 6:2032. [PubMed: 17675664]
346. Ballif BA, Carey GR, Sunyaev SR, Gygi SP. *J Proteome Res*. 2008; 7:311. [PubMed: 18034455]
347. Rush J, Moritz A, Lee KA, Guo A, Goss VL, Spek EJ, Zhang H, Zha XM, Polakiewicz RD, Comb MJ. *Nat Biotechnol*. 2005; 23:94. [PubMed: 15592455]
348. Mamone G, Picariello G, Ferranti P, Addeo F. *Proteomics*. 2010; 10:380. [PubMed: 19953538]
349. Fonslow BR, Niessen SM, Singh M, Wong CC, Xu T, Carvalho PC, Choi J, Park SK, Yates JR 3rd. *J Proteome Res*. 2012; 11:2697. [PubMed: 22509746]
350. Bodenmiller B, Mueller LN, Pedrioli PG, Pflieger D, Junger MA, Eng JK, Aebersold R, Tao WA. *Mol Biosyst*. 2007; 3:275. [PubMed: 17372656]
351. Jin WH, Dai J, Zhou H, Xia QC, Zou HF, Zeng R. *Rapid Commun Mass Spectrom*. 2004; 18:2169. [PubMed: 15378723]
352. Torres MP, Thapar R, Marzluff WF, Borchers CH. *J Proteome Res*. 2005; 4:1628. [PubMed: 16212415]
353. Ishihama Y, Wei FY, Aoshima K, Sato T, Kuromitsu J, Oda Y. *J Proteome Res*. 2007; 6:1139. [PubMed: 17330947]
354. Stensballe A, Jensen ON, Olsen JV, Haselmann KF, Zubarev RA. *Rapid Commun Mass Spectrom*. 2000; 14:1793. [PubMed: 11006587]
355. Chi A, Huttenhower C, Geer LY, Coon JJ, Syka JE, Bai DL, Shabanowitz J, Burke DJ, Troyanskaya OG, Hunt DF. *Proc Natl Acad Sci U S A*. 2007; 104:2193. [PubMed: 17287358]
356. Molina H, Horn DM, Tang N, Mathivanan S, Pandey A. *Proc Natl Acad Sci U S A*. 2007; 104:2199. [PubMed: 17287340]
357. Lu B, Ruse C, Xu T, Park SK, Yates J 3rd. *Anal Chem*. 2007; 79:1301. [PubMed: 17297928]
358. Diella F, Gould CM, Chica C, Via A, Gibson TJ. *Nucleic Acids Res*. 2008; 36:D240. [PubMed: 17962309]
359. Gnad F, Ren S, Cox J, Olsen JV, Macek B, Orosi M, Mann M. *Genome Biol*. 2007; 8:R250. [PubMed: 18039369]
360. Obenaus JC, Cantley LC, Yaffe MB. *Nucleic Acids Res*. 2003; 31:3635. [PubMed: 12824383]
361. Wong YH, Lee TY, Liang HK, Huang CM, Wang TY, Yang YH, Chu CH, Huang HD, Ko MT, Hwang JK. *Nucleic Acids Res*. 2007; 35:W588. [PubMed: 17517770]
362. Xue Y, Li A, Wang L, Feng H, Yao X. *BMC Bioinformatics*. 2006; 7:163. [PubMed: 16549034]

363. Hershko A, Ciechanover A, Rose IA. *Proc Natl Acad Sci U S A*. 1979; 76:3107. [PubMed: 290989]
364. Ciechanover A, Hod Y, Hershko A. *Biochem Biophys Res Commun*. 1978; 81:1100. [PubMed: 666810]
365. Pickart CM. *Annu Rev Biochem*. 2001; 70:503. [PubMed: 11395416]
366. Dikic I, Wakatsuki S, Walters KJ. *Nat Rev Mol Cell Biol*. 2009; 10:659. [PubMed: 19773779]
367. Newton K, Matsumoto ML, Wertz IE, Kirkpatrick DS, Lill JR, Tan J, Dugger D, Gordon N, Sidhu SS, Fellouse FA, Komuves L, French DM, Ferrando RE, Lam C, Compaan D, Yu C, Bosanac I, Hymowitz SG, Kelley RF, Dixit VM. *Cell*. 2008; 134:668. [PubMed: 18724939]
368. Schwartz DC, Hochstrasser M. *Trends Biochem Sci*. 2003; 28:321. [PubMed: 12826404]
369. Welchman RL, Gordon C, Mayer RJ. *Nat Rev Mol Cell Biol*. 2005; 6:599. [PubMed: 16064136]
370. Grabbe C, Dikic I. *Chem Rev*. 2009; 109:1481. [PubMed: 19253967]
371. Boja ES, Fales HM. *Anal Chem*. 2001; 73:3576. [PubMed: 11510821]
372. Nielsen ML, Vermeulen M, Bonaldi T, Cox J, Moroder L, Mann M. *Nat Methods*. 2008; 5:459. [PubMed: 18511913]
373. Xu P, Duong DM, Seyfried NT, Cheng D, Xie Y, Robert J, Rush J, Hochstrasser M, Finley D, Peng J. *Cell*. 2009; 137:133. [PubMed: 19345192]
374. Vasilescu J, Smith JC, Ethier M, Figeys D. *J Proteome Res*. 2005; 4:2192. [PubMed: 16335966]
375. Peng J, Schwartz D, Elias JE, Thoreen CC, Cheng D, Marsischky G, Roelofs J, Finley D, Gygi SP. *Nat Biotechnol*. 2003; 21:921. [PubMed: 12872131]
376. Xu G, Paige JS, Jaffrey SR. *Nat Biotechnol*. 2010; 28:868. [PubMed: 20639865]
377. Hicke L, Schubert HL, Hill CP. *Nat Rev Mol Cell Biol*. 2005; 6:610. [PubMed: 16064137]
378. Jones J, Wu K, Yang Y, Guerrero C, Nillegoda N, Pan ZQ, Huang L. *J Proteome Res*. 2008; 7:1274. [PubMed: 18247557]
379. Giannakopoulos NV, Luo JK, Papov V, Zou W, Lenschow DJ, Jacobs BS, Borden EC, Li J, Virgin HW, Zhang DE. *Biochem Biophys Res Commun*. 2005; 336:496. [PubMed: 16139798]
380. Wohlschlegel JA, Johnson ES, Reed SI, Yates JR 3rd. *J Proteome Res*. 2006; 5:761. [PubMed: 16602682]
381. Kirkpatrick DS, Denison C, Gygi SP. *Nat Cell Biol*. 2005; 7:750. [PubMed: 16056266]
382. Huang H, Jeon MS, Liao L, Yang C, Elly C, Yates JR 3rd, Liu YC. *Immunity*. 2010; 33:60. [PubMed: 20637659]
383. Bause E. *Biochem J*. 1983; 209:331. [PubMed: 6847620]
384. Gerken TA, Gupta R, Jentoft N. *Biochemistry*. 1992; 31:639. [PubMed: 1310044]
385. Greis KD, Hayes BK, Comer FI, Kirk M, Barnes S, Lowary TL, Hart GW. *Anal Biochem*. 1996; 234:38. [PubMed: 8742080]
386. Gonzalez-Begne M, Lu B, Liao L, Xu T, Bedi G, Melvin JE, Yates JR 3rd. *J Proteome Res*. 2011; 10:5031. [PubMed: 21936497]
387. Mechref Y, Madera M, Novotny MV. *Methods Mol Biol*. 2008; 424:373. [PubMed: 18369876]
388. Zhang H, Li XJ, Martin DB, Aebersold R. *Nat Biotechnol*. 2003; 21:660. [PubMed: 12754519]
389. Hagglund P, Bunkenborg J, Elortza F, Jensen ON, Roepstorff P. *J Proteome Res*. 2004; 3:556. [PubMed: 15253437]
390. Yeh CH, Chen SH, Li DT, Lin HP, Huang HJ, Chang CI, Shih WL, Chern CL, Shi FK, Hsu JL. *J Chromatogr A*. 2012; 1224:70. [PubMed: 22226559]
391. Monzo A, Bonn GK, Guttman A. *Anal Bioanal Chem*. 2007; 389:2097. [PubMed: 17909756]
392. Krisp C, Kubutat C, Kyas A, Steinstrasser L, Jacobsen F, Wolters D. *J Proteomics*. 2011; 74:502. [PubMed: 21278005]
393. Hakansson K, Cooper HJ, Emmett MR, Costello CE, Marshall AG, Nilsson CL. *Anal Chem*. 2001; 73:4530. [PubMed: 11575803]
394. Hakansson K, Chalmers MJ, Quinn JP, McFarland MA, Hendrickson CL, Marshall AG. *Anal Chem*. 2003; 75:3256. [PubMed: 12964777]
395. Catalina MI, Koeleman CA, Deelder AM, Wührer M. *Rapid Commun Mass Spectrom*. 2007; 21:1053. [PubMed: 17311219]

396. Kuster B, Mann M. *Anal Chem.* 1999; 71:1431. [PubMed: 10204042]
397. Atwood JA 3rd, Sahoo SS, Alvarez-Manilla G, Weatherly DB, Kolli K, Orlando R, York WS. *Rapid Commun Mass Spectrom.* 2005; 19:3002. [PubMed: 16196021]
398. Hwang CS, Shemorry A, Varshavsky A. *Science.* 2010; 327:973. [PubMed: 20110468]
399. Polevoda B, Sherman F. *J Biol Chem.* 2000; 275:36479. [PubMed: 11013267]
400. Allfrey VG, Faulkner R, Mirsky AE. *Proc Natl Acad Sci U S A.* 1964; 51:786. [PubMed: 14172992]
401. Fu M, Wang C, Zhang X, Pestell RG. *Biochem Pharmacol.* 2004; 68:1199. [PubMed: 15313417]
402. Zhao S, Xu W, Jiang W, Yu W, Lin Y, Zhang T, Yao J, Zhou L, Zeng Y, Li H, Li Y, Shi J, An W, Hancock SM, He F, Qin L, Chin J, Yang P, Chen X, Lei Q, Xiong Y, Guan KL. *Science.* 2010; 327:1000. [PubMed: 20167786]
403. Butler R, Bates GP. *Nat Rev Neurosci.* 2006; 7:784. [PubMed: 16988654]
404. Yang XJ, Seto E. *Mol Cell.* 2008; 31:449. [PubMed: 18722172]
405. Choudhary C, Kumar C, Gnad F, Nielsen ML, Rehman M, Walther TC, Olsen JV, Mann M. *Science.* 2009; 325:834. [PubMed: 19608861]
406. Eriksen JG, Beavis AW, Coffey MA, Leer JW, Magrini SM, Benstead K, Boelling T, Hjalmeriksson M, Kantor G, Maciejewski B, Mezeckis M, Oliveira A, Thirion P, Vitek P, Olsen DR, Eudaldo T, Enghardt W, Francois P, Garibaldi C, Heijmen B, Josipovic M, Major T, Nikolettopoulos S, Rijnders A, Waligorski M, Wasilewska-Radwanska M, Mullaney L, Boejen A, Vaandering A, Vandevelde G, Verfaillie C, Potter R. *Radiother Oncol.* 2012; 103:103. [PubMed: 22444243]
407. Stallcup MR. *Oncogene.* 2001; 20:3014. [PubMed: 11420716]
408. Ong SE, Mittler G, Mann M. *Nat Methods.* 2004; 1:119. [PubMed: 15782174]
409. Pang CN, Gasteiger E, Wilkins MR. *BMC Genomics.* 2010; 11:92. [PubMed: 20137074]
410. Vermeulen M, Eberl HC, Matarese F, Marks H, Denissov S, Butter F, Lee KK, Olsen JV, Hyman AA, Stunnenberg HG, Mann M. *Cell.* 2010; 142:967. [PubMed: 20850016]
411. Lee JC, Kang SU, Jeon Y, Park JW, You JS, Ha SW, Bae N, Lubec G, Kwon SH, Lee JS, Cho EJ, Han JW. *Nat Commun.* 2012; 3:927. [PubMed: 22735455]
412. Bachi A, Dalle-Donne I, Scaloni A. *Chem Rev.* 2012
413. Lindahl M, Mata-Cabana A, Kieselbach T. *Antioxid Redox Signal.* 2011; 14:2581. [PubMed: 21275844]
414. Nakamura T, Cho DH, Lipton SA. *Exp Neurol.* 2012; 238:12. [PubMed: 22771760]
415. Wiseman DA, Thurmond DC. *Curr Diabetes Rev.* 2012; 8:303. [PubMed: 22587517]
416. Wang Z. *Cancer Lett.* 2012; 320:123. [PubMed: 22425962]
417. Murphy E, Kohr M, Sun J, Nguyen T, Steenbergen C. *J Mol Cell Cardiol.* 2012; 52:568. [PubMed: 21907718]
418. Jaffrey SR, Snyder SH. *Sci STKE.* 2001; 2001:pl1. [PubMed: 11752655]
419. Forrester MT, Thompson JW, Foster MW, Nogueira L, Moseley MA, Stamler JS. *Nat Biotechnol.* 2009; 27:557. [PubMed: 19483679]
420. Baerenfaller K, Grossmann J, Grobei MA, Hull R, Hirsch-Hoffmann M, Yalovsky S, Zimmermann P, Grossniklaus U, Gruissem W, Baginsky S. *Science.* 2008; 320:938. [PubMed: 18436743]
421. Beck M, Schmidt A, Malmstroem J, Claassen M, Ori A, Szymborska A, Herzog F, Rinner O, Ellenberg J, Aebersold R. *Mol Syst Biol.* 2011; 7:549. [PubMed: 22068332]
422. Nagaraj N, Wisniewski JR, Geiger T, Cox J, Kircher M, Kelso J, Paabo S, Mann M. *Mol Syst Biol.* 2011; 7:548. [PubMed: 22068331]
423. Corthals GL, Wasinger VC, Hochstrasser DF, Sanchez JC. *Electrophoresis.* 2000; 21:1104. [PubMed: 10786884]
424. Bondarenko PV, Chelius D, Shaler TA. *Anal Chem.* 2002; 74:4741. [PubMed: 12349978]
425. Chelius D, Bondarenko PV. *J Proteome Res.* 2002; 1:317. [PubMed: 12645887]
426. Wang W, Zhou H, Lin H, Roy S, Shaler TA, Hill LR, Norton S, Kumar P, Anderle M, Becker CH. *Anal Chem.* 2003; 75:4818. [PubMed: 14674459]

427. Wiener MC, Sachs JR, Deyanova EG, Yates NA. *Anal Chem.* 2004; 76:6085. [PubMed: 15481957]
428. Higgs RE, Knierman MD, Gelfanova V, Butler JP, Hale JE. *Methods Mol Biol.* 2008; 428:209. [PubMed: 18287776]
429. Liu H, Sadygov RG, Yates JR 3rd. *Anal Chem.* 2004; 76:4193. [PubMed: 15253663]
430. Gilchrist A, Au CE, Hiding J, Bell AW, Fernandez-Rodriguez J, Lesimple S, Nagaya H, Roy L, Gosline SJ, Hallett M, Paiement J, Kearney RE, Nilsson T, Bergeron JJ. *Cell.* 2006; 127:1265. [PubMed: 17174899]
431. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. *Nat Biotechnol.* 1999; 17:994. [PubMed: 10504701]
432. Smolka MB, Zhou H, Purkayastha S, Aebersold R. *Anal Biochem.* 2001; 297:25. [PubMed: 11567524]
433. Regnier FE, Riggs L, Zhang R, Xiong L, Liu P, Chakraborty A, Seeley E, Sioma C, Thompson RA. *J Mass Spectrom.* 2002; 37:133. [PubMed: 11857757]
434. Oda Y, Owa T, Sato T, Boucher B, Daniels S, Yamanaka H, Shinohara Y, Yokoi A, Kuromitsu J, Nagasu T. *Anal Chem.* 2003; 75:2159. [PubMed: 12720356]
435. Li J, Steen H, Gygi SP. *Mol Cell Proteomics.* 2003; 2:1198. [PubMed: 14506205]
436. Hansen KC, Schmitt-Ulms G, Chalkley RJ, Hirsch J, Baldwin MA, Burlingame AL. *Mol Cell Proteomics.* 2003; 2:299. [PubMed: 12766231]
437. Schmidt A, Kellermann J, Lottspeich F. *Proteomics.* 2005; 5:4. [PubMed: 15602776]
438. Hsu JL, Huang SY, Chow NH, Chen SH. *Anal Chem.* 2003; 75:6843. [PubMed: 14670044]
439. Boersema PJ, Raijmakers R, Lemeer S, Mohammed S, Heck AJ. *Nat Protoc.* 2009; 4:484. [PubMed: 19300442]
440. Reynolds KJ, Yao X, Fenselau C. *J Proteome Res.* 2002; 1:27. [PubMed: 12643523]
441. Reynolds KJ, Yao XD, Fenselau C. *Journal of Proteome Research.* 2002; 1:27. [PubMed: 12643523]
442. Yao X, Freas A, Ramirez J, Demirev PA, Fenselau C. *Anal Chem.* 2001; 73:2836. [PubMed: 11467524]
443. Julka S, Regnier F. *J Proteome Res.* 2004; 3:350. [PubMed: 15253416]
444. Johnson KL, Muddiman DC. *J Am Soc Mass Spectrom.* 2004; 15:437. [PubMed: 15047049]
445. Ramos-Fernandez A, Lopez-Ferrer D, Vazquez J. *Mol Cell Proteomics.* 2007; 6:1274. [PubMed: 17322307]
446. Yao X, Afonso C, Fenselau C. *J Proteome Res.* 2003; 2:147. [PubMed: 12716128]
447. Mason CJ, Therneau TM, Eckel-Passow JE, Johnson KL, Oberg AL, Olson JE, Nair KS, Muddiman DC, Bergen HR 3rd. *Mol Cell Proteomics.* 2007; 6:305. [PubMed: 17068186]
448. Langen, H.; Fountoulakis, M.; Evers, S.; Wipf, B.; Berndt, P. Wiley-VCH; Weinheim, Germany: 1998.
449. Oda Y, Huang K, Cross FR, Cowburn D, Chait BT. *Proc Natl Acad Sci U S A.* 1999; 96:6591. [PubMed: 10359756]
450. Conrads TP, Alving K, Veenstra TD, Belov ME, Anderson GA, Anderson DJ, Lipton MS, Pasatolic L, Udseth HR, Chrisler WB, Thrall BD, Smith RD. *Anal Chem.* 2001; 73:2132. [PubMed: 11354501]
451. Ishihama Y, Sato T, Tabata T, Miyamoto N, Sagane K, Nagasu T, Oda Y. *Nat Biotechnol.* 2005; 23:617. [PubMed: 15834404]
452. Palmblad M, Bindschedler LV, Cramer R. *Proteomics.* 2007; 7:3462. [PubMed: 17726679]
453. Huttlin EL, Hegeman AD, Harms AC, Sussman MR. *Mol Cell Proteomics.* 2007; 6:860. [PubMed: 17293592]
454. Schaff JE, Mbeunkui F, Blackburn K, Bird DM, Goshe MB. *Plant J.* 2008; 56:840. [PubMed: 18665915]
455. Nelson CJ, Huttlin EL, Hegeman AD, Harms AC, Sussman MR. *Proteomics.* 2007; 7:1279. [PubMed: 17443642]

