Published in final edited form as:

Wiley Interdiscip Rev Syst Biol Med. 2012 March ; 4(2): 141–162. doi:10.1002/wsbm.166.

Mass Spectrometry-based Proteomics: Qualitative Identification to Activity-based Protein Profiling

Job D. Cardoza^{1,3}, Jignesh R. Parikh^{1,4}, Scott B. Ficarro^{1,2,3}, and Jarrod A. Marto^{1,2,3,§}

¹Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA 02115

²Blais Proteomics Center, Dana-Farber Cancer Institute, Boston, MA 02115

³Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115

⁴Bioinformatics Program, Boston University, Boston, MA 02115

Abstract

Mass spectrometry has become the method of choice for proteome characterization, including multi-component protein complexes (typically tens to hundreds of proteins) and total protein expression (up to tens of thousands of proteins), in biological samples. Qualitative sequence assignment based on MS/MS spectra is relatively well-defined, while statistical metrics for relative quantification have not completely stabilized. Nonetheless, proteomics studies have progressed to the point whereby various gene-, pathway-, or network-oriented computational frameworks may be used to place mass spectrometry data into biological context. Despite this progress, the dynamic range of protein expression remains a significant hurdle, and impedes comprehensive proteome analysis. Methods designed to enrich specific protein classes have emerged as an effective means to characterize enzymes or other catalytically active proteins that are otherwise difficult to detect in typical discovery mode proteomics experiments. Collectively, these approaches will facilitate identification of biomarkers and pathways relevant to diagnosis and treatment of human disease.

1.1: Introduction

Mass spectrometry has transitioned from an esoteric, low-throughput analytical method typically utilized only by highly specialized labs to the technique of choice for systematic identification and quantification of proteins studied in the context of numerous biological systems. Progress in large-scale protein sequencing has been driven largely by technological developments in mass spectrometry, protein/peptide separations, biochemical enrichment methods and data analysis algorithms. One long-term goal of efforts directed at comprehensive proteome sequencing is the development of models based on primary proteomics measurements that support prediction of cellular activity, clinical outcomes, or other biological responses “personalized” to individuals or clinical cohorts [1].

We will focus here on bottom-up analysis of proteins, which deals with proteolytic fragments of proteins (or ‘peptides’) of approximately 3,000 Da or less, and is the most widely practiced technique in proteomics today. Several mass spectrometers are well-suited to bottom-up analysis of proteins. The quadrupole ion trap, which is capable of mass-

[§]To whom correspondence should be addressed: Department of Cancer Biology, Dana-Farber Cancer Institute, 450 Brookline Avenue, Smith Building, 1158A, Boston, MA, 02215-5450, USA. Phone: (617) 632-3150, Fax: (617) 582-7737, jarrod_marto@dfci.harvard.edu.

selected ion storage, can function as a stand-alone mass spectrometer for MS and MS/MS measurements [6, 7] or be coupled to a high resolution mass analyzer in a hybrid geometry [8–10]. Time-of-flight mass analyzers have been coupled to matrix assisted laser desorption [11–16] or electrospray [17–20] ionization sources for MS and MS/MS analyses of proteins and peptides.

1.2: Mass Concepts

The fundamental observable in a mass spectrometer is analyte mass-to-charge ratio (denoted m/z). In protein digests of whole cell lysates, hundreds of thousands of distinct peptides may be generated, many of which will have relative m/z differences of 0.1 Da or less, owing to the fact that highly variable peptide sequences can be constructed using the 20 common amino acids which are either isomeric (i.e., same molecular formulae) or nominally isobaric (i.e., same integral mass, but different molecular formulae) to other peptides. Drawing from combinatorial theory, He *et al.* [21] showed that of all possible di- and tri- peptides, assuming only the 20 naturally occurring amino acids, only 52% and 19%, respectively, were compositionally distinct; moreover, 29% and 53% of each sub-class were isomers. The consequence of this pseudo-degeneracy is that mass alone may not be sufficient for peptide identification, but can be used to limit the number of candidate peptides during sequence assignment of MS/MS spectra, illustrating the value of high resolution mass spectrometry platforms.

A number of concepts related to resolution have evolved that are useful to introduce here. Monoisotopic mass is calculated using the most abundant isotope for each atom in an ion. Average mass is calculated using the abundance-weighted average mass of each atom. Both definitions are based on the unified atomic mass unit, or 'u', which defines 1 u as 1/12 the mass of carbon-12, and is equivalent to 1 Da [22]. Resolution (R_z) is defined by IUPAC [23] as:

$$R_z = \frac{m}{\Delta m_z},$$

where m is the m/z ratio at a peak's maximum and Δm_z is the peak width at Z percent of the maximum intensity. Z is commonly assumed to be 50%, and this value will be used herein. In most time-of-flight instruments, R can approach 5,000 without a reflectron [24, 25] and 15,000 with a reflectron [24, 26], although the latest generation time-of-flight mass spectrometers can achieve $R = 50,000$ [27] for small molecules and $R = 40,000$ for peptides [28]. Radio frequency ion trap mass spectrometers have about 1.0 Da resolution for ions up to 2000 m/z in a typical implementation [6] (or $R \leq 2,000$), though $R = 20,000$ is attainable with slower scan rates [29]. Fourier transform instruments can achieve significantly higher values, such as the Orbitrap ($R = 150,000$ [2]) and FTICR ($R = 3,300,000$ [30]) mass spectrometers.

Mass accuracy is important for mass spectrometry-based proteomics because accurate measurement of a peptide's mass greatly reduces the putative number of sequence matches in the associated database of proteins. Zubarev *et al.* [31] showed that a mass accuracy of 1 ppm can eliminate 99% of nominally isobaric peptides. Mass accuracy is commonly reported in two ways: parts per million or Da. Parts per million (ppm) specifies error of the measured mass relative to a peptide's theoretical mass. For example, consider a peptide with a theoretical m/z of 1570.58 Da and a measured m/z of 1570.56 Da. The mass accuracy is calculated as:

$$\left| \frac{(1570.58 - 1570.56\text{Da})}{1570.58\text{Da}} \right| \cdot 1\text{e}6\text{ppm} = 12.7\text{ppm}.$$

Alternatively, the mass accuracy can be specified relative to the theoretical mass in absolute units, commonly Daltons. For an ion trap mass spectrometer, the accuracy is typically 0.6 Da for MS precursor peaks or MS/MS fragment peaks. For a time-of-flight instrument, equipped with a reflectron ion mirror, the mass accuracy is 0.1–0.25 Da for MS or MS/MS. Fourier transform-based instruments typically have mass accuracies at or below 10 ppm for both MS and MS/MS.

1.3.1: De Novo Sequencing of Low-Energy CAD Mass Spectra

Peptide sequence information can be derived from a variety of MS/MS data, such as that from low energy collisionally activated dissociation (CAD) in an ion trap [32] or collision cell [18, 33, 34] configuration, electron transfer dissociation (ETD) [35, 36] and electron capture dissociation (ECD) [37, 38]. In CAD, peptide ions are fragmented by low-energy (≤ 200 eV) collisions with a neutral seed gas to form b- and y-type ions (Fig. 2a). In ETD and ECD, peptide ions undergo non-ergodic fragmentation upon incorporation of a thermalized electron, either directly (ECD) or via transfer from an electron-donor anion (ETD), to form c/z ions [38, 39] (Fig. 2a). CAD will be the focus of the following sections.

Typical rates of data generation of state-of-the-art mass spectrometers require computational database search tools such as Mascot [40], SEQUEST [41], Phenyx [42, 43], and X!Tandem [44] that provide for high-throughput, automated spectral sequence assignment. The details of the underlying search algorithms are described in several recent reviews [45–47]. In most cases a theoretical ‘peptidome’ resulting from enzymatic digestion (*e.g.*, using trypsin) is readily generated *in silico* based on the known genome sequence of the organism from which the sample was derived. Computational algorithms match peaks in each MS/MS spectrum with fragment ions derived from the mass-filtered set of theoretical peptides in accordance with the accuracy of the mass spectrometer. Sequences are scored based on the numbers of matched peaks, often with a penalty for non-matching signals. In this way, thousands of spectra can be rapidly sequenced and mapped back to their source proteins.