456. Krijgsveld J, Ketting RF, Mahmoudi T, Johansen J, Artal-Sanz M, Verrijzer CP, Plasterk RH, Heck AJ. *Nat Biotechnol.* 2003; 21:927. [PubMed: 12858183]
457. Dong MQ, Venable JD, Au N, Xu T, Park SK, Cociorva D, Johnson JR, Dillin A, Yates JR 3rd. *Science.* 2007; 317:660. [PubMed: 17673661]
458. McClatchy DB, Dong MQ, Wu CC, Venable JD, Yates JR. *Journal of Proteome Research.* 2007; 6:2005. [PubMed: 17375949]
459. Wu CC, MacCoss MJ, Howell KE, Matthews DE, Yates JR 3rd. *Anal Chem.* 2004; 76:4951. [PubMed: 15373428]
460. Zhang Y, Filiou MD, Reckow S, Gormanns P, Maccarrone G, Kessler MS, Frank E, Hamsch B, Holsboer F, Landgraf R, Turck CW. *Mol Cell Proteomics.* 2011; 10:M111 008110.
461. Liao L, McClatchy DB, Park SK, Xu T, Lu B, Yates JR 3rd. *J Proteome Res.* 2008; 7:4743. [PubMed: 18823140]
462. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M. *Mol Cell Proteomics.* 2002; 1:376. [PubMed: 12118079]
463. Kruger M, Moser M, Ussar S, Thievessen I, Lubner CA, Forner F, Schmidt S, Zanivan S, Fassler R, Mann M. *Cell.* 2008; 134:353. [PubMed: 18662549]
464. Sury MD, Chen JX, Selbach M. *Mol Cell Proteomics.* 2010; 9:2173. [PubMed: 20525996]
465. Hilger M, Mann M. *J Proteome Res.* 2012; 11:982. [PubMed: 22011079]
466. Tomazela DM, Patterson BW, Hanson E, Spence KL, Kanion TB, Salinger DH, Vicini P, Barret H, Heins HB, Cole FS, Hamvas A, MacCoss MJ. *Anal Chem.* 2010; 82:2561. [PubMed: 20178338]
467. Liao L, Sando RC, Farnum JB, Vanderklish PW, Maximov A, Yates JR. *J Proteome Res.* 2012; 11:1341. [PubMed: 22070516]
468. Geiger T, Cox J, Ostasiewicz P, Wisniewski JR, Mann M. *Nat Methods.* 2010; 7:383. [PubMed: 20364148]
469. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlet-Jones M, He F, Jacobson A, Pappin DJ. *Mol Cell Proteomics.* 2004; 3:1154. [PubMed: 15385600]
470. Thompson A, Schafer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, Neumann T, Johnstone R, Mohammed AK, Hamon C. *Anal Chem.* 2003; 75:1895. [PubMed: 12713048]
471. Dayon L, Hainard A, Licker V, Turck N, Kuhn K, Hochstrasser DF, Burkhard PR, Sanchez JC. *Anal Chem.* 2008; 80:2921. [PubMed: 18312001]
472. Ow SY, Cardona T, Taton A, Magnuson A, Lindblad P, Stensjo K, Wright PC. *J Proteome Res.* 2008; 7:1615. [PubMed: 18290607]
473. Louris JN, Cooks RG, Syka JEP, Kelley PE, Stafford GC, Todd JFJ. *Analytical Chemistry.* 1987; 59:1677.
474. Cunningham C, Glish GL, Burinsky DJ. *Journal of the American Society for Mass Spectrometry.* 2006; 17:81. [PubMed: 16352436]
475. Kocher T, Pichler P, Schutzbier M, Stingl C, Kaul A, Teucher N, Hasenfuss G, Penninger JM, Mechtler K. *J Proteome Res.* 2009; 8:4743. [PubMed: 19663507]
476. Ow SY, Salim M, Noirel J, Evans C, Rehman I, Wright PC. *J Proteome Res.* 2009; 8:5347. [PubMed: 19754192]
477. Karp NA, Huber W, Sadowski PG, Charles PD, Hester SV, Lilley KS. *Mol Cell Proteomics.* 2010; 9:1885. [PubMed: 20382981]
478. Ting L, Rad R, Gygi SP, Haas W. *Nat Methods.* 2011; 8:937. [PubMed: 21963607]
479. Wenger CD, Lee MV, Hebert AS, McAlister GC, Phanstiel DH, Westphall MS, Coon JJ. *Nat Methods.* 2011; 8:933. [PubMed: 21963608]
480. Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B. *Anal Bioanal Chem.* 2007; 389:1017. [PubMed: 17668192]
481. Asara JM, Christofk HR, Freemark LM, Cantley LC. *Proteomics.* 2008; 8:994. [PubMed: 18324724]
482. Nesvizhskii AI, Keller A, Kolker E, Aebersold R. *Anal Chem.* 2003; 75:4646. [PubMed: 14632076]

483. Cutillas PR, Vanhaesebroeck B. *Mol Cell Proteomics*. 2007; 6:1560. [PubMed: 17565973]
484. Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M. *Mol Cell Proteomics*. 2005; 4:1265. [PubMed: 15958392]
485. Braisted JC, Kuntumalla S, Vogel C, Marcotte EM, Rodrigues AR, Wang R, Huang ST, Ferlanti ES, Saeed AI, Fleischmann RD, Peterson SN, Pieper R. *BMC Bioinformatics*. 2008; 9:529. [PubMed: 19068132]
486. Malmstrom J, Beck M, Schmidt A, Lange V, Deutsch EW, Aebersold R. *Nature*. 2009; 460:762. [PubMed: 19606093]
487. Podwojski K, Eisenacher M, Kohl M, Turewicz M, Meyer HE, Rahnenfuhrer J, Stephan C. *Expert Rev Proteomics*. 2010; 7:249. [PubMed: 20377391]
488. Listgarten J, Emili A. *Mol Cell Proteomics*. 2005; 4:419. [PubMed: 15741312]
489. Silva JC, Denny R, Dorschel CA, Gorenstein M, Kass IJ, Li GZ, McKenna T, Nold MJ, Richardson K, Young P, Geromanos S. *Anal Chem*. 2005; 77:2187. [PubMed: 15801753]
490. Griffin NM, Yu J, Long F, Oh P, Shore S, Li Y, Koziol JA, Schnitzer JE. *Nat Biotechnol*.
491. Carvalho PC, Han X, Xu T, Cociorva D, da Carvalho MG, Barbosa VC, Yates JR 3rd . *Bioinformatics*. 2010; 26:847. [PubMed: 20106817]
492. Mueller LN, Brusniak MY, Mani DR, Aebersold R. *J Proteome Res*. 2008; 7:51. [PubMed: 18173218]
493. Mann B, Madera M, Sheng Q, Tang H, Mechref Y, Novotny MV. *Rapid Commun Mass Spectrom*. 2008; 22:3823. [PubMed: 18985620]
494. Tsou CC, Tsai CF, Tsui YH, Sudhir PR, Wang YT, Chen YJ, Chen JY, Sung TY, Hsu WL. *Mol Cell Proteomics*. 2010; 9:131. [PubMed: 19752006]
495. Eng JK, McCormack AL, Yates JR III . *J Am Soc Mass Spectrom*. 1994; 5:976.
496. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. *Electrophoresis*. 1999; 20:3551. [PubMed: 10612281]
497. Craig R, Beavis RC. *Bioinformatics*. 2004; 20:1466. [PubMed: 14976030]
498. Mueller LN, Rinner O, Schmidt A, Letarte S, Bodenmiller B, Brusniak MY, Vitek O, Aebersold R, Muller M. *Proteomics*. 2007; 7:3470. [PubMed: 17726677]
499. Jaffe JD, Mani DR, Leptos KC, Church GM, Gillette MA, Carr SA. *Mol Cell Proteomics*. 2006; 5:1927. [PubMed: 16857664]
500. Park SK, Venable JD, Xu T, Yates JR 3rd . *Nat Methods*. 2008; 5:319. [PubMed: 18345006]
501. Lai X, Wang L, Tang H, Witzmann FA. *J Proteome Res*. 2011; 10:4799. [PubMed: 21888428]
502. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, Kern R, Tabb DL, Liebler DC, MacCoss MJ. *Bioinformatics*. 2010; 26:966. [PubMed: 20147306]
503. Schilling B, Rardin MJ, MacLean BX, Zawadzka AM, Frewen BE, Cusack MP, Sorensen DJ, Bereman MS, Jing E, Wu CC, Verdin E, Kahn CR, Maccoss MJ, Gibson BW. *Mol Cell Proteomics*. 2012; 11:202. [PubMed: 22454539]
504. Bowers GN Jr, Fassett JD, White Et. *Anal Chem*. 1993; 65:475R.
505. Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP. *Proc Natl Acad Sci U S A*. 2003; 100:6940. [PubMed: 12771378]
506. Beynon RJ, Doherty MK, Pratt JM, Gaskell SJ. *Nat Methods*. 2005; 2:587. [PubMed: 16094383]
507. Brun V, Dupuis A, Adrait A, Marcellin M, Thomas D, Court M, Vandenesch F, Garin J. *Mol Cell Proteomics*. 2007; 6:2139. [PubMed: 17848587]
508. Hanke S, Besir H, Oesterhelt D, Mann M. *J Proteome Res*. 2008; 7:1118. [PubMed: 18271523]
509. Coppinger JA, Hutt DM, Razvi A, Koulov AV, Pankow S, Yates JR 3rd, Balch WE. *PLoS One*. 2012; 7:e37682. [PubMed: 22701530]
510. Kirkpatrick DS, Gerber SA, Gygi SP. *Methods*. 2005; 35:265. [PubMed: 15722223]
511. Gerber SA, Kettenbach AN, Rush J, Gygi SP. *Methods Mol Biol*. 2007; 359:71. [PubMed: 17484111]
512. Mayya V, Rezual K, Wu L, Fong MB, Han DK. *Mol Cell Proteomics*. 2006; 5:1146. [PubMed: 16546994]
513. Havlis J, Shevchenko A. *Anal Chem*. 2004; 76:3029. [PubMed: 15167779]

514. Brownridge P, Holman SW, Gaskell SJ, Grant CM, Harman VM, Hubbard SJ, Lanthaler K, Lawless C, O’Cualain R, Sims P, Watkins R, Beynon RJ. *Proteomics*. 2011; 11:2957. [PubMed: 21710569]
515. Zhao Y, Jia W, Sun W, Jin W, Guo L, Wei J, Ying W, Zhang Y, Xie Y, Jiang Y, He F, Qian X. *J Proteome Res*. 2010; 9:3319. [PubMed: 20420461]
516. Dephoure N, Gygi SP. *Sci Signal*. 2012; 5:rs2. [PubMed: 22457332]
517. Noble WS. *Nat Biotechnol*. 2009; 27:1135. [PubMed: 20010596]
518. Pavelka N, Fournier ML, Swanson SK, Pelizzola M, Ricciardi-Castagnoli P, Florens L, Washburn MP. *Mol Cell Proteomics*. 2008; 7:631. [PubMed: 18029349]
519. Hill EG, Schwacke JH, Comte-Walters S, Slate EH, Oberg AL, Eckel-Passow JE, Therneau TM, Schey KL. *J Proteome Res*. 2008; 7:3091. [PubMed: 18578521]
520. Oberg AL, Mahoney DW, Eckel-Passow JE, Malone CJ, Wolfinger RD, Hill EG, Cooper LT, Onuma OK, Spiro C, Therneau TM, Bergen HR 3rd. *J Proteome Res*. 2008; 7:225. [PubMed: 18173221]
521. Bonferroni CE. *Publicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*. 1936; 8:3.
522. Benjamini Y, Hochberg Y. *Journal of the Royal Statistical Society Series B-Methodological*. 1995; 57:289.
523. Storey JD, Tibshirani R. *Proc Natl Acad Sci U S A*. 2003; 100:9440. [PubMed: 12883005]
524. Margolin AA, Ong SE, Schenone M, Gould R, Schreiber SL, Carr SA, Golub TR. *PLoS One*. 2009; 4:e7454. [PubMed: 19829701]
525. Zhang B, VerBerkmoes NC, Langston MA, Uberbacher E, Hettich RL, Samatova NF. *J Proteome Res*. 2006; 5:2909. [PubMed: 17081042]
526. Choi H, Fermin D, Nesvizhskii AI. *Mol Cell Proteomics*. 2008; 7:2373. [PubMed: 18644780]
527. Dai L, Li C, Shedden KA, Lee CJ, Quoc H, Simeone DM, Lubman DM. *J Proteome Res*. 2010; 9:3394. [PubMed: 20486718]
528. Wang LH, Li DQ, Fu Y, Wang HP, Zhang JF, Yuan ZF, Sun RX, Zeng R, He SM, Gao W. *Rapid Commun Mass Spectrom*. 2007; 21:2985. [PubMed: 17702057]
529. Cox J, Matic I, Hilger M, Nagaraj N, Selbach M, Olsen JV, Mann M. *Nat Protoc*. 2009; 4:698. [PubMed: 19373234]
530. Deutsch EW, Mendoza L, Shteynberg D, Farrah T, Lam H, Tasman N, Sun Z, Nilsson E, Pratt B, Prazan B, Eng JK, Martin DB, Nesvizhskii AI, Aebersold R. *Proteomics*. 2010; 10:1150. [PubMed: 20101611]
531. Bell AW, Deutsch EW, Au CE, Kearney RE, Beavis R, Sechi S, Nilsson T, Bergeron JJ. *Nat Methods*. 2009; 6:423. [PubMed: 19448641]
532. Kinsinger CR, Apffel J, Baker M, Bian X, Borchers CH, Bradshaw R, Brusniak MY, Chan DW, Deutsch EW, Domon B, Gorman J, Grimm R, Hancock W, Hermjakob H, Horn D, Hunter C, Kolar P, Kraus HJ, Langen H, Linding R, Moritz RL, Omenn GS, Orlando R, Pandey A, Ping P, Rahbar A, Rivers R, Seymour SL, Simpson RJ, Slotta D, Smith RD, Stein SE, Tabb DL, Tagle D, Yates JR, Rodriguez H. *J Proteome Res*. 2012; 11:1412. [PubMed: 22053864]
533. Han MJ, Lee SY. *Microbiol Mol Biol Rev*. 2006; 70:362. [PubMed: 16760308]
534. Paulovich AG, Billheimer D, Ham AJ, Vega-Montoto L, Rudnick PA, Tabb DL, Wang P, Blackman RK, Bunk DM, Cardasis HL, Clauser KR, Kinsinger CR, Schilling B, Tegeler TJ, Variyath AM, Wang M, Whiteaker JR, Zimmerman LJ, Fenyó D, Carr SA, Fisher SJ, Gibson BW, Mesri M, Neubert TA, Regnier FE, Rodriguez H, Spiegelman C, Stein SE, Tempst P, Liebler DC. *Mol Cell Proteomics*. 2010; 9:242. [PubMed: 19858499]
535. Bohrer BC, Li YF, Reilly JP, Clemmer DE, DiMarchi RD, Radivojac P, Tang H, Arnold RJ. *Anal Chem*. 2010; 82:6559. [PubMed: 20669997]
536. Vaudel M, Burkhardt JM, Breiter D, Zahedi RP, Sickmann A, Martens L. *J Proteome Res*. 2012; 11:5065. [PubMed: 22489649]
537. Rudnick PA, Clauser KR, Kilpatrick LE, Tchekhovskoi DV, Neta P, Blonder N, Billheimer DD, Blackman RK, Bunk DM, Cardasis HL, Ham AJ, Jaffe JD, Kinsinger CR, Mesri M, Neubert TA, Schilling B, Tabb DL, Tegeler TJ, Vega-Montoto L, Variyath AM, Wang M, Wang P, Whiteaker