Although database-driven sequencing has become an industry standard in proteomics [40, 48–50], and has identified tens of thousands of peptides for a broad range of biomedical applications [51], large fractions of MS/MS spectra remain unidentified [52]. This is partly driven by experimental factors [49], but computational limitations also play a role. Non-quantitative amino acid modifications (*i.e.*, chemical or post-translational modifications which occur on specific side-chains, but often with low stoichiometry), for example, require searching both modified and unmodified forms of all amino acids targeted by the specific modification in a protein database. This effectively increases the search space and the chance of random matches; hence in practice one must compromise between the number of simultaneous post-translational modifications that adequately capture the nature of the sample being analyzed and an acceptable rate of false positive identifications [49].

In light of such considerations, we provide a brief tutorial on manual sequencing of MS/MS spectra. This approach is labor intensive and time-consuming, but is not restricted by modifications, genome annotation, or enzymatic cleavage specificity, and therefore is very useful for targeted validation and confirmation of subsets of peptide hits.

1.3.2: Example Spectrum

Sample MS and MS/MS spectra are shown in Fig. 1A and 1B, respectively. These data were generated on an LTQ-Orbitrap hybrid mass spectrometer [8] with online ESI, where $[M+2H]^{2+}$ and $[M+3H]^{3+}$ ions are expected. The 785.84 Da, +2 charge state precursor will be used for manual sequencing below. The charge state is determined from the 0.5 Da peak separation in the isotope series of the precursor in the MS spectrum (Fig. 1A inset). The peak separation corresponds to a single neutron addition to a peptide atom (*e.g.*, ^{13}C , ^{15}N , ^{32}S). The ratio of the neutron rest mass (approx. 1 Da) to the peak separation is equal to the charge state of the precursor:

$$Z = \frac{\text{Neutron Rest Mass}}{\text{Peak Separation}} \approx \frac{1 \text{ Da}}{0.5 \text{ Da}} = 2$$

When such a precursor is subjected to CAD, singly charged fragment ions are typically formed. As a result, calculation of the singly-charged precursor m/z is a useful first step in manual interpretation. In this example, we remove a proton (H^+): $[M+2H]^{2+} = 785.84 \text{ Da}$, $[M+H]^+ = [2 \times 785.84] - 1.007825 = 1570.67 \text{ Da}$.

The MS/MS spectrum of the 785.84 Da precursor shown in Fig. 1B was generated in the LTQ ion trap mass analyzer via low-energy CAD. A peptide's structure is heteropolymeric, consisting of n amino acids of the following form [53]:

Fragmentation in CAD typically cleaves the amide bond between adjacent amino acids in a peptide (Fig. 2A). A unified nomenclature for the resulting fragment ions was originally proposed by Roepstorff and Fohlman [54], and later modified by Johnson and colleagues [55]. In this scheme, the b series builds up ordinally from b_1 at the N-terminus and the y series builds up ordinally from y_1 at the C-terminus, with addition of single amino acids creating a “ladder” for each ion series. The cleavage reaction is initiated by protonation of an amide nitrogen on the peptide backbone to form fragments containing the original N-terminus (b - and potentially a -type ions) and C-terminus (y -type ions) [56, 57] (Fig. 2B). The gas phase basicity of amide nitrogens is roughly equivalent, meaning that fragment ions corresponding to multiple amide bond cleavages in a peptide will be observed in the associated MS/MS spectrum. Alternative mechanisms have been proposed for generation of b - and y -type ions [58, 59], that yield different fragment ion structures. These products are isobaric to the ones shown in Fig. 2, and hence are indistinguishable by mass measurement.

The vast majority of peptides produced by digestion with trypsin, the most commonly used enzyme in proteomics, will have Lys or Arg at their carboxy-terminus [60]. As drawn above, a single amino acid has one open valence at the N-terminus and one at the C-terminus. To

calculate a y-type ion mass, the ion must first be neutralized by addition of an H atom at the N-terminus and an OH group at the C-terminus, and subsequently protonated to receive a formal charge. Addition of an H atom, an H⁺ atom and one OH group yields a net change in fragment ion mass of: $3 \times 1.007825 + 15.9949146 = 19.01839$ Da. The resultant y₁ masses for Lys and Arg are 147 Da and 175 Da, respectively. The presence of either ion is a good indicator of the identity of the C-terminal amino acid. Neither of these ions is found in Fig. 1B. The alternative method of identifying the C-terminal amino acid is to search for the complementary b ion to y₁ using the equation $b_{n-m} = [M+H]^+ - y_m + H^+$, where n is the length of the peptide and m refers to the ordinal number of an ion in its series. This equation can be used to calculate any b/y complementary pair, as illustrated in Figure 2B for fragmentation of a generic, doubly-charged tryptic peptide. For $[M+H]^+ = 1570.67$ Da:

$$b_{n-1} = 1570.67 - 147 + 1.007825 = 1424.7 \text{ Da} \quad \text{or}$$

$$b_{n-1} = 1570.67 - 175 + 1.007825 = 1396.7 \text{ Da}.$$

There is an ion of mass 1396.4 Da, indicating that Arg is the C-terminal amino acid. Further, there is an ion of similar intensity at 1378.4 Da, corresponding to H₂O loss from b_{n-1}, indicating the presence of Ser, Thr, Glu and/or Asp in the peptide sequence. In this exercise, the b – H₂O peaks are more intense overall than their b series counterparts.

To concatenate the y series, a fragment must be located that corresponds to addition of an amino acid residue mass to the y₁ ion. There is a peak at 246.3 Da, which is 71 Da above the y₁ ion for Arg, corresponding to Ala. The b_{n-2} complementary ion is calculated as:

$$b_{n-2} = 1570.67 - 246.3 + 1.007825 = 1325.4 \text{ Da}$$

There is an ion at 1325.2 Da, corroborated by a 1307.3 Da H₂O loss ion, which confirms the assignment of the 246.3 Da peak as y₂. C-terminal sequencing proceeds in this fashion through the entire amino acid backbone.

The final ion series assignments and partial peptide sequence are shown below. It was not possible to get some of the low mass b ions and the high mass y ions, though often spectra do not offer complete ion series simply because fragmentation favors some ions over others, and as is plain to see in Fig. 1B, the intensity distribution is not even across the m/z range, particularly at the high and low ends of the spectrum. Still, an incomplete de novo sequence tag can be used to find reasonable-probability matches for completion of the sequence and spectral comparison [61, 62].

b _{n-11}	b _{n-10}	b _{n-9}	b _{n-8}	b _{n-7}	b _{n-6}	b _{n-5}	b _{n-4}	b _{n-3}	b _{n-2}	b _{n-1}	
ΔH ₂ O	ΔH ₂ O	ΔH ₂ O	ΔH ₂ O	ΔH ₂ O	ΔH ₂ O	ΔH ₂ O	ΔH ₂ O	ΔH ₂ O	ΔH ₂ O	ΔH ₂ O	
268.2	382.3	497.3	611.3	740.4	869.3	926.3	1073.3	1220.4	1307.3	1378.4	–
b _{n-11}	b _{n-10}	b _{n-9}	b _{n-8}	b _{n-7}	b _{n-6}	b _{n-5}	b _{n-4}	b _{n-3}	b _{n-2}	b _{n-1}	
286.3	400.3	515.4	629.4	758.6	887.4	944.3	1091.3	1238.4	1325.2	1396.4	–
	Asn	Asp	Asn	Glu	Glu	Gly	Phe	Phe	Ser	Ala	Arg
	1285.5	1171.5	1056.5	942.5	813.5	684.5	627.4	480.3	333.3	246.3	–
	y ₁₁	y ₁₀	y ₉	y ₈	y ₇	y ₆	y ₅	y ₄	y ₃	y ₂	

The sequence above, in one-letter code, is NDNEEGFFSAR. This sequence tag was submitted to a species-independent online protein blast search of the NCBI non-redundant protein database at <http://blast.ncbi.nlm.nih.gov/Blast.cgi>. This resulted in a list of 21 hits

with 100% sequence tag coverage, producing two candidate peptides: QGVNDNEEGFFSAR and EGVNDNEEGFFSAR. The $[M+H]^+$ for the precursor to Fig. 1B, determined to within 10 ppm accuracy in the Orbitrap (about 0.015 Da for $[M+H]^+ = 1570.7$ Da), was calculated above as 1570.67 Da. Using the monoisotopic masses in Table 1, the $[M+H]^+$ masses for QGVNDNEEGFFSAR and EGVNDNEEGFFSAR are 1569.68 Da and 1570.66 Da, respectively, uniquely identifying EGVNDNEEGFFSAR as the correct sequence for Fig. 1B, with $n = 14$. Figure 3 shows the assigned MS/MS spectrum.