- JR, Zimmerman LJ, Carr SA, Fisher SJ, Gibson BW, Paulovich AG, Regnier FE, Rodriguez H, Spiegelman C, Tempst P, Liebler DC, Stein SE. *Mol Cell Proteomics*. 2010; 9:225. [PubMed: 19837981]
538. Sigmon I, Lee LW, Chang DK, Krusberski N, Cohen D, Eng JK, Martin DB. *Anal Chem*. 2010; 82:5060. [PubMed: 20503988]
539. Scheltema RA, Mann M. *J Proteome Res*. 2012
540. Barboza R, Cociorva D, Xu T, Barbosa VC, Perales J, Valente RH, Franca FM, Yates JR 3rd, Carvalho PC. *Proteomics*. 2011; 11:4105. [PubMed: 21834134]
541. Yates JR 3rd, Park SK, Delahunty CM, Xu T, Savas JN, Cociorva D, Carvalho PC. *Nat Methods*. 2012; 9:455. [PubMed: 22543378]
542. Hunt DF, Yates JR 3rd, Shabanowitz J, Winston S, Hauer CR. *Proc Natl Acad Sci U S A*. 1986; 83:6233. [PubMed: 3462691]
543. Sadygov RG, Cociorva D, Yates JR 3rd. *Nat Methods*. 2004; 1:195. [PubMed: 15789030]
544. Choi H, Ghosh D, Nesvizhskii AI. *J Proteome Res*. 2008; 7:286. [PubMed: 18078310]
545. Choi H, Nesvizhskii AI. *J Proteome Res*. 2008; 7:254. [PubMed: 18159924]
546. Fenyo D, Beavis RC. *Anal Chem*. 2003; 75:768. [PubMed: 12622365]
547. Craig R, Beavis RC. *Rapid Commun Mass Spectrom*. 2003; 17:2310. [PubMed: 14558131]
548. Alves G, Wu WW, Wang G, Shen RF, Yu YK. *J Proteome Res*. 2008; 7:3102. [PubMed: 18558733]
549. Searle BC, Turner M, Nesvizhskii AI. *J Proteome Res*. 2008; 7:245. [PubMed: 18173222]
550. Narasimhan C, Tabb DL, Verberkmoes NC, Thompson MR, Hettich RL, Uberbacher EC. *Anal Chem*. 2005; 77:7581. [PubMed: 16316165]
551. Tabb DL, Fernando CG, Chambers MC. *J Proteome Res*. 2007; 6:654. [PubMed: 17269722]
552. Li W, Ji L, Goya J, Tan G, Wysocki VH. *J Proteome Res*. 2011; 10:1593. [PubMed: 21204564]
553. Sadygov RG, Eng J, Durr E, Saraf A, McDonald H, MacCoss MJ, Yates JR 3rd. *J Proteome Res*. 2002; 1:211. [PubMed: 12645897]
554. Tabb DL, Narasimhan C, Strader MB, Hettich RL. *Anal Chem*. 2005; 77:2464. [PubMed: 15828782]
555. Eng JK, Fischer B, Grossmann J, MacCoss MJ. *J Proteome Res*. 2008; 7:4598. [PubMed: 18774840]
556. Diamant BJ, Noble WS. *J Proteome Res*. 2011; 10:3871. [PubMed: 21761931]
557. Xu T, Venable JD, Park SK, Cociorva D, Lu B, Liao L, Wohlschlegel J, Hewel J, Yates JR 3rd. *Mol Cell Proteomics*. 2006; 5:S174.
558. Lu B, Motoyama A, Ruse C, Venable J, Yates JR 3rd. *Anal Chem*. 2008; 80:2018. [PubMed: 18275164]
559. Savitski MM, Mathieson T, Becher I, Bantscheff M. *J Proteome Res*. 2010; 9:5511. [PubMed: 20836569]
560. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. *J Proteome Res*. 2011; 10:1794. [PubMed: 21254760]
561. Brosch M, Swamy S, Hubbard T, Choudhary J. *Mol Cell Proteomics*. 2008; 7:962. [PubMed: 18216375]
562. Arifuzzaman M, Maeda M, Itoh A, Nishikata K, Takita C, Saito R, Ara T, Nakahigashi K, Huang HC, Hirai A, Tsuzuki K, Nakamura S, Altaf-Ul-Amin M, Oshima T, Baba T, Yamamoto N, Kawamura T, Ioka-Nakamichi T, Kitagawa M, Tomita M, Kanaya S, Wada C, Mori H. *Genome Res*. 2006; 16:686. [PubMed: 16606699]
563. Graumann J, Scheltema RA, Zhang Y, Cox J, Mann M. *Mol Cell Proteomics*. 2012; 11:M111013185. [PubMed: 22171319]
564. Bailey DJ, Rose CM, McAlister GC, Brumbaugh J, Yu P, Wenger CD, Westphall MS, Thomson JA, Coon JJ. *Proc Natl Acad Sci U S A*. 2012; 109:8411. [PubMed: 22586074]
565. Xu M, Li L. *J Proteome Res*. 2011; 10:3632. [PubMed: 21726074]
566. Tabb DL, Saraf A, Yates JR 3rd. *Anal Chem*. 2003; 75:6415. [PubMed: 14640709]
567. Frank A, Tanner S, Bafna V, Pevzner P. *J Proteome Res*. 2005; 4:1287. [PubMed: 16083278]

568. Cao X, Nesvizhskii AI. *J Proteome Res.* 2008; 7:4422. [PubMed: 18785767]
569. Tabb DL, Ma ZQ, Martin DB, Ham AJ, Chambers MC. *J Proteome Res.* 2008; 7:3838. [PubMed: 18630943]
570. Tanner S, Shu H, Frank A, Wang LC, Zandi E, Mumby M, Pevzner PA, Bafna V. *Anal Chem.* 2005; 77:4626. [PubMed: 16013882]
571. Dasari S, Chambers MC, Slebos RJ, Zimmerman LJ, Ham AJ, Tabb DL. *J Proteome Res.* 2010; 9:1716. [PubMed: 20131910]
572. Dasari S, Chambers MC, Codreanu SG, Liebler DC, Collins BC, Pennington SR, Gallagher WM, Tabb DL. *Chem Res Toxicol.* 2011; 24:204. [PubMed: 21214251]
573. Sweet SM, Jones AW, Cunningham DL, Heath JK, Creese AJ, Cooper HJ. *J Proteome Res.* 2009; 8:5475. [PubMed: 19821632]
574. Bailey CM, Sweet SM, Cunningham DL, Zeller M, Heath JK, Cooper HJ. *J Proteome Res.* 2009; 8:1965. [PubMed: 19275241]
575. Sun RX, Dong MQ, Song CQ, Chi H, Yang B, Xiu LY, Tao L, Jing ZY, Liu C, Wang LH, Fu Y, He SM. *J Proteome Res.* 2010; 9:6354. [PubMed: 20883037]
576. Sadygov RG, Good DM, Swaney DL, Coon JJ. *J Proteome Res.* 2009; 8:3198. [PubMed: 19354237]
577. Good DM, Wenger CD, McAlister GC, Bai DL, Hunt DF, Coon JJ. *J Am Soc Mass Spectrom.* 2009; 20:1435. [PubMed: 19362853]
578. Liu WT, Ng J, Meluzzi D, Bandeira N, Gutierrez M, Simmons TL, Schultz AW, Linington RG, Moore BS, Gerwick WH, Pevzner PA, Dorrestein PC. *Anal Chem.* 2009; 81:4200. [PubMed: 19413302]
579. Ng J, Bandeira N, Liu WT, Ghassemian M, Simmons TL, Gerwick WH, Linington R, Dorrestein PC, Pevzner PA. *Nat Methods.* 2009; 6:596. [PubMed: 19597502]
580. Hoopmann MR, Finney GL, MacCoss MJ. *Anal Chem.* 2007; 79:5620. [PubMed: 17580982]
581. Luethy R, Kessner DE, Katz JE, Maclean B, Grothe R, Kani K, Faca V, Pitteri S, Hanash S, Agus DB, Mallick P. *J Proteome Res.* 2008; 7:4031. [PubMed: 18707148]
582. Cox J, Mann M. *J Am Soc Mass Spectrom.* 2009; 20:1477. [PubMed: 19553133]
583. Clauser KR, Baker P, Burlingame AL. *Anal Chem.* 1999; 71:2871. [PubMed: 10424174]
584. Wenger CD, Coon JJ. *J Proteome Res.* 2013
585. Carvalho PC, Cociorva D, Wong CC, da Carvalho MG, Barbosa VC, Yates JR 3rd. *Anal Chem.* 2009; 81:1996. [PubMed: 19203245]
586. Sharma V, Eng JK, Feldman S, von Haller PD, MacCoss MJ, Noble WS. *J Proteome Res.* 2010; 9:5438. [PubMed: 20731383]
587. Keller A, Nesvizhskii AI, Kolker E, Aebersold R. *Anal Chem.* 2002; 74:5383. [PubMed: 12403597]
588. Hansen TA, Kryuchkov F, Kjeldsen F. *Anal Chem.* 2012; 84:6638. [PubMed: 22799558]
589. Tabb DL, Smith LL, Brechi LA, Wysocki VH, Lin D, Yates JR 3rd. *Anal Chem.* 2003; 75:1155. [PubMed: 12641236]
590. Tabb DL, Huang Y, Wysocki VH, Yates JR 3rd. *Anal Chem.* 2004; 76:1243. [PubMed: 14987077]
591. Chalkley RJ, Medzihradzky KF, Lynn AJ, Baker PR, Burlingame AL. *Anal Chem.* 2010; 82:579. [PubMed: 20028093]
592. Li W, Song C, Bailey DJ, Tseng GC, Coon JJ, Wysocki VH. *Anal Chem.* 2011; 83:9540. [PubMed: 22022956]
593. Michalski A, Neuhauser N, Cox J, Mann M. *J Proteome Res.* 2012; 11:5479. [PubMed: 22998608]
594. Taylor JA, Johnson RS. *Anal Chem.* 2001; 73:2594. [PubMed: 11403305]
595. Frank A, Pevzner P. *Anal Chem.* 2005; 77:964. [PubMed: 15858974]
596. Waridel P, Frank A, Thomas H, Surendranath V, Sunyaev S, Pevzner P, Shevchenko A. *Proteomics.* 2007; 7:2318. [PubMed: 17623296]
597. Taylor JA, Johnson RS. *Rapid Commun Mass Spectrom.* 1997; 11:1067. [PubMed: 9204580]

598. Kim S, Gupta N, Bandeira N, Pevzner PA. *Mol Cell Proteomics*. 2009; 8:53. [PubMed: 18703573]
599. Bern M, Cai Y, Goldberg D. *Anal Chem*. 2007; 79:1393. [PubMed: 17243770]
600. Zhang J, Xin L, Shan B, Chen W, Xie M, Yuen D, Zhang W, Zhang Z, Lajoie GA, Ma B. *Mol Cell Proteomics*. 2012; 11:M111 010587.
601. Chi H, Sun RX, Yang B, Song CQ, Wang LH, Liu C, Fu Y, Yuan ZF, Wang HP, He SM, Dong MQ. *J Proteome Res*. 2010; 9:2713. [PubMed: 20329752]
602. van Breukelen B, Georgiou A, Drugan MM, Taouatas N, Mohammed S, Heck AJ. *Proteomics*. 2010; 10:1196. [PubMed: 20077410]
603. Elias JE, Gygi SP. *Nat Methods*. 2007; 4:207. [PubMed: 17327847]
604. Gupta N, Bandeira N, Keich U, Pevzner PA. *J Am Soc Mass Spectrom*. 2011; 22:1111. [PubMed: 21953092]
605. Gupta N, Pevzner PA. *J Proteome Res*. 2009; 8:4173. [PubMed: 19627159]
606. Kall L, Canterbury JD, Weston J, Noble WS, MacCoss MJ. *Nat Methods*. 2007; 4:923. [PubMed: 17952086]
607. Spivak M, Weston J, Bottou L, Kall L, Noble WS. *J Proteome Res*. 2009; 8:3737. [PubMed: 19385687]
608. Kall L, Storey JD, MacCoss MJ, Noble WS. *J Proteome Res*. 2008; 7:29. [PubMed: 18067246]
609. Spivak M, Weston J, Tomazela D, MacCoss MJ, Noble WS. *Mol Cell Proteomics*. 2012; 11:M111 012161. [PubMed: 22052992]
610. Bern MW, Kil YJ. *J Proteome Res*. 2011; 10:5296. [PubMed: 22010998]
611. Reiter L, Claassen M, Schrimpf SP, Jovanovic M, Schmidt A, Buhmann JM, Hengartner MO, Aebersold R. *Mol Cell Proteomics*. 2009; 8:2405. [PubMed: 19608599]
612. Wang X, Slebos RJ, Wang D, Halvey PJ, Tabb DL, Liebler DC, Zhang B. *J Proteome Res*. 2012; 11:1009. [PubMed: 22103967]
613. Volkening JD, Bailey DJ, Rose CM, Grimsrud PA, Howes-Podoll M, Venkateshwaran M, Westphall MS, Ane JM, Coon JJ, Sussman MR. *Mol Cell Proteomics*. 2012; 11:933. [PubMed: 22774004]
614. McCormack AL, Schieltz DM, Goode B, Yang S, Barnes G, Drubin D, Yates JR 3rd. *Anal Chem*. 1997; 69:767. [PubMed: 9043199]
615. Borchert N, Dieterich C, Krug K, Schutz W, Jung S, Nordheim A, Sommer RJ, Macek B. *Genome Res*. 2010; 20:837. [PubMed: 20237107]
616. Link AJ, Hays LG, Carmack EB, Yates JR 3rd. *Electrophoresis*. 1997; 18:1314. [PubMed: 9298647]
617. Pawar H, Sahasrabudhe NA, Renuse S, Keerthikumar S, Sharma J, Kumar GS, Venugopal A, Sekhar NR, Kelkar DS, Nemade H, Khobragade SN, Muthusamy B, Kandasamy K, Harsha HC, Chaerkady R, Patole MS, Pandey A. *Proteomics*. 2012; 12:832. [PubMed: 22539434]
618. Cappellini E, Jensen LJ, Szklarczyk D, Ginolhac A, da Fonseca RA, Stafford TW, Holen SR, Collins MJ, Orlando L, Willerslev E, Gilbert MT, Olsen JV. *J Proteome Res*. 2012; 11:917. [PubMed: 22103443]
619. Asara JM, Garavelli JS, Slatter DA, Schweitzer MH, Freimark LM, Phillips M, Cantley LC. *Science*. 2007; 317:1324. [PubMed: 17823333]
620. Asara JM, Schweitzer MH, Freimark LM, Phillips M, Cantley LC. *Science*. 2007; 316:280. [PubMed: 17431180]
621. Liska AJ, Popov AV, Sunyaev S, Coughlin P, Habermann B, Shevchenko A, Bork P, Karsenti E. *Proteomics*. 2004; 4:2707. [PubMed: 15352245]
622. Grossmann J, Fischer B, Baerenfaller K, Owiti J, Buhmann JM, Gruissem W, Baginsky S. *Proteomics*. 2007; 7:4245. [PubMed: 18040981]
623. Oshiro G, Wodicka LM, Washburn MP, Yates JR 3rd, Lockhart DJ, Winzeler EA. *Genome Res*. 2002; 12:1210. [PubMed: 12176929]
624. Merrihew GE, Davis C, Ewing B, Williams G, Kall L, Frewen BE, Noble WS, Green P, Thomas JH, MacCoss MJ. *Genome Res*. 2008; 18:1660. [PubMed: 18653799]

625. Brosch M, Saunders GI, Frankish A, Collins MO, Yu L, Wright J, Verstraten R, Adams DJ, Harrow J, Choudhary JS, Hubbard T. *Genome Res.* 2011; 21:756. [PubMed: 21460061]
626. Castellana NE, Payne SH, Shen Z, Stanke M, Bafna V, Briggs SP. *Proc Natl Acad Sci U S A.* 2008; 105:21034. [PubMed: 19098097]
627. Li J, Zimmerman LJ, Park BH, Tabb DL, Liebler DC, Zhang B. *Mol Syst Biol.* 2009; 5:303. [PubMed: 19690572]
628. Carr S, Aebersold R, Baldwin M, Burlingame A, Clauser K, Nesvizhskii A. *Mol Cell Proteomics.* 2004; 3:531. [PubMed: 15075378]
629. Latterich M. *Proteome Sci.* 2006; 4:8. [PubMed: 16643670]
630. Wilkins MR, Appel RD, Van Eyk JE, Chung MC, Gorg A, Hecker M, Huber LA, Langen H, Link AJ, Paik YK, Patterson SD, Pennington SR, Rabilloud T, Simpson RJ, Weiss W, Dunn MJ. *Proteomics.* 2006; 6:4. [PubMed: 16400714]
631. Tabb DL. *J Proteome Res.* 2008; 7:45. [PubMed: 18081243]
632. Binz PA, Barkovich R, Beavis RC, Creasy D, Horn DM, Julian RK Jr, Seymour SL, Taylor CF, Vandenbrouck Y. *Nat Biotechnol.* 2008; 26:862. [PubMed: 18688233]
633. Tabb DL, McDonald WH, Yates JR 3rd. *J Proteome Res.* 2002; 1:21. [PubMed: 12643522]
634. Resing KA, Meyer-Arendt K, Mendoza AM, Aveline-Wolf LD, Jonscher KR, Pierce KG, Old WM, Cheung HT, Russell S, Wattawa JL, Goehle GR, Knight RD, Ahn NG. *Anal Chem.* 2004; 76:3556. [PubMed: 15228325]
635. Zhang B, Chambers MC, Tabb DL. *J Proteome Res.* 2007; 6:3549. [PubMed: 17676885]
636. Ma ZQ, Dasari S, Chambers MC, Litton MD, Sobecki SM, Zimmerman LJ, Halvey PJ, Schilling B, Drake PM, Gibson BW, Tabb DL. *J Proteome Res.* 2009; 8:3872. [PubMed: 19522537]
637. Qeli E, Ahrens CH. *Nat Biotechnol.* 2010; 28:647. [PubMed: 20622826]
638. Feng J, Naiman DQ, Cooper B. *Anal Chem.* 2007; 79:3901. [PubMed: 17441689]
639. Bandeira N, Tang H, Bafna V, Pevzner P. *Anal Chem.* 2004; 76:7221. [PubMed: 15595863]
640. Bandeira N, Clauser KR, Pevzner PA. *Mol Cell Proteomics.* 2007; 6:1123. [PubMed: 17446555]
641. Bandeira N, Tsur D, Frank A, Pevzner PA. *Proc Natl Acad Sci U S A.* 2007; 104:6140. [PubMed: 17404225]
642. Frank AM, Bandeira N, Shen Z, Tanner S, Briggs SP, Smith RD, Pevzner PA. *J Proteome Res.* 2008; 7:113. [PubMed: 18067247]
643. Nesvizhskii AI, Aebersold R. *Mol Cell Proteomics.* 2005; 4:1419. [PubMed: 16009968]
644. Fermin D, Basur V, Yocum AK, Nesvizhskii AI. *Proteomics.* 2011; 11:1340. [PubMed: 21360675]
645. Zhang Y, Wen Z, Washburn MP, Florens L. *Anal Chem.* 2010; 82:2272. [PubMed: 20166708]
646. Lange V, Picotti P, Domon B, Aebersold R. *Mol Syst Biol.* 2008; 4:222. [PubMed: 18854821]
647. Picotti P, Rinner O, Stallmach R, Dautel F, Farrah T, Domon B, Wenschuh H, Aebersold R. *Nat Methods.* 2010; 7:43. [PubMed: 19966807]
648. Domon B, Aebersold R. *Nat Biotechnol.* 2010; 28:710. [PubMed: 20622845]
649. Jin S, Daly DS, Springer DL, Miller JH. *J Proteome Res.* 2008; 7:164. [PubMed: 18001079]
650. Dost B, Bandeira N, Li X, Shen Z, Briggs SP, Bafna V. *J Comput Biol.* 2012; 19:337. [PubMed: 22414154]
651. Noble WS, MacCoss MJ. *PLoS Comput Biol.* 2012; 8:e1002296. [PubMed: 22291580]
652. Wu SL, Kim J, Hancock WS, Karger B. *J Proteome Res.* 2005; 4:1155. [PubMed: 16083266]
653. Boyne MT, Garcia BA, Li M, Zamdborg L, Wenger CD, Babai S, Kelleher NL. *J Proteome Res.* 2009; 8:374. [PubMed: 19053528]
654. Cannon J, Lohnes K, Wynne C, Wang Y, Edwards N, Fenselau C. *J Proteome Res.* 2010; 9:3886. [PubMed: 20557100]
655. Fields S, Song O. *Nature.* 1989; 340:245. [PubMed: 2547163]
656. Ito T, Tashiro K, Muta S, Ozawa R, Chiba T, Nishizawa M, Yamamoto K, Kuhara S, Sakaki Y. *Proc Natl Acad Sci U S A.* 2000; 97:1143. [PubMed: 10655498]
657. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y. *Proc Natl Acad Sci U S A.* 2001; 98:4569. [PubMed: 11283351]

658. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, Qureshi-Emili A, Li Y, Godwin B, Conover D, Kalbfleisch T, Vijayadamodar G, Yang M, Johnston M, Fields S, Rothberg JM. *Nature*. 2000; 403:623. [PubMed: 10688190]
659. Walhout AJ, Sordella R, Lu X, Hartley JL, Temple GF, Brasch MA, Thierry-Mieg N, Vidal M. *Science*. 2000; 287:116. [PubMed: 10615043]
660. Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, Li Y, Hao YL, Ooi CE, Godwin B, Vitols E, Vijayadamodar G, Pochart P, Machineni H, Welsh M, Kong Y, Zerhusen B, Malcolm R, Varrone Z, Collis A, Minto M, Burgess S, McDaniel L, Stimpson E, Spriggs F, Williams J, Neurath K, Ioime N, Agee M, Voss E, Furtak K, Renzulli R, Aanensen N, Carrolla S, Bickelhaupt E, Lazovatsky Y, DaSilva A, Zhong J, Stanyon CA, Finley RL Jr, White KP, Braverman M, Jarvie T, Gold S, Leach M, Knight J, Shimkets RA, McKenna MP, Chant J, Rothberg JM. *Science*. 2003; 302:1727. [PubMed: 14605208]
661. Li S, Armstrong CM, Bertin N, Ge H, Milstein S, Boxem M, Vidalain PO, Han JD, Chesneau A, Hao T, Goldberg DS, Li N, Martinez M, Rual JF, Lamesch P, Xu L, Tewari M, Wong SL, Zhang LV, Berriz GF, Jacotot L, Vaglio P, Reboul J, Hirozane-Kishikawa T, Li Q, Gabel HW, Elewa A, Baumgartner B, Rose DJ, Yu H, Bosak S, Sequerra R, Fraser A, Mango SE, Saxton WM, Strome S, Van Den Heuvel S, Piano F, Vandenhaute J, Sardet C, Gerstein M, Doucette-Stamm L, Gunsalus KC, Harper JW, Cusick ME, Roth FP, Hill DE, Vidal M. *Science*. 2004; 303:540. [PubMed: 14704431]
662. Colland F, Jacq X, Trouplin V, Mouglin C, Groizeleau C, Hamburger A, Meil A, Wojcik J, Legrain P, Gauthier JM. *Genome Res*. 2004; 14:1324. [PubMed: 15231748]
663. Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N, Klitgord N, Simon C, Boxem M, Milstein S, Rosenberg J, Goldberg DS, Zhang LV, Wong SL, Franklin G, Li S, Albala JS, Lim J, Fraughton C, Llamas E, Cevik S, Bex C, Lamesch P, Sikorski RS, Vandenhaute J, Zoghbi HY, Smolyar A, Bosak S, Sequerra R, Doucette-Stamm L, Cusick ME, Hill DE, Roth FP, Vidal M. *Nature*. 2005; 437:1173. [PubMed: 16189514]
664. Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, Stroedicke M, Zenkner M, Schoenherr A, Koeppen S, Timm J, Mintzlaff S, Abraham C, Bock N, Kietzmann S, Goedde A, Toksoz E, Droege A, Krobitsch S, Korn B, Birchmeier W, Lehrach H, Wanker EE. *Cell*. 2005; 122:957. [PubMed: 16169070]
665. Venkatesan K, Rual JF, Vazquez A, Stelzl U, Lemmens I, Hirozane-Kishikawa T, Hao T, Zenkner M, Xin X, Goh KI, Yildirim MA, Simonis N, Heinzmann K, Gebreab F, Sahalie JM, Cevik S, Simon C, de Smet AS, Dann E, Smolyar A, Vinayagam A, Yu H, Szeto D, Borick H, Dricot A, Klitgord N, Murray RR, Lin C, Lalowski M, Timm J, Rau K, Boone C, Braun P, Cusick ME, Roth FP, Hill DE, Tavernier J, Wanker EE, Barabasi AL, Vidal M. *Nat Methods*. 2009; 6:83. [PubMed: 19060904]
666. Yang B, Wu YJ, Zhu M, Fan SB, Lin J, Zhang K, Li S, Chi H, Li YX, Chen HF, Luo SK, Ding YH, Wang LH, Hao Z, Xiu LY, Chen S, Ye K, He SM, Dong MQ. *Nat Methods*. 2012; 9:904. [PubMed: 22772728]
667. Heck AJ. *Nat Methods*. 2008; 5:927. [PubMed: 18974734]
668. Fonslow BR, Kang SA, Gestaut DR, Graczyk B, Davis TN, Sabatini DM, Yates JR 3rd. *Anal Chem*. 2010; 82:6643. [PubMed: 20614870]
669. Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B. *Nat Biotechnol*. 1999; 17:1030. [PubMed: 10504710]
670. Gavin AC, Maeda K, Kuhner S. *Curr Opin Biotechnol*. 2011; 22:42. [PubMed: 20934865]
671. Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, Millar A, Taylor P, Bennett K, Boutilier K, Yang L, Wolting C, Donaldson I, Schandorff S, Shewnarane J, Vo M, Taggart J, Goudreau M, Muskut B, Alfarano C, Dewar D, Lin Z, Michalickova K, Willems AR, Sassi H, Nielsen PA, Rasmussen KJ, Andersen JR, Johansen LE, Hansen LH, Jespersen H, Podtelejnikov A, Nielsen E, Crawford J, Poulsen V, Sorensen BD, Matthiesen J, Hendrickson RC, Gleeson F, Pawson T, Moran MF, Durocher D, Mann M, Hogue CW, Figeys D, Tyers M. *Nature*. 2002; 415:180. [PubMed: 11805837]

672. Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM, Remor M, Hofert C, Schelder M, Brajenovic M, Ruffner H, Merino A, Klein K, Hudak M, Dickson D, Rudi T, Gnau V, Bauch A, Bastuck S, Huhse B, Leutwein C, Heurtier MA, Copley RR, Edelmann A, Querfurth E, Rybin V, Drewes G, Raida M, Bouwmeester T, Bork P, Seraphin B, Kuster B, Neubauer G, Superti-Furga G. *Nature*. 2002; 415:141. [PubMed: 11805826]
673. Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, Jensen LJ, Bastuck S, Dumpelfeld B, Edelmann A, Heurtier MA, Hoffman V, Hoefert C, Klein K, Hudak M, Michon AM, Schelder M, Schirle M, Remor M, Rudi T, Hooper S, Bauer A, Bouwmeester T, Casari G, Drewes G, Neubauer G, Rick JM, Kuster B, Bork P, Russell RB, Superti-Furga G. *Nature*. 2006; 440:631. [PubMed: 16429126]
674. Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, Li J, Pu S, Datta N, Tikuisis AP, Punna T, Peregrin-Alvarez JM, Shales M, Zhang X, Davey M, Robinson MD, Paccanaro A, Bray JE, Sheung A, Beattie B, Richards DP, Canadien V, Lalev A, Mena F, Wong P, Starostine A, Canete MM, Vlasblom J, Wu S, Orsi C, Collins SR, Chandran S, Haw R, Rilstone JJ, Gandi K, Thompson NJ, Musso G, St Onge P, Ghanny S, Lam MH, Butland G, Altaf-Ul AM, Kanaya S, Shilatifard A, O'Shea E, Weissman JS, Ingles CJ, Hughes TR, Parkinson J, Gerstein M, Wodak SJ, Emili A, Greenblatt JF. *Nature*. 2006; 440:637. [PubMed: 16554755]
675. Butland G, Peregrin-Alvarez JM, Li J, Yang W, Yang X, Canadien V, Starostine A, Richards D, Beattie B, Krogan N, Davey M, Parkinson J, Greenblatt J, Emili A. *Nature*. 2005; 433:531. [PubMed: 15690043]
676. Hu P, Janga SC, Babu M, Diaz-Mejia JJ, Butland G, Yang W, Pogoutse O, Guo X, Phanse S, Wong P, Chandran S, Christopoulos C, Nazarians-Armavil A, Nasser NK, Musso G, Ali M, Nazemof N, Eroukova V, Golshani A, Paccanaro A, Greenblatt JF, Moreno-Hagelsieb G, Emili A. *PLoS Biol*. 2009; 7:e96. [PubMed: 19402753]
677. Ewing RM, Chu P, Elisma F, Li H, Taylor P, Climie S, McBroom-Cerajewski L, Robinson MD, O'Connor L, Li M, Taylor R, Dharsee M, Ho Y, Heilbut A, Moore L, Zhang S, Ornatsky O, Bukhman YV, Ethier M, Sheng Y, Vasilescu J, Abu-Farha M, Lambert JP, Duedel HS, Stewart II, Kuehl B, Hogue K, Colwill K, Gladwish K, Muskat B, Kinach R, Adams SL, Moran MF, Morin GB, Topaloglou T, Figeys D. *Mol Syst Biol*. 2007; 3:89. [PubMed: 17353931]
678. Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Goodlett DR, Aebersold R, Hood L. *Science*. 2001; 292:929. [PubMed: 11340206]
679. Kolkman A, Daran-Lapujade P, Fullaondo A, Olsthoorn MM, Pronk JT, Slijper M, Heck AJ. *Mol Syst Biol*. 2006; 2:0026. [PubMed: 16738570]
680. Lu P, Vogel C, Wang R, Yao X, Marcotte EM. *Nat Biotechnol*. 2007; 25:117. [PubMed: 17187058]
681. Schirmer EC, Florens L, Guan T, Yates JR 3rd, Gerace L. *Science*. 2003; 301:1380. [PubMed: 12958361]
682. Arnesen T, Van Damme P, Polevoda B, Helsens K, Evjenth R, Colaert N, Varhaug JE, Vandekerckhove J, Lillehaug JR, Sherman F, Gevaert K. *Proc Natl Acad Sci U S A*. 2009; 106:8157. [PubMed: 19420222]
683. Kung LA, Tao SC, Qian J, Smith MG, Snyder M, Zhu H. *Mol Syst Biol*. 2009; 5:308. [PubMed: 19756047]
684. Lundberg E, Fagerberg L, Klevebring D, Matic I, Geiger T, Cox J, Algenas C, Lundberg J, Mann M, Uhlen M. *Mol Syst Biol*. 2010; 6:450. [PubMed: 21179022]
685. Lee MV, Topper SE, Hubler SL, Hose J, Wenger CD, Coon JJ, Gasch AP. *Mol Syst Biol*. 2011; 7:514. [PubMed: 21772262]
686. Munoz J, Low TY, Kok YJ, Chin A, Frese CK, Ding V, Choo A, Heck AJ. *Mol Syst Biol*. 2011; 7:550. [PubMed: 22108792]
687. van Noort V, Seebacher J, Bader S, Mohammed S, Vonkova I, Betts MJ, Kuhner S, Kumar R, Maier T, O'Flaherty M, Rybin V, Schmeisky A, Yus E, Stulke J, Serrano L, Russell RB, Heck AJ, Bork P, Gavin AC. *Mol Syst Biol*. 2012; 8:571. [PubMed: 22373819]
688. Jensen ON. *Nat Rev Mol Cell Biol*. 2006; 7:391. [PubMed: 16723975]
689. Witze ES, Old WM, Resing KA, Ahn NG. *Nat Methods*. 2007; 4:798. [PubMed: 17901869]