2.1: Peptide quantitation from mass spectrometry data

There are two strategies for mass-spectrometry based quantification of protein expression, generally referred to as “label” [63–71] and “label-free” [72, 73]. The former utilizes a single analysis of mixed (and isotopically-labeled) samples and the latter relies on comparison of extracted ion chromatograms (XIC) or spectral counting for samples analyzed in a series of LC-MS/MS acquisitions. Ultimately the data are represented as fold changes (i.e., relative ratios) of peptide or reporter ion signal intensities across several samples/conditions. Increasingly these measurements are accompanied by confidence intervals and other relevant statistics [74–77].

In label-based methods, the method of label incorporation and the type of mass spectrometer used for detection may vary significantly. In SILAC [63, 64], for example, cells are cultured in a growth medium containing either a ‘light’ amino acid, with the natural isotopic abundance of each atom, or a heavy analog, enriched with multiple heavy atoms (e.g., ^{13}C , ^{15}N). The growing cells utilize these amino acids in protein synthesis. When differentially labeled cells are combined, MS analysis of protein digests from these cells reveals a user-defined peptide mass shift, allowing the relative quantitation of different biological samples. In AQUA [65], heavy amino acid analogs of peptides from biological samples are synthesized, combined with cell lysate, and digested. In addition to amino acid-based labeling approaches, chemical labels which target amino acid functional groups are also common. ICAT [66, 67] is used for MS-level peptide quantitation. It can be used for cysteine peptide enrichment, which confers the added benefit of reduced mixture complexity, as cysteine is a relatively rare amino acid [78]. iTRAQ [68, 69] and TMT [70, 71] provide quantification at the MS/MS level. The iTRAQ label contains a reporter group that fragments easily during MS/MS. Once combined, the differentially labeled peptides co-elute during LC/MS, and are quantified by well-defined reporter ion masses (e.g., spanning 114–117 Da for the four-plex reagent) in a single MS/MS spectrum. TMT reagents have a similar design, but the reporter ions fall in a higher-mass region of an MS/MS spectrum (126–131 Da for the six-plex reagent).

Challenges in quantification are based on the nature of specific labeling chemistries [79], in addition to fundamental differences in the operation of mass spectrometry detectors and how these translate into measurement variance and significance thresholds for observed peptide and protein ratios. These issues are exacerbated for analysis of post-translational modifications, in which it is necessary to quantify sites on individual peptides, rather than aggregating evidence across multiple peptides from the same protein. Although models have been proposed for specific platforms [76, 77], a universal solution has not yet gained widespread acceptance. In all cases, computation of changes at the protein level involve combining the mean or median of the associated peptide ratios [80].

2.2: Analysis of protein phosphorylation by mass spectrometry

Reversible protein phosphorylation is an important and tightly-regulated catalyst of many cellular functions, including growth, proliferation and transcription. Phosphorylation often

occurs with low stoichiometry or in low-abundance proteins, making it difficult to identify and quantify phosphorylation sites *en masse* from complex biological systems. Strategies to circumvent this problem generally fall into three categories: chemical affinity tag derivatization, selective chromatographic enrichment of phosphorylated side chains and linked-scan mass spectrometer acquisition methods that rely on diagnostic ions specific to phosphorylated peptides.

Chemical derivatization techniques exploit the reactivity of the phosphate functional group. For example, high pH conditions cause β -elimination of phosphoserine and phosphothreonine, yielding dehydroalanine or β -methyldehydroalanine, respectively. The products of β -elimination can be modified through Michael addition to introduce affinity handles (i.e. biotin) for isolation [81] or stable isotopes to afford quantitation across biological states [82]. Although this approach demonstrated enrichment of phosphopeptides from whole yeast tryptic digest [81], deleterious side reactions can be problematic [83]. An alternative strategy uses carbodiimide-catalyzed thiolation of phosphorylated amino acid side chains [84], followed by solid phase capture using an iodoacetyl functionalized resin. Following the initial report, a simplified method coupling modified phosphopeptides to a dendrimer support was described [85], and successfully applied to profile phosphorylation in T-cells.