690. Blagoev B, Ong SE, Kratchmarova I, Mann M. *Nat Biotechnol.* 2004; 22:1139. [PubMed: 15314609]
691. Schmelzle K, Kane S, Gridley S, Lienhard GE, White FM. *Diabetes.* 2006; 55:2171. [PubMed: 16873679]
692. Kratchmarova I, Blagoev B, Haack-Sorensen M, Kassem M, Mann M. *Science.* 2005; 308:1472. [PubMed: 15933201]
693. Van Hoof D, Munoz J, Braam SR, Pinkse MW, Linding R, Heck AJ, Mummery CL, Krijgsveld J. *Cell Stem Cell.* 2009; 5:214. [PubMed: 19664995]
694. Brill LM, Xiong W, Lee KB, Ficarro SB, Crain A, Xu Y, Terskikh A, Snyder EY, Ding S. *Cell Stem Cell.* 2009; 5:204. [PubMed: 19664994]
695. Weintz G, Olsen JV, Fruhauf K, Niedzielska M, Amit I, Jantsch J, Mages J, Frech C, Dolken L, Mann M, Lang R. *Mol Syst Biol.* 2010; 6:371. [PubMed: 20531401]
696. Swaney DL, Wenger CD, Thomson JA, Coon JJ. *Proc Natl Acad Sci U S A.* 2009; 106:995. [PubMed: 19144917]
697. Gruhler A, Olsen JV, Mohammed S, Mortensen P, Faergeman NJ, Mann M, Jensen ON. *Mol Cell Proteomics.* 2005; 4:310. [PubMed: 15665377]
698. Bao MZ, Schwartz MA, Cantin GT, Yates JR 3rd, Madhani HD. *Cell.* 2004; 119:991. [PubMed: 15620357]
699. Jones MH, Keck JM, Wong CC, Xu T, Yates JR 3rd, Winey M. *Cell Cycle.* 2011; 10:3435. [PubMed: 22031224]
700. Villen J, Beausoleil SA, Gerber SA, Gygi SP. *Proc Natl Acad Sci U S A.* 2007; 104:1488. [PubMed: 17242355]
701. Godl K, Wissing J, Kurtenbach A, Habenberger P, Blencke S, Gutbrod H, Salassidis K, Steingerlach M, Missio A, Cotten M, Daub H. *Proc Natl Acad Sci U S A.* 2003; 100:15434. [PubMed: 14668439]
702. Bantscheff M, Eberhard D, Abraham Y, Bastuck S, Boesche M, Hobson S, Mathieson T, Perrin J, Raida M, Rau C, Reader V, Sweetman G, Bauer A, Bouwmeester T, Hopf C, Kruse U, Neubauer G, Ramsden N, Rick J, Kuster B, Drewes G. *Nat Biotechnol.* 2007; 25:1035. [PubMed: 17721511]
703. Daub H, Olsen JV, Bairlein M, Gnad F, Oppermann FS, Korner R, Greff Z, Keri G, Stemmann O, Mann M. *Mol Cell.* 2008; 31:438. [PubMed: 18691976]
704. Yu Y, Anjum R, Kubota K, Rush J, Villen J, Gygi SP. *Proc Natl Acad Sci U S A.* 2009; 106:11606. [PubMed: 19564600]
705. Kubota K, Anjum R, Yu Y, Kunz RC, Andersen JN, Kraus M, Keilhack H, Nagashima K, Krauss S, Paweletz C, Hendrickson RC, Feldman AS, Wu CL, Rush J, Villen J, Gygi SP. *Nat Biotechnol.* 2009; 27:933. [PubMed: 19801977]
706. Hart GW, Housley MP, Slawson C. *Nature.* 2007; 446:1017. [PubMed: 17460662]
707. Arnold CS, Johnson GV, Cole RN, Dong DL, Lee M, Hart GW. *J Biol Chem.* 1996; 271:28741. [PubMed: 8910513]
708. Kim HS, Kim EM, Lee J, Yang WH, Park TY, Kim YM, Cho JW. *FEBS Lett.* 2006; 580:2311. [PubMed: 16579988]
709. Wells L, Vosseller K, Cole RN, Cronshaw JM, Matunis MJ, Hart GW. *Mol Cell Proteomics.* 2002; 1:791. [PubMed: 12438562]
710. Wang Z, Park K, Comer F, Hsieh-Wilson LC, Saudek CD, Hart GW. *Diabetes.* 2009; 58:309. [PubMed: 18984734]
711. Wang Z, Udeshi ND, O'Malley M, Shabanowitz J, Hunt DF, Hart GW. *Mol Cell Proteomics.* 2010; 9:153. [PubMed: 19692427]
712. Vosseller K, Trinidad JC, Chalkley RJ, Specht CG, Thalhammer A, Lynn AJ, Snedecor JO, Guan S, Medzihradsky KF, Maltby DA, Schoepfer R, Burlingame AL. *Mol Cell Proteomics.* 2006; 5:923. [PubMed: 16452088]
713. Wang Z, Gucek M, Hart GW. *Proc Natl Acad Sci U S A.* 2008; 105:13793. [PubMed: 18779572]
714. Dias WB, Cheung WD, Wang Z, Hart GW. *J Biol Chem.* 2009; 284:21327. [PubMed: 19506079]

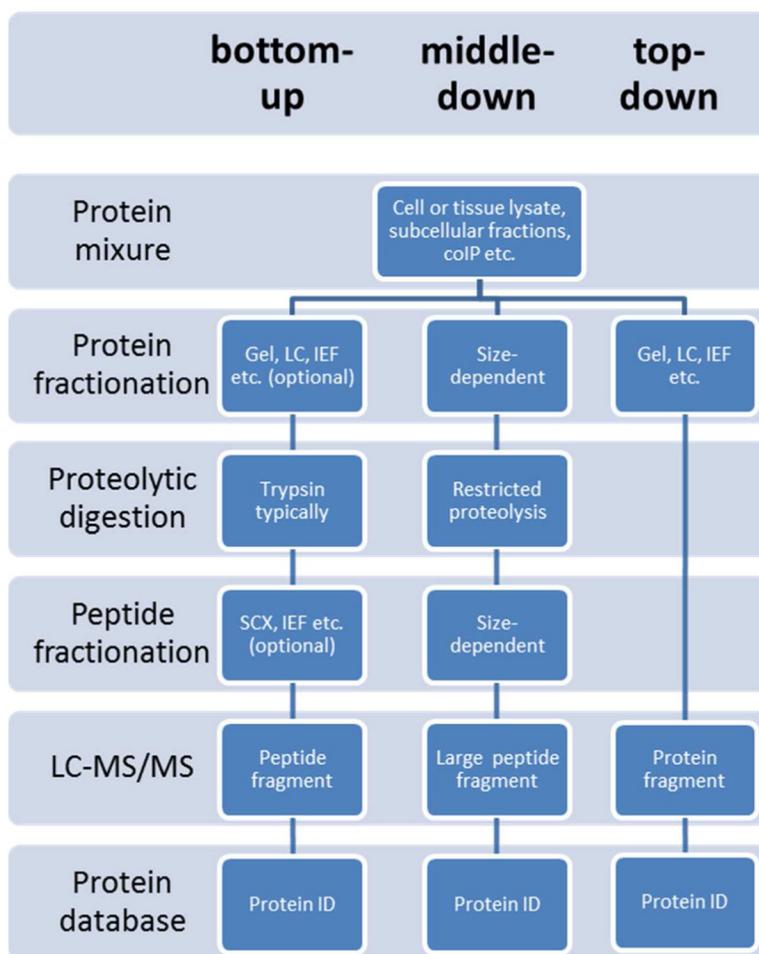
715. Kawauchi K, Araki K, Tobiume K, Tanaka N. *Proc Natl Acad Sci U S A*. 2009; 106:3431. [PubMed: 19202066]
716. Slawson C, Zachara NE, Vosseller K, Cheung WD, Lane MD, Hart GW. *J Biol Chem*. 2005; 280:32944. [PubMed: 16027160]
717. Wang Z, Udeshi ND, Slawson C, Compton PD, Sakabe K, Cheung WD, Shabanowitz J, Hunt DF, Hart GW. *Sci Signal*. 2010; 3:ra2. [PubMed: 20068230]
718. Huang F, Kirkpatrick D, Jiang X, Gygi S, Sorkin A. *Mol Cell*. 2006; 21:737. [PubMed: 16543144]
719. Kirkpatrick DS, Hathaway NA, Hanna J, Elsasser S, Rush J, Finley D, King RW, Gygi SP. *Nat Cell Biol*. 2006; 8:700. [PubMed: 16799550]
720. Strahl BD, Allis CD. *Nature*. 2000; 403:41. [PubMed: 10638745]
721. Chi P, Allis CD, Wang GG. *Nat Rev Cancer*. 2010; 10:457. [PubMed: 20574448]
722. Zhou Q, Chaerkady R, Shaw PG, Kensler TW, Pandey A, Davidson NE. *Proteomics*. 2010; 10:1029. [PubMed: 20049865]
723. Deribe YL, Wild P, Chandrashaker A, Curak J, Schmidt MH, Kalaidzidis Y, Milutinovic N, Kratchmarova I, Buerkle L, Fetchko MJ, Schmidt P, Kittanakom S, Brown KR, Jurisica I, Blagoev B, Zerial M, Stagljar I, Dikic I. *Sci Signal*. 2009; 2:ra84. [PubMed: 20029029]
724. Spickett CM, Pitt AR, Morrice N, Kolch W. *Biochim Biophys Acta*. 2006; 1764:1823. [PubMed: 17070740]
725. Jaffrey SR, Erdjument-Bromage H, Ferris CD, Tempst P, Snyder SH. *Nat Cell Biol*. 2001; 3:193. [PubMed: 11175752]
726. Mannick JB, Schonhoff CM. *Methods Enzymol*. 2008; 440:231. [PubMed: 18423221]
727. Lindermayr C, Saalbach G, Durner J. *Plant Physiol*. 2005; 137:921. [PubMed: 15734904]
728. Chen G, Gharib TG, Huang CC, Taylor JM, Misek DE, Kardina SL, Giordano TJ, Iannettoni MD, Orringer MB, Hanash SM, Beer DG. *Mol Cell Proteomics*. 2002; 1:304. [PubMed: 12096112]
729. Griffin TJ, Gygi SP, Ideker T, Rist B, Eng J, Hood L, Aebersold R. *Mol Cell Proteomics*. 2002; 1:323. [PubMed: 12096114]
730. Rao PK, Rodriguez GM, Smith I, Li Q. *Anal Chem*. 2008; 80:6860. [PubMed: 18690695]
731. Pupim LB, Flakoll PJ, Ikizler TA. *J Am Soc Nephrol*. 2004; 15:1920. [PubMed: 15213282]
732. Hinkson IV, Elias JE. *Trends Cell Biol*. 2011; 21:293. [PubMed: 21474317]
733. Belle A, Tanay A, Bitincka L, Shamir R, O'Shea EK. *Proc Natl Acad Sci U S A*. 2006; 103:13004. [PubMed: 16916930]
734. Eden E, Geva-Zatorsky N, Issaeva I, Cohen A, Dekel E, Danon T, Cohen L, Mayo A, Alon U. *Science*. 2011; 331:764. [PubMed: 21233346]
735. Yen HC, Xu Q, Chou DM, Zhao Z, Elledge SJ. *Science*. 2008; 322:918. [PubMed: 18988847]
736. Pratt JM, Petty J, Riba-Garcia I, Robertson DH, Gaskell SJ, Oliver SG, Beynon RJ. *Mol Cell Proteomics*. 2002; 1:579. [PubMed: 12376573]
737. Selbach M, Schwanhaussner B, Thierfelder N, Fang Z, Khanin R, Rajewsky N. *Nature*. 2008; 455:58. [PubMed: 18668040]
738. Boisvert FM, Ahmad Y, Gierlinski M, Charriere F, Lamont D, Scott M, Barton G, Lamond AI. *Mol Cell Proteomics*. 2012; 11:M111 011429.
739. Walrand S, Guillet C, Gachon P, Rousset P, Giraudet C, Vasson MP, Boirie Y. *Am J Physiol Cell Physiol*. 2004; 286:C1474. [PubMed: 14749219]
740. Caso G, Ford GC, Nair KS, Vosswinkel JA, Garlick PJ, McNurlan MA. *Am J Physiol Endocrinol Metab*. 2001; 280:E937. [PubMed: 11350775]
741. Bregendahl K, Liu L, Cant JP, Bayley HS, McBride BW, Milligan LP, Yen JT, Fan MZ. *J Nutr*. 2004; 134:2722. [PubMed: 15465773]
742. Rachdaoui N, Austin L, Kramer E, Previs MJ, Anderson VE, Kasumov T, Previs SF. *Mol Cell Proteomics*. 2009; 8:2653. [PubMed: 19724074]
743. Price JC, Holmes WE, Li KW, Floreani NA, Neese RA, Turner SM, Hellerstein MK. *Anal Biochem*. 2012; 420:73. [PubMed: 21964502]

744. Kim TY, Wang D, Kim AK, Lau E, Lin AJ, Liem DA, Zhang J, Zong NC, Lam MP, Ping P. *Mol Cell Proteomics*. 2012; 11:1586. [PubMed: 22915825]
745. Smith K, Reynolds N, Downie S, Patel A, Rennie MJ. *Am J Physiol*. 1998; 275:E73. [PubMed: 9688876]
746. Kipping M, Schierhorn A. *J Mass Spectrom*. 2003; 38:271. [PubMed: 12644988]
747. Mandell JG, Falick AM, Komives EA. *Anal Chem*. 1998; 70:3987. [PubMed: 9784743]
748. Doherty MK, Whitehead C, McCormack H, Gaskell SJ, Beynon RJ. *Proteomics*. 2005; 5:522. [PubMed: 15627957]
749. Zhang Y, Reckow S, Webhofer C, Boehme M, Gormanns P, Egge-Jacobsen WM, Turck CW. *Anal Chem*. 2011; 83:1665. [PubMed: 21314131]
750. Price JC, Guan S, Burlingame A, Prusiner SB, Ghaemmaghami S. *Proc Natl Acad Sci U S A*. 2010; 107:14508. [PubMed: 20699386]
751. Savas JN, Toyama BH, Xu T, Yates JR 3rd, Hetzer MW. *Science*. 2012; 335:942. [PubMed: 22300851]
752. Zhang DS, Piazza V, Perrin BJ, Rzadzinska AK, Poczatek JC, Wang M, Prosser HM, Ervasti JM, Corey DP, Lechene CP. *Nature*. 2012; 481:520. [PubMed: 22246323]
753. Atkinson AJ, Colburn WA, Degruittola VG. *Clin Pharmacol Ther*. 2001; 69:89. [PubMed: 11240971]
754. Zhang X, Li L, Wei D, Yap Y, Chen F. *Trends Biotechnol*. 2007; 25:166. [PubMed: 17316853]
755. Zhang H, Liu AY, Loriaux P, Wollscheid B, Zhou Y, Watts JD, Aebersold R. *Mol Cell Proteomics*. 2007; 6:64. [PubMed: 17030953]
756. Decramer S, Gonzalez de Peredo A, Breuil B, Mischak H, Monsarrat B, Bascands JL, Schanstra JP. *Mol Cell Proteomics*. 2008; 7:1850. [PubMed: 18667409]
757. Moon PG, You S, Lee JE, Hwang D, Baek MC. *Mass Spectrom Rev*. 2011; 30:1185. [PubMed: 21544848]
758. Moon PG, Lee JE, You S, Kim TK, Cho JH, Kim IS, Kwon TH, Kim CD, Park SH, Hwang D, Kim YL, Baek MC. *Proteomics*. 2011; 11:2459. [PubMed: 21595033]
759. Tumani H, Pfeifle M, Lehmsiek V, Rau D, Mogel H, Ludolph AC, Brettschneider J. *J Neuroimmunol*. 2009; 214:109. [PubMed: 19589605]
760. Wrensch MR, Petrakis NL, Gruenke LD, Ernster VL, Miike R, King EB, Hauck WW. *Breast Cancer Res Treat*. 1990; 15:39. [PubMed: 2183892]
761. Alexander H, Stegner AL, Wagner-Mann C, Du Bois GC, Alexander S, Sauter ER. *Clin Cancer Res*. 2004; 10:7500. [PubMed: 15569980]
762. Farina A, Dumonceau JM, Lescuyer P. *Expert Rev Proteomics*. 2009; 6:285. [PubMed: 19489700]
763. VanMeter AJ, Rodriguez AS, Bowman ED, Jen J, Harris CC, Deng J, Calvert VS, Silvestri A, Fredolini C, Chandhoke V, Petricoin EF 3rd, Liotta LA, Espina V. *Mol Cell Proteomics*. 2008; 7:1902. [PubMed: 18687633]
764. Hood BL, Darfler MM, Guiel TG, Furusato B, Lucas DA, Ringeisen BR, Sesterhenn IA, Conrads TP, Veenstra TD, Krizman DB. *Mol Cell Proteomics*. 2005; 4:1741. [PubMed: 16091476]
765. Simpson KL, Whetton AD, Dive C. *J Chromatogr B Analyt Technol Biomed Life Sci*. 2009; 877:1240.
766. Veenstra TD. *J Chromatogr B Analyt Technol Biomed Life Sci*. 2007; 847:3.
767. Liu T, Qian WJ, Mottaz HM, Gritsenko MA, Norbeck AD, Moore RJ, Purvine SO, Camp DG 2nd, Smith RD. *Mol Cell Proteomics*. 2006; 5:2167. [PubMed: 16854842]
768. Comegys MM, Lin SH, Rand D, Britt D, Flanagan D, Callanan H, Brilliant K, Hixson DC. *J Biol Chem*. 2004; 279:35063. [PubMed: 15184366]
769. Kui Wong N, Easton RL, Panico M, Sutton-Smith M, Morrison JC, Lattanzio FA, Morris HR, Clark GF, Dell A, Patankar MS. *J Biol Chem*. 2003; 278:28619. [PubMed: 12734200]
770. Ludwig JA, Weinstein JN. *Nat Rev Cancer*. 2005; 5:845. [PubMed: 16239904]
771. Tajiri M, Ohyama C, Wada Y. *Glycobiology*. 2008; 18:2. [PubMed: 17956937]

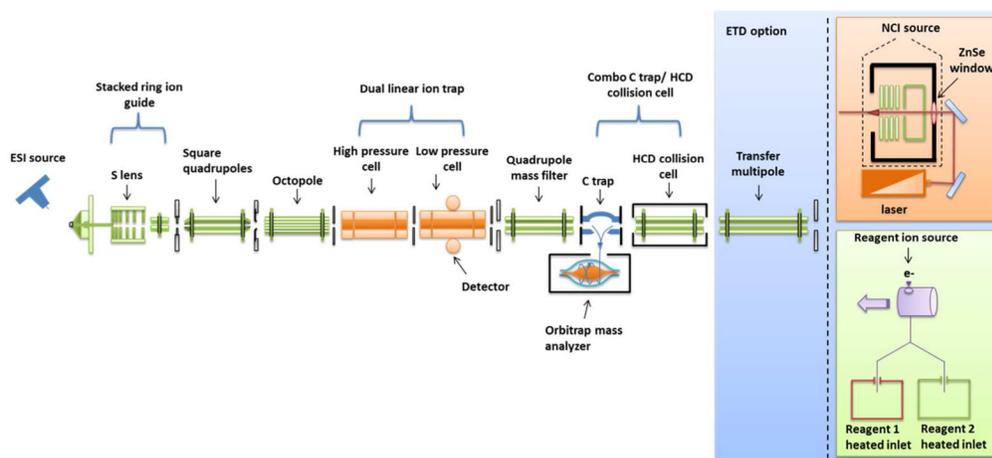
772. Brooks SA, Carter TM, Royle L, Harvey DJ, Fry SA, Kinch C, Dwek RA, Rudd PM. *Anticancer Agents Med Chem.* 2008; 8:2. [PubMed: 18220502]
773. Tian Y, Zhou Y, Elliott S, Aebersold R, Zhang H. *Nat Protoc.* 2007; 2:334. [PubMed: 17406594]
774. Palmisano G, Lendal SE, Engholm-Keller K, Leth-Larsen R, Parker BL, Larsen MR. *Nat Protoc.* 2010; 5:1974. [PubMed: 21127490]
775. Nwosu CC, Seipert RR, Strum JS, Hua SS, An HJ, Zivkovic AM, German BJ, Lebrilla CB. *J Proteome Res.* 2011; 10:2612. [PubMed: 21469647]
776. Leiserowitz GS, Lebrilla C, Miyamoto S, An HJ, Duong H, Kirmiz C, Li B, Liu H, Lam KS. *Int J Gynecol Cancer.* 2008; 18:470. [PubMed: 17655680]
777. de Leoz ML, An HJ, Kronewitter S, Kim J, Beecroft S, Vinnall R, Miyamoto S, de Vere White R, Lam KS, Lebrilla C. *Dis Markers.* 2008; 25:243. [PubMed: 19126968]
778. de Leoz ML, Young LJ, An HJ, Kronewitter SR, Kim J, Miyamoto S, Borowsky AD, Chew HK, Lebrilla CB. *Mol Cell Proteomics.* 2011; 10:M110 002717. [PubMed: 21097542]
779. Aldredge D, An HJ, Tang N, Waddell K, Lebrilla CB. *J Proteome Res.* 2012; 11:1958. [PubMed: 22320385]
780. Hua S, Lebrilla C, An HJ. *Bioanalysis.* 2011; 3:2573. [PubMed: 22122604]
781. Hua S, An HJ, Ozcan S, Ro GS, Soares S, DeVere-White R, Lebrilla CB. *Analyst.* 2011; 136:3663. [PubMed: 21776491]
782. Backstrom M, Thomsson KA, Karlsson H, Hansson GC. *J Proteome Res.* 2009; 8:538. [PubMed: 19072658]
783. Joshi HJ, Harrison MJ, Schulz BL, Cooper CA, Packer NH, Karlsson NG. *Proteomics.* 2004; 4:1650. [PubMed: 15174134]
784. Lohmann KK, von der Lieth CW. *Nucleic Acids Res.* 2004; 32:W261. [PubMed: 15215392]
785. Ceroni A, Maass K, Geyer H, Geyer R, Dell A, Haslam SM. *J Proteome Res.* 2008; 7:1650. [PubMed: 18311910]
786. Goldberg D, Bern M, Li B, Lebrilla CB. *J Proteome Res.* 2006; 5:1429. [PubMed: 16739994]
787. Zhang H, Yi EC, Li XJ, Mallick P, Kelly-Spratt KS, Masselon CD, Camp DG 2nd, Smith RD, Kemp CJ, Aebersold R. *Mol Cell Proteomics.* 2005; 4:144. [PubMed: 15608340]
788. Zolotarjova N, Mrozinski P, Chen H, Martosella J. *J Chromatogr A.* 2008; 1189:332. [PubMed: 18154976]
789. Pang J, Liu WP, Liu XP, Li LY, Fang YQ, Sun QP, Liu SJ, Li MT, Su ZL, Gao X. *J Proteome Res.* 2010; 9:216. [PubMed: 19894759]
790. Kim HJ, Kang HJ, Lee H, Lee ST, Yu MH, Kim H, Lee C. *J Proteome Res.* 2009; 8:1368. [PubMed: 19186948]
791. Ralhan R, Desouza LV, Matta A, Chandra Tripathi S, Ghanny S, Datta Gupta S, Bahadur S, Sui KW. *Mol Cell Proteomics.* 2008; 7:1162. [PubMed: 18339795]
792. Xue H, Lu B, Zhang J, Wu M, Huang Q, Wu Q, Sheng H, Wu D, Hu J, Lai M. *J Proteome Res.* 2010; 9:545. [PubMed: 19924834]
793. Bauer JW, Baechler EC, Petri M, Batliwalla FM, Crawford D, Ortmann WA, Espe KJ, Li W, Patel DD, Gregersen PK, Behrens TW. *PLoS Med.* 2006; 3:e491. [PubMed: 17177599]
794. Liao H, Wu J, Kuhn E, Chin W, Chang B, Jones MD, O'Neil S, Clauser KR, Karl J, Hasler F, Roubenoff R, Zolg W, Guild BC. *Arthritis Rheum.* 2004; 50:3792. [PubMed: 15593230]
795. Rithidech KN, Honikel L, Milazzo M, Madigan D, Troxell R, Krupp LB. *Mult Scler.* 2009; 15:455. [PubMed: 19324981]
796. Nanni P, Levander F, Roda G, Caponi A, James P, Roda A. *J Chromatogr B Analyt Technol Biomed Life Sci.* 2009; 877:3127.
797. Zimmermann-Ivol CG, Burkhard PR, Le Floch-Rohr J, Allard L, Hochstrasser DF, Sanchez JC. *Mol Cell Proteomics.* 2004; 3:66. [PubMed: 14581522]
798. Bell LN, Theodorakis JL, Vuppalanchi R, Saxena R, Bemis KG, Wang M, Chalasani N. *Hepatology.* 2010; 51:111. [PubMed: 19885878]
799. Kim HJ, Cho EH, Yoo JH, Kim PK, Shin JS, Kim MR, Kim CW. *J Proteome Res.* 2007; 6:735. [PubMed: 17269729]

800. Kolla V, Jenö P, Moes S, Tercanli S, Lapaire O, Choolani M, Hahn S. *J Biomed Biotechnol*. 2010; 2010:952047. [PubMed: 19902006]
801. Bons JA, Drent M, Bouwman FG, Mariman EC, van Dieijen-Visser MP, Wodzig WK. *Respir Med*. 2007; 101:1687. [PubMed: 17446058]
802. Weissinger EM, Schiffer E, Hertenstein B, Ferrara JL, Holler E, Stadler M, Kolb HJ, Zander A, Zurbig P, Kellmann M, Ganser A. *Blood*. 2007; 109:5511. [PubMed: 17339419]
803. Stanley BA, Gundry RL, Cotter RJ, Van Eyk JE. *Dis Markers*. 2004; 20:167. [PubMed: 15502250]
804. McDonough JL, Van Eyk JE. *Prog Cardiovasc Dis*. 2004; 47:207. [PubMed: 15736586]
805. Agranoff D, Fernandez-Reyes D, Papadopoulos MC, Rojas SA, Herbster M, Loosemore A, Tarelli E, Sheldon J, Schwenk A, Pollok R, Rayner CF, Krishna S. *Lancet*. 2006; 368:1012. [PubMed: 16980117]
806. Chen JH, Chang YW, Yao CW, Chiueh TS, Huang SC, Chien KY, Chen A, Chang FY, Wong CH, Chen YJ. *Proc Natl Acad Sci U S A*. 2004; 101:17039. [PubMed: 15572443]
807. He QY, Lau GK, Zhou Y, Yuen ST, Lin MC, Kung HF, Chiu JF. *Proteomics*. 2003; 3:666. [PubMed: 12748946]
808. Rifai N, Gillette MA, Carr SA. *Nat Biotechnol*. 2006; 24:971. [PubMed: 16900146]
809. Paulovich AG, Whiteaker JR, Hoofnagle AN, Wang P. *Proteomics Clin Appl*. 2008; 2:1386. [PubMed: 20976028]
810. Anderson NL. *Clin Chem*. 2010; 56:177. [PubMed: 19884488]
811. Selevsek N, Matondo M, Sanchez Carbayo M, Aebersold R, Domon B. *Proteomics*. 2011; 11:1135. [PubMed: 21360671]
812. Kitteringham NR, Jenkins RE, Lane CS, Elliott VL, Park BK. *J Chromatogr B Analyt Technol Biomed Life Sci*. 2009; 877:1229.
813. Latterich M, Abramovitz M, Leyland-Jones B. *Eur J Cancer*. 2008; 44:2737. [PubMed: 18977654]
814. Kim K, Kim SJ, Yu HG, Yu J, Park KS, Jang IJ, Kim Y. *J Proteome Res*. 2010; 9:689. [PubMed: 20020744]
815. Kuzyk MA, Smith D, Yang J, Cross TJ, Jackson AM, Hardie DB, Anderson NL, Borchers CH. *Mol Cell Proteomics*. 2009; 8:1860. [PubMed: 19411661]
816. Fortin T, Salvador A, Charrier JP, Lenz C, Bettsworth F, Lacoux X, Choquet-Kastylevsky G, Lemoine J. *Anal Chem*. 2009; 81:9343. [PubMed: 19839594]
817. Hossain M, Kaleta DT, Robinson EW, Liu T, Zhao R, Page JS, Kelly RT, Moore RJ, Tang K, Camp DG 2nd, Qian WJ, Smith RD. *Mol Cell Proteomics*. 2011; 10:M000062. [PubMed: 20410378]
818. Martin DB, Holzman T, May D, Peterson A, Eastham A, Eng J, McIntosh M. *Mol Cell Proteomics*. 2008; 7:2270. [PubMed: 18641041]
819. Anderson NL, Anderson NG, Haines LR, Hardie DB, Olafson RW, Pearson TW. *J Proteome Res*. 2004; 3:235. [PubMed: 15113099]
820. Anderson NL, Jackson A, Smith D, Hardie D, Borchers C, Pearson TW. *Mol Cell Proteomics*. 2009; 8:995. [PubMed: 19196707]
821. Liu R, Li Z, Bai S, Zhang H, Tang M, Lei Y, Chen L, Liang S, Zhao YL, Wei Y, Huang C. *Mol Cell Proteomics*. 2009; 8:70. [PubMed: 18723843]
822. Shi T, Fillmore TL, Sun X, Zhao R, Schepmoes AA, Hossain M, Xie F, Wu S, Kim JS, Jones N, Moore RJ, Pasa-Tolic L, Kagan J, Rodland KD, Liu T, Tang K, Camp DG 2nd, Smith RD, Qian WJ. *Proc Natl Acad Sci U S A*. 2012; 109:15395. [PubMed: 22949669]
823. Zhang Z, Chan DW. *Cancer Epidemiol Biomarkers Prev*. 2010; 19:2995. [PubMed: 20962299]
824. Savas JN, Stein BD, Wu CC, Yates JR 3rd. *Trends Biochem Sci*. 2011; 36:388. [PubMed: 21616670]
825. Blackler AR, Speers AE, Ladinsky MS, Wu CC. *J Proteome Res*. 2008; 7:3028. [PubMed: 18537282]
826. Hu Q, Noll RJ, Li H, Makarov A, Hardman M, Graham Cooks R. *J Mass Spectrom*. 2005; 40:430. [PubMed: 15838939]

827. Bohrer BC, Merenbloom SI, Koeniger SL, Hilderbrand AE, Clemmer DE. *Annu Rev Anal Chem* (Palo Alto Calif). 2008; 1:293. [PubMed: 20636082]
828. Cooper B. *J Proteome Res.* 2011; 10:1432. [PubMed: 21229988]
829. Geromanos SJ, Vissers JP, Silva JC, Dorschel CA, Li GZ, Gorenstein MV, Bateman RH, Langridge JI. *Proteomics.* 2009; 9:1683. [PubMed: 19294628]

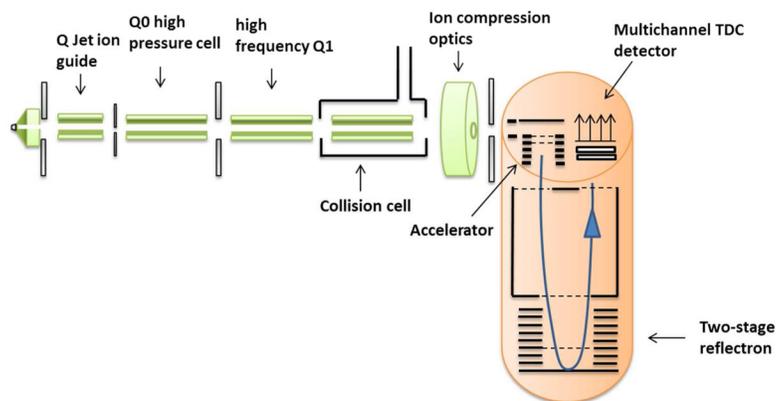
**Figure 1.**

Proteomic strategies: bottom-up vs. top-down vs. middle-down. The bottom-up approach analyzes proteolytic peptides. The top-down method measures the intact proteins. The middle-down strategy analyzes larger peptides resulted from limited digestion or more selective proteases. One or more protein or peptide fractionation techniques can be applied prior to MS analysis and database searching.