Chromatographic methods for phosphopeptide enrichment involve selective and reversible binding of the phosphate group directly. Immobilized metal affinity chromatography (IMAC) relies on the high affinity of phosphate toward various metal ions. The most widely used IMAC resins typically employ Fe^{3+} or Ga^{3+} chelated to iminodiacetic acid (IDA) or nitrilotriacetic acid (NTA) to selectively retain phosphorylated molecules [86–88]. Chemical derivatization of carboxyl groups is often required to obtain high specificity for phosphopeptides with Fe^{3+} -IDA [89–91], whereas Fe^{3+} -NTA can achieve >95% specificity from complex mixtures without any derivatization [92]. A seminal report in 2004 demonstrated enrichment of phosphopeptides using titanium dioxide [93], spurring rapid development of metal oxide affinity chromatography (MOAC) in the field of phosphoproteomics. Subsequent reports revealed that small molecule chemical competitors like 2,5-dihydroxybenzoic acid [94] and other acids [95, 96] increased the specificity significantly. Since these initial reports, several metal oxides including zirconium [97] and niobium [98], amongst others [99], have also been shown to enrich phosphopeptides. Ultimately, the different oxides appear to have unique selectivities that may provide complementary phosphoproteome coverage.

Additional modes of chromatography for phosphopeptide enrichment are selective because the acidic phosphate group will bear a formal negative charge at all but very low pH values [100], enabling charge-based separation of non-phosphorylated peptides from their more anionic phosphopeptide counterparts. In strong cation exchange chromatography, phosphopeptides tend to appear in the flowthrough or elute early in a salt concentration gradient at low pH [101, 102]. In strong anion exchange chromatography, phosphopeptides bind more tightly and elute only at lower pH values [103]. In hydrophilic interaction chromatography (HILIC), the polarity of phosphopeptides makes them less hydrophobic, resulting in later elution in a normal phase gradient [104, 105]. ERLIC, an extension of HILIC, can isolate phosphopeptides by polarity under normal phase conditions [106], but exhibits different elution properties from HILIC due to electrostatic repulsion effects between peptides and their charge-matched stationary phase. These techniques have been used with [102, 104, 105, 107] or without [101, 103] other phosphopeptide enrichment methods described above, producing as many as 14,000 phosphorylation site identifications in a single experiment [102].

Bioaffinity separations also play an important role in phosphoproteomics, especially for profiling tyrosine phosphorylation. This subset of phosphorylation is particularly important in regulating cell growth and differentiation [108], but is difficult to detect as it represents <1% of total cellular phosphorylation [109]. Protein [110] and peptide [111] level phosphotyrosine immunoprecipitation have been used to study the signaling of several receptor tyrosine kinases [112].

Mass spectrometry – based phosphopeptide identification methods capitalize on unique features of the ion chemistry of gas phase phosphopeptides. For example, PO_2^- (63 Da) and PO_3^- (79 Da) are fragmentation products derived from the phosphate moiety and can be used to detect the associated phosphopeptide precursors in the context of linked scans on a triple quadrupole mass spectrometer [113]. Another common fragmentation feature of phosphopeptides is the H_3PO_4 neutral loss product, which can be observed indirectly as a precursor mass shift of $(-98 \text{ Da}/|Z|)$ in MS/MS spectra, where $|Z|$ is the magnitude of the peptide charge state. Frequently the neutral loss product dominates many fragmentation pathways in MS/MS, which makes it a convenient and intense ion for detection of phosphorylation [89, 90, 114], but also eliminates structurally informative ions for MS/MS sequence assignment. As a solution, MS/MS/MS (or neutral loss product fragmentation) has been used to identify these product ions [101, 115]. Neutral loss can be especially problematic for basic and/or multiply phosphorylated peptides, and hence these are particularly amenable to sequence identification by non-ergodic dissociation techniques such as ECD [38] or ETD [35].