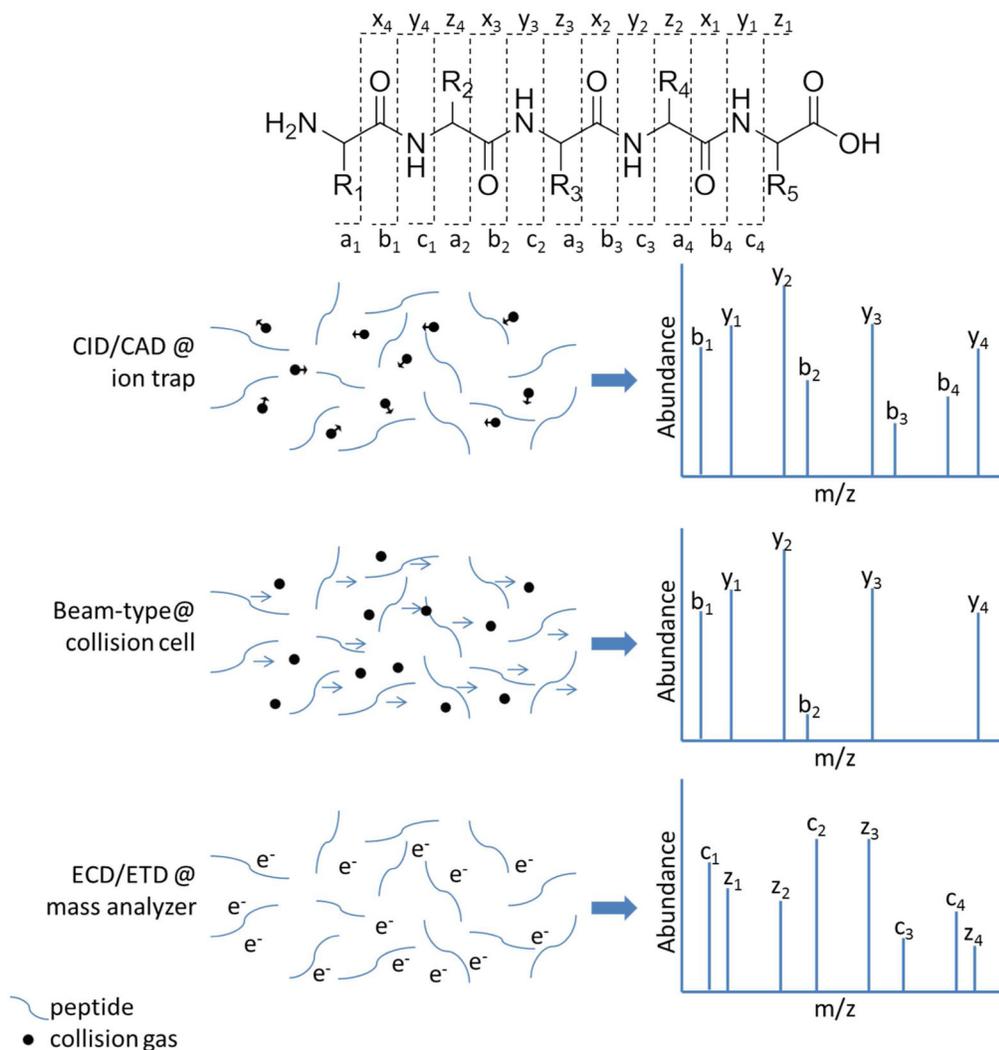


**Figure 2.**

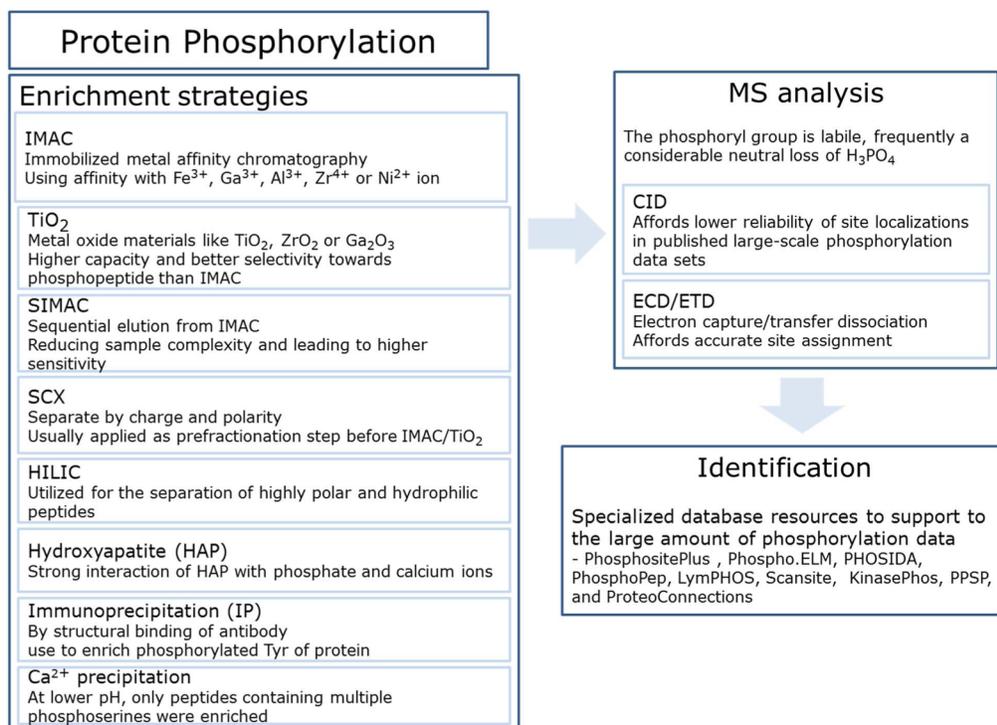
Diagrammatic summary of LIT- and Orbitrap-based mass spectrometers for proteomics. All but the IR fragmentation (orange box) is commercially available either as a single mass spectrometer or a combination of some essential components. Peptides are electrosprayed directly from a LC separation into the heated inlet source where desolvated peptide ions are focused by a stacked ring ion guide (S-lens), then focused, filtered (typically 300–2000  $m/z$ ), and transferred by the square quadrupole and octapole to the dual LIT. Peptide ions are collected, isolated, and fragmented by CID in the high pressure cell. Isolated precursor and fragment ions are passed to the low pressure trap for detection. This sequence of events is currently considered a state-of-the-art LIT experiment primarily for protein identification and label-free protein quantification. For higher resolution and mass accuracy detection, precursor or fragment ions can be passed to the Orbitrap mass analyzer via the second quadrupole and C-trap. Beam-type collision can be performed in the HCD collision cell instead of the ion trap for detection with Orbitrap. The Orbitrap detects ion currents of peptide ions that process around an orbital electrode. A Fourier transform is used to convert the frequency-based ion current to a  $m/z$  value. Either of these fragmentation schemes with Orbitrap detection are considered state-of-the-art for high mass accuracy peptide analysis and are most often used for protein and post-translational modification identification and quantification with isotopic labeling. Alternative fragmentation methods of ETD, IRMPD, or the combination IR-ETD can be used in conjunction with either the LIT or Orbitrap mass analyzers. The simplest configuration most recently demonstrated is a Q-Orbitrap with HCD detection which consists of only the S-lens, quadrupole mass filter, C-trap, Orbitrap, and HCD collision cell. This configuration and experiment design is quite similar to the QToF illustrated in Figure 3.



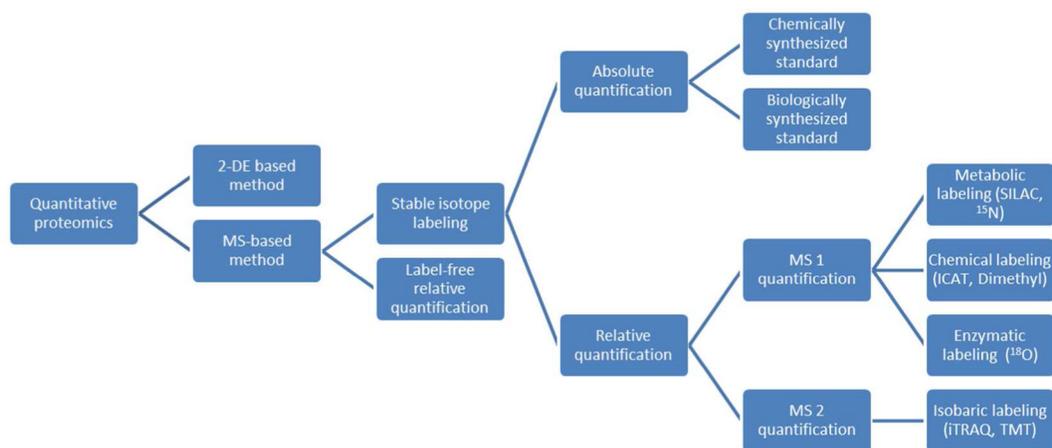
**Figure 3.** Schematic diagram of a QToF. The QToF allows for consecutive precursor mass filtering and fragmentation, an essential for shotgun proteomics which measures both intact and fragment ion masses of peptides. Essentially the instrument is a high mass accuracy QQQ frequently used in proteomics for targeted protein quantification experiments. Peptide ions from electrospray ionization are continuously focused by a Q jet ion guide, accelerated through quadrupole zero (Q0), Q1, and the collision cell for precursor peptide mass detection by time-of-flight and multichannel time-to-digital conversion (TDC) detector. Subsequently, abundant precursor ions can be selected by Q1 and fragmented in the collision cell prior to orthogonal acceleration for ToF detection.

**Figure 4.**

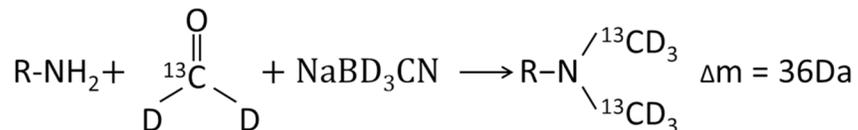
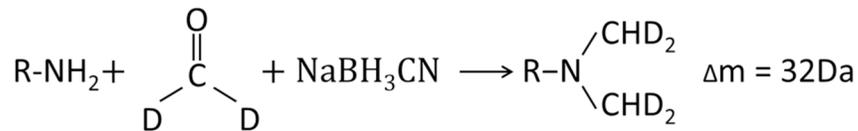
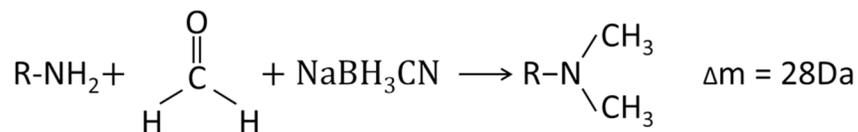
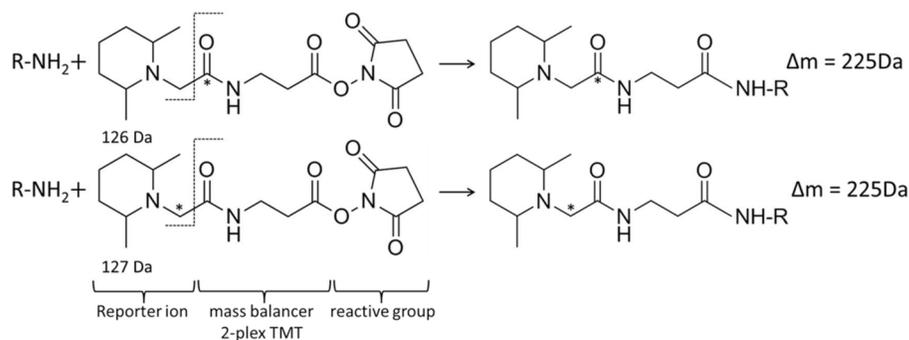
Common peptide fragmentation methods. Peptides are cleaved along the peptide backbone to sequence the peptide. CID/CAD entails acceleration of the kinetic energy of ions to promote energetic collisions with a target gas, thus causing conversion of the ion's kinetic energy to internal energy and ultimately resulting in ion fragmentation. Most fragmentation pathways rely on proton transfer. For trapping instruments that employ resonant excitation, the waveform used to accelerate a precursor ion is specific for a particular  $m/z$  value, thus only the selected precursor ion is activated. The resulting fragment ions are not excited and thus do not dissociate further, allowing detection of an array of predominantly both b and y ions (as well as others). In contrast to resonant excitation in trapping instruments, in beam-type CID both the selected precursor ions and any resulting fragment ions are passed through a collision region which enables further activation and dissociation of the fragment ions. This means that the less stable b ions will frequently decompose to very small ions. Thus, both peptide precursors and their fragment ions are activated and mostly y-type fragment ions persist. Unlike CID, ECD and ETD fragmentation relies on the gas phase reaction of the peptide ions with a thermal electron or an ETD reagent, respectively. The electron transfer-driven fragmentation mechanisms create mostly c- and z-type ions.



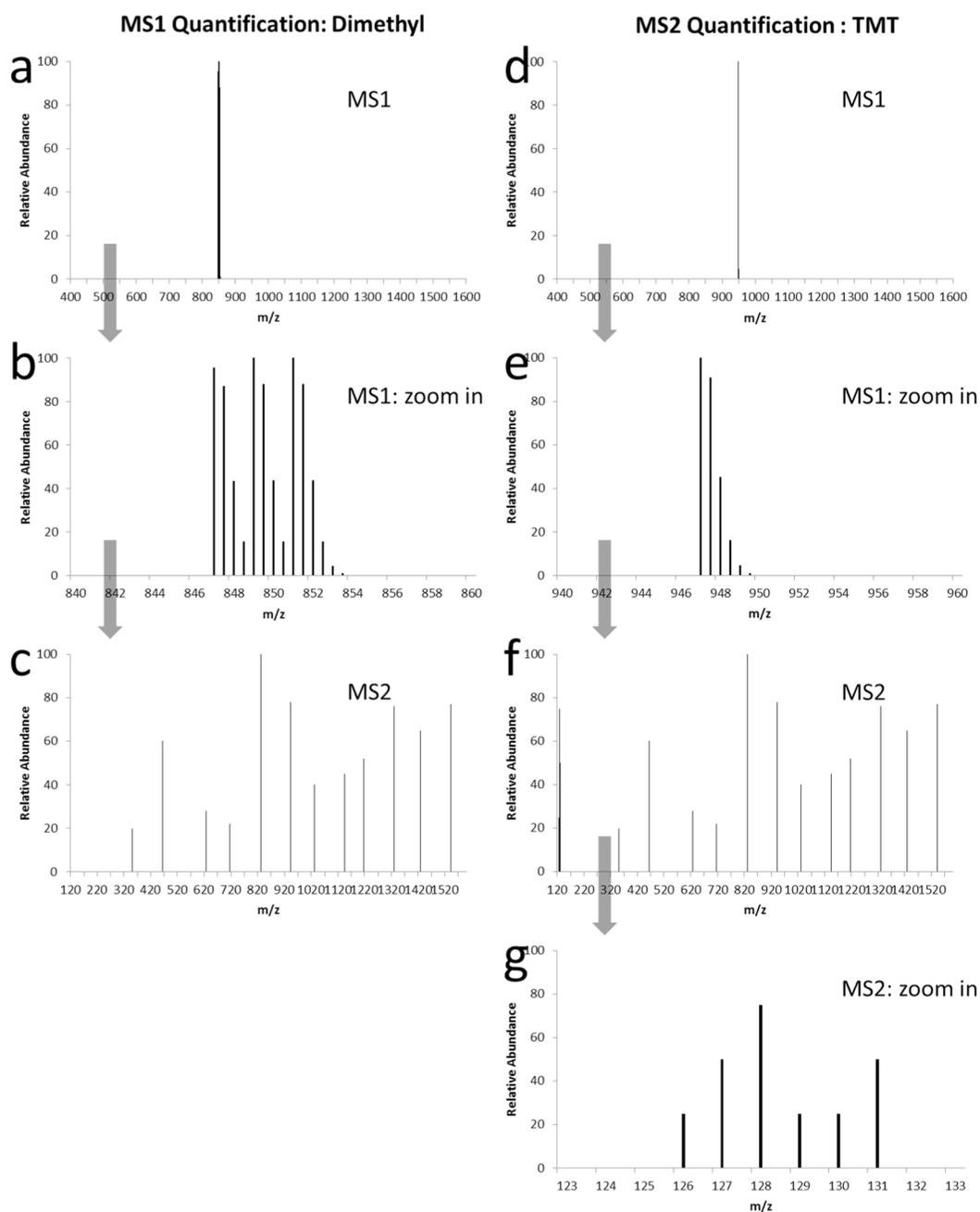
**Figure 5.**  
Overview of phosphoproteomic methods



**Figure 6.** Categorization of the main strategies for quantitative proteomics.

**a: Dimethyl****b: 2-plex TMT****Figure 7.**

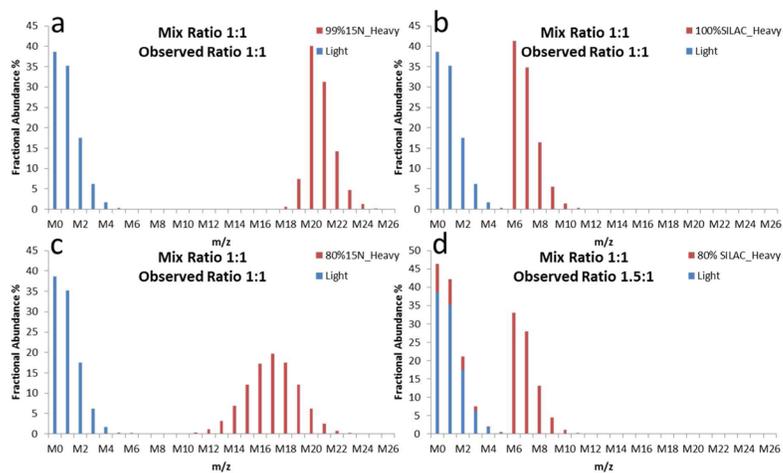
Labeling reactions of triplex stable isotope dimethyl labels and duplex TMT labels. Both labeling methods target the N-terminus and Lys of peptides with a reactive isotopic group. a: Dimethyl labeling introduces different mass increases via different isotopic reagents. b: TMT labeling introduces equal mass increases via different reagents. The reporter ions with different masses are released during peptide fragmentation. Asterisk indicates a  $^{13}\text{C}$  atom replacement.



**Figure 8.**

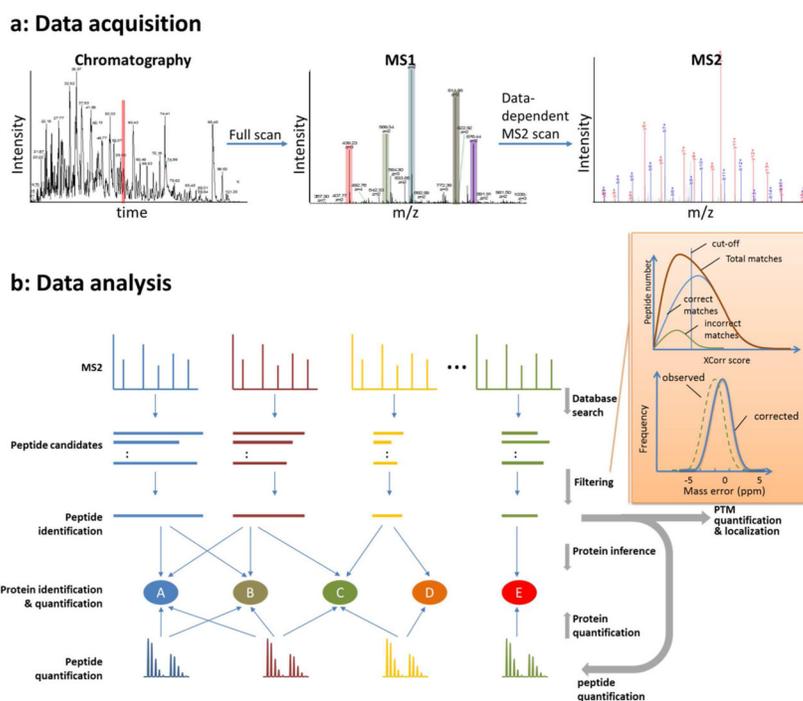
Two quantitative strategies based on chemical reaction. The representative peptide is doubly charged, and consisting of 15 molecules of averagine, a hypothetical amino acid. Averagine has a molecular formula of  $C_{4.9384} H_{7.7583} N_{1.3577} O_{1.4773} S_{0.0417}$ , which is derived from statistical occurrence of amino acids in the protein identification resource protein database. a, b, c: Dimethyl labeling. The quantification is calculated based on the ion intensities in MS1. The MS2 is only used to identify the peptide sequence (c). This given peptide has slight isotopic peak overlap between the three groups (b), the overlapped areas will become more significant for longer peptides. d, e, f, g: 6-plex TMT labeling. The quantification is calculated based on the report ion intensities in MS2 (g). The MS2 is used for both peptide

sequence identification (f) and quantification (g). The mixed peptide groups have the same  $m/z$  in MS1 (e).

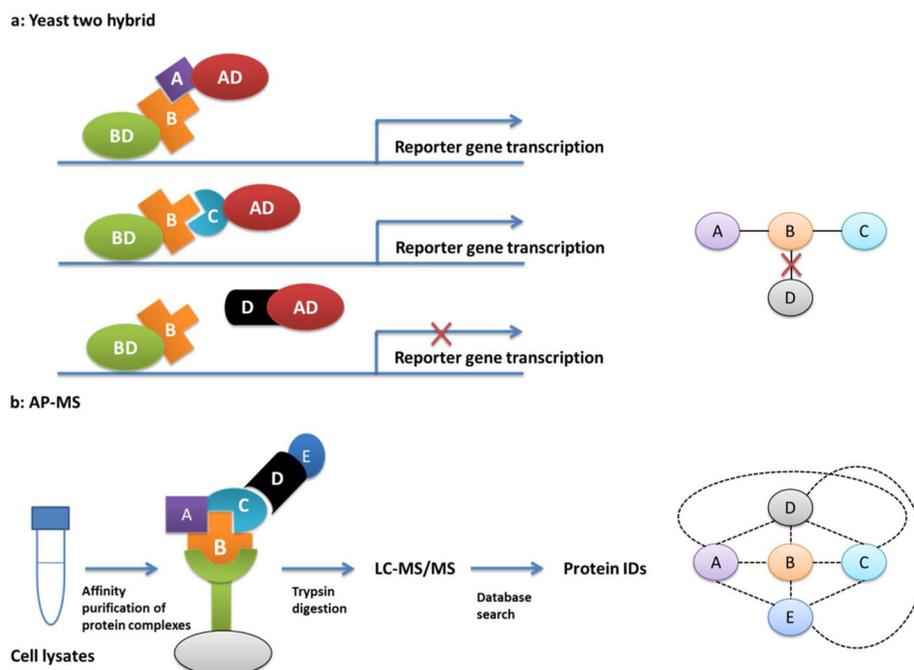


**Figure 9.**

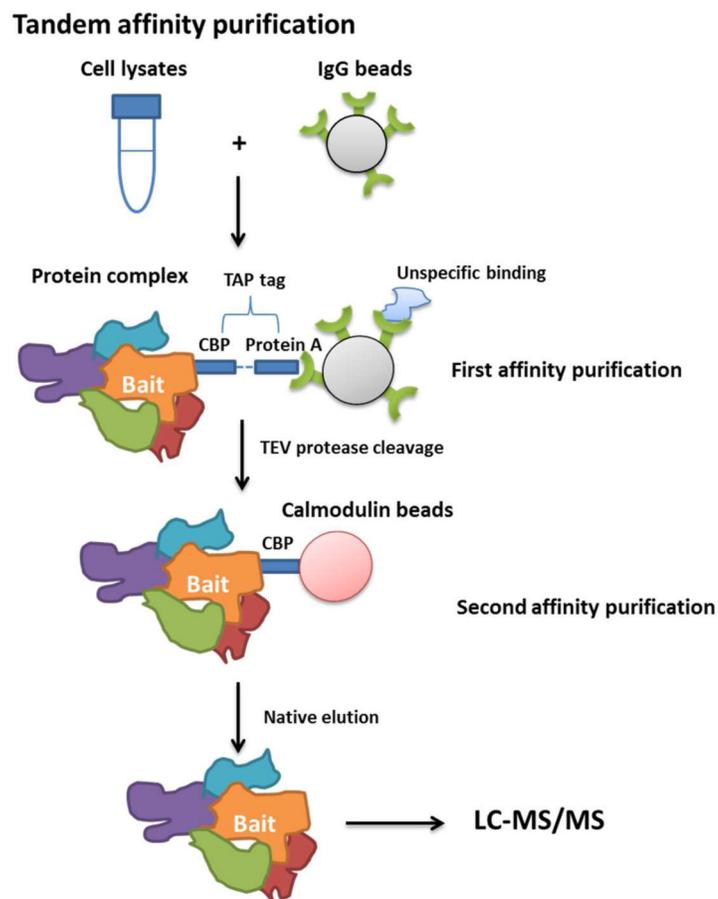
Comparison of  $^{15}\text{N}$  and SILAC quantifications. The peptide presented here consists of 15 molecules of hypothetical amino acid, averagine. Averagine has a molecular formula of  $\text{C}_{4.9384}\text{H}_{7.7583}\text{N}_{1.3577}\text{O}_{1.4773}\text{S}_{0.0417}$ , which is derived from statistical occurrence of amino acids in the protein identification resource protein database. Each spectrum represents a 1:1 mixture of unlabeled and labeled peptides. (a), (b): The mass spectra observed when the labeling efficiency is or nearly complete for  $^{15}\text{N}$  and SILAC quantifications.  $^{15}\text{N}$  labeling typically generates bigger mass increase compared to SILAC labeling. (c), (d): The mass spectra observed when the labeling efficiency is 80% for  $^{15}\text{N}$  and SILAC quantifications.  $^{15}\text{N}$  is still able to result in an accurate quantitative ratio, whereby the SILAC labeling leads to a false ratio. SILAC quantitative labeling is not compatible with partial labeling because the observed unlabeled peaks do not exclusively originate from unlabeled sample.

**Figure 10.**

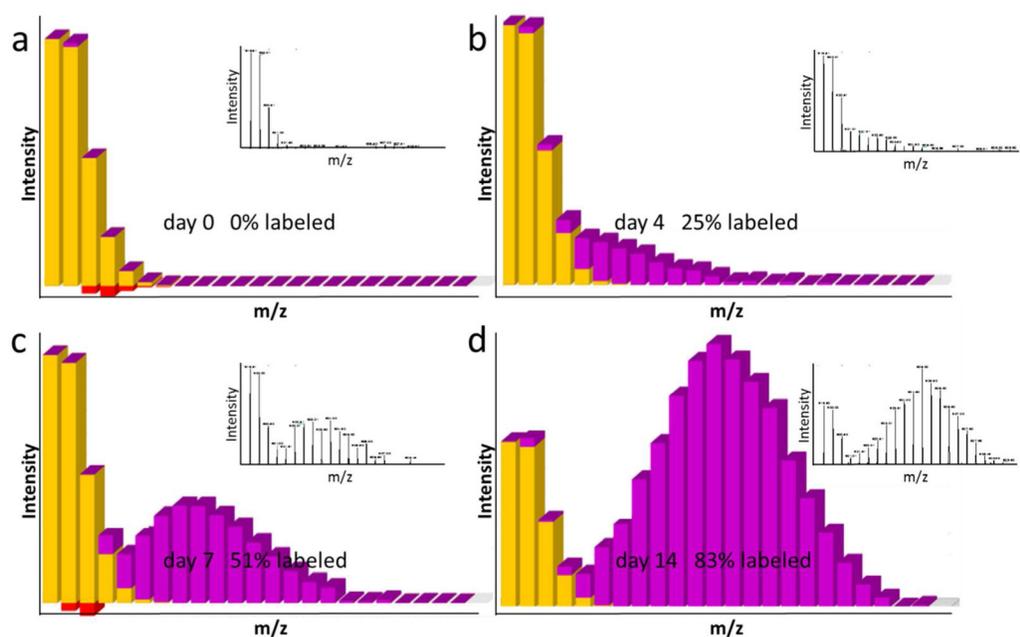
Representative LC-MS/MS data and a generalized bioinformatic analysis pipeline for protein identification and quantification in shotgun proteomics. (a) As a total ion chromatogram is acquired by nESI-MS from the nLC separation of peptides, the mass spectrometer acquires both full scan MS (MS1) precursor spectra and data-dependent MS/MS (MS2) fragmentation spectra. All ions, typically between 300 – 2000 m/z, are detected during nLC in the full MS scans. The full scan defines the most abundant peptide precursor ions which are sampled by data-dependent for MS/MS. (b) The acquired data is then processed through a bioinformatics pipeline. A database search is used to match theoretically generated peptide fragmentation spectra to experimental MS2 fragmentation spectra, creating a list of peptide candidates for each experimental spectrum. The peptide candidates are ranked and filtered to create peptide spectrum matches (PSMs) and peptide identifications. PSMs can be filtered by XCorr using the “incorrect” reverse PSMs as an estimation of false discovery rate (FDR). In high mass accuracy experiments, prior to XCorr filtering a mass error window can be used to prefilter PSMs based on the precursor mass measurement from MS1 scans. If a systematic mass error is “observed”, it can be “corrected” by adding or subtracting the average mass error. Identified peptides are assigned to proteins by inference to create a list of proteins present within the sample. The relative protein quantification is then performed by averaging the peptide ratio measurements for peptides assigned to the protein. These same strategies can be used for post-translational modification (PTM) identification and quantification, by using a peptide as a surrogate measurement of the PTM.



**Figure 11.** The protein-protein interaction detected by Y2H and AP-MS. a: Y2H is used to detect binary interaction. The results here demonstrate the protein B can interact with protein A and C, but not D. b: AP-MS is used to identify the whole protein complex. All the interactors binding to protein B, including both direct and indirect binders, are identified by shotgun proteomics.

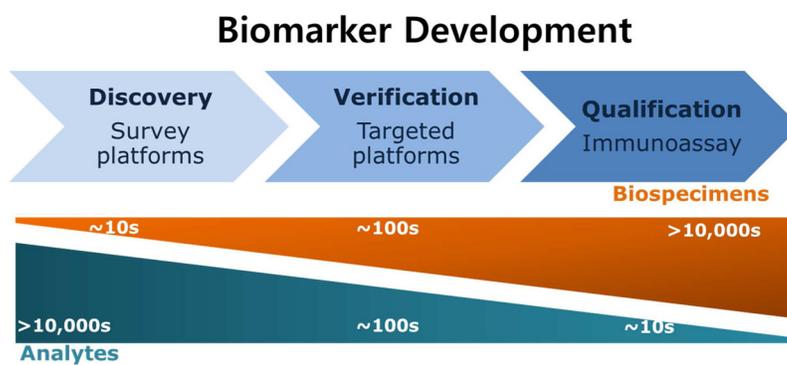


**Figure 12.** Scheme of tandem affinity purification (TAP). Two steps of purification significantly remove the unspecific binding proteins.



**Figure 13.**

The protein turnover measured by  $^{15}\text{N}$  partial labeling with the software *ProTurnyzer*.<sup>749</sup> The representative peptide has an amino acid sequence of LDKSQIHDI<sup>749</sup>LVGGSTR derived from mouse heat shock cognate 71 kDa protein. The yellow peaks represent the pre-existing peptide with natural isotopic composition, whereby the purple peaks represent peptide synthesized during  $^{15}\text{N}$  feeding. The gain of purple and removal of yellow reflect protein synthesis and protein degradation, respectively. The red bars indicate the intensity error between theoretical and measured peptide isotope distributions. The inset figures are the observed mass spectra for which the unlabeled and labeled peptide peaks were analyzed.



**Figure 14.**  
Steps of biomarker development

**Table 1**

Common proteases used for shotgun proteomics.

Protease	Cleavage Specificity <sup>a</sup>	Common proteomic usage
Trypsin	-K,R-↑-Z- not -K,R-↑-P-	General protein digestion
Endoproteinase Lys-C	-K-↑-Z-	Alternative to trypsin for increased peptide length; multiple protease digestion; <sup>18</sup> O labeling
Chymotrypsin	-W,F,Y-↑-Z- and -L,M,A,D,E-↑-Z- at a slower rate	Multiple protease digestion
Subtilisin	Broad specificity to native and denatured proteins	Multiple protease digestion
Elastase	-B-↑-Z-	Multiple protease digestion
Endoproteinase Lys-N	-Z-↑-K-	Increase peptide length; create higher charge state for ETD
Endoproteinase Glu-C	-E-↑-Z- and 3000 times slower at -D-↑-Z-	Multiple protease digestion; <sup>18</sup> O labeling
Endoproteinase Arg-C	-R-↑-Z-	Multiple protease digestion
Endoproteinase Asp-N	-Z-↑-D- and -Z-↑-cysteic acid- but not -Z-↑-C-	Multiple protease digestion
Proteinase K	-X-↑-Y-	Non-specific digestion of membrane-bound proteins
OmpT	-K,R-↑-K,R-	Increased peptide length for middle-down proteomics

<sup>a</sup>B – uncharged, non-aromatic amino acids (i.e. A, V, L, I, G, S); X – aliphatic, aromatic, or hydrophobic amino acids; and Z – any amino acid.

**Table 2**

Mass analyzers used in shotgun proteomics and their commonly-achieved analytical metrics for peptide analysis.

<b>Analyzer</b>	<b>Instruments</b>	<b>Type</b>	<b>Resolution</b>	<b>Mass accuracy</b>	<b>Dynamic Range</b>
<b>Quadrupole</b>	QQ QToF	Beam	1 – 2 K	~ 1 % <sup>a</sup>	5 – 6
<b>Ion trap</b>	LIT	Trapping (Electric field)	1 – 2 K	~ 1 % <sup>a</sup>	3 – 4
<b>ToF</b>	QToF	Beam	10 – 50 K	5 – 10 ppm	4
<b>Orbitrap</b>	FT	Trapping (Electric field)	7.5 – 240 K	500 ppb – 10 ppm	4
<b>ICR</b>	FT	Trapping (Magnetic and axial DC fields)	100 – 500 K	~100 ppb	3

<sup>a</sup> parts-per-thousand.

**Table 3**

Common post-translational modifications detected by mass spectrometry.

PTM	Residues	Chemical group	$\Delta$ mass (Da) <sup>a</sup>
<b>Phosphorylation</b>	Ser, Thr, Tyr	HPO <sub>3</sub>	79.9663
<b>N-Glycosylation</b>	Asn	Glycan	132.0432 <sup>b</sup> (-0.9840 and 2.9890) <sup>c</sup>
<b>O-Glycosylation</b>	Ser, Thr	Glycan	132.0432 <sup>bc</sup>
<b>Oxidation</b>	Met	O	15.9949
<b>Methylation</b>	N- C- terminus, Lys, Ser, Thr, Asn, Gln, (Iso)Asp <sup>d</sup>	CH <sub>2</sub>	14.0156
<b>Dimethylation</b>	Arg, Lys	CH <sub>2</sub> CH <sub>2</sub>	28.0313
<b>Trimethylation</b>	Arg, Lys	CH <sub>2</sub> CH <sub>2</sub> CH <sub>2</sub>	42.0470
<b>S-Nitrosylation</b>	Cys	NO	28.9902
<b>Citrullination</b>	Arg	O	0.9840
<b>Ubiquitination</b>	Lys	Ubiquitin	8564.8448 (114.0429) <sup>e</sup>
<b>Acetylation</b>	N terminus, Lys, Ser	CH <sub>3</sub> CO	42.0106
<b>Carbamylation</b>	N-terminus, Lys, Arg	CONH <sub>2</sub>	43.0058
<b>Biotinylation (amide bond to)</b>	N-terminus, Lys	Biotin	226.0776

<sup>a</sup>  $\Delta$  mass (Da) is the change in mass of the peptide and amino acid in Daltons due to the addition of a PTM.

<sup>b</sup> Mono- and polyglycosylation result in variable  $\Delta$  masses. The lowest mass increase from monoglycosylation by pentose is 132.0432. The remaining monoglycosylation masses are 146.0579 (deoxyhexose), 162.0528 (hexose), 203.0794 (N-acetylhexosamine), 291.0954 (N-acetylneuraminic acid), and 307.0903 (N-glycolylneuraminic acid). Polyglycosylation creates combinatorial addition of these masses.

<sup>c</sup> The presence and localization of an N-linked glycan can be determined by treatment with PNGase and observation of a 0.9840 Da mass loss. The use of heavy H<sub>2</sub><sup>18</sup>O during the PNGase treatment produces a more easily detectable mass increase of 2.9890 that can be differentiated from the mass loss associated with Asn deamidation (-0.9840 Da).

<sup>d</sup> IsoAsp is racemized aspartic acid.

<sup>e</sup> Ubiquitinylation or polyubiquitinylation of a single lysine residue results in  $\Delta$  masses equal to or greater than 8564.8448 Da, respectively. From tryptic digestion, the majority of the ubiquitin molecule is cleaved from the lysine residue, leaving a signature Gly-Gly dipeptide on the lysine side chain that has a detectable mass shift of 114.0429 Da.

**Table 4**

## Isotope labeling quantitative strategies

Methods	Labeling residues	Multiplexing	MS1/MS2	Mass shift or reporter ionmasses
<b>ICAT</b>	Cys	2	MS1	+8 Da in heavy
<b>ICPL</b>	Lys	2	MS1	+6 Da in heavy
<b>Dimethyl</b>	N-terminus, Lys	3	MS1	+4 Da in medium +8 Da in heavy
<b><sup>18</sup>O</b>	C-terminus	2	MS1	+2,+4 Da in heavy
<b>SILAC</b>	Lys, Arg	3	MS1	+4,+6,+8,+10 Da in heavy
<b>SILAM</b>	All AAs	2	MS1	Typically >10 Da <sup>a</sup>
<b>iTRAQ</b>	N-terminus, Lys	8	MS2	113–119, 121 Da for 8-plex 126–127 Da for 2-plex
<b>TMT</b>	N-terminus, Lys	6	MS2	126–131 Da for 6-plex

<sup>a</sup>The mass shift is determined by the number of <sup>15</sup>N and/or <sup>13</sup>C atoms in the amino acid and peptide.