Coupling multidimensional separations [104, 116] with various ion dissociation techniques [35], improved scoring [117], site localization methods [118, 119] and the quantitative methods described above, thousands of phosphopeptides are routinely profiled across multiple biological states. Further technological advances in instrumentation, as well as methodology for enrichment and fractionation of phosphopeptides will provide for more complete phosphoproteome coverage, and serve to enrich our understanding of the role of cell signaling in normal physiology and human disease.

2.3: Clustering of quantitative data

Although the computational approaches for quantitative proteomics data are not fully resolved, it is nonetheless possible to organize data based on similarity of response (*e.g.*, “upregulated or downregulated”) irrespective of the underlying confidence for a given ratio measurement. Recently the concept of multiplexing has enabled the comparison of up to eight biological conditions simultaneously [120, 121]. In such cases, peptides are clustered based on similar quantitative profiles across all the conditions. The underlying assumption is that peptides exhibiting similar behavior within a cluster will likely share a common biological pathway or response. For example, Wolf-Yadlin *et al.* used a method called self-organizing maps to cluster 62 phosphopeptides across 4 time points and 4 conditions to identify the upstream stimulus, such as EGF stimulation or HER2 activation, responsible for the regulation of phosphorylation activity [122]. Tang *et al.* also used this approach with quantitative phosphoproteomic data to generate 5 clusters based on a classical k-means algorithm to identify groups of effectors, positive regulators, and negative regulators in the Wnt signaling pathway [123]. Clustering not only separates quantitative mass spectrometry data into meaningful groups but is also used to identify recurring patterns that may yield further biological insight.

2.4: Integrating biological knowledge with MS/MS data

Biological insight can be gleaned from high-throughput data by integrating prior knowledge stored in publicly available databases. These biological annotation databases include: (i) Gene Ontology (GO) for mining protein function, localization, and process [124], (ii) literature-curated pathway databases such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway database [125] and BioCarta [126], in addition to (iii) experimental data warehouses such as BioGRID [127] that contains protein interaction data from high-throughput experiments and the Human Protein Reference Database (HPRD) [128] that includes curated enzyme-substrate data.

2.5: Enrichment analyses

GO is divided into three categories: (i) molecular function, (ii) cellular component, and (iii) biological processes. Molecular function distinguishes proteins based on their functional role in the cell and includes terms such as “MAP kinase activity.” The cellular component category groups proteins based on cellular localization and, at a higher resolution, complexes such as “IkappaB kinase complex.” Finally, biological processes include terms such as “MAPKKK cascade” that consist of proteins participating in the particular biological event. Similar to GO biological process, pathway databases such as KEGG Pathway and BioCarta assemble groups of proteins that belong to the same signaling or metabolic pathway. Unlike GO, which only contains lists of functional protein groups, the aforementioned pathway databases also store images of how the proteins interact within each pathway. The typical use-case of these annotation databases in automated analysis is to quantify the enrichment of a particular functional group of proteins within an experimental dataset. Statistical methods such as Fisher’s Exact Test calculate a significance level for the degree of enrichment of a particular protein group such as a pathway, complex, or enzyme class within an experimentally determined list of proteins (Figure 4). Given the role of mass spectrometry in interrogating signaling pathways via phosphoproteomic experiments, identifying members of a complex [129], or targeting enzyme classes via activity-based profiling assays, enrichment meta-analyses can facilitate the assessment of the experimental result. For example, Ficarro *et al.* used KEGG Pathway enrichment analysis to identify cell adhesion, junction, and matrix proteins as potentially involved in embryonic stem cell self-renewal [130]. Several online tools exist for conducting meta-analyses given an input list of experimentally determined proteins [131–133].

2.6: Networks

Cell signaling involves a complex network of protein-protein interactions and enzymatic reactions that form feedforward and feedback loops, along with a significant degree of “crosstalk” between various cascades [134–137]. Enrichment analyses treat pathways as segregated entities, ignoring the effect of shared protein members. In addition, only a fraction of proteins relative to the entire proteome are assigned to pathways in the public databases, such that many of the proteins detected in a large-scale experiment are ignored in subsequent network or pathway analyses. The same problem occurs with other annotations such as GO categories where proteins are often excluded due to incomplete curation. In such scenarios it is often necessary to gain a higher resolution picture of the experimental data. Networks where nodes represent proteins and edges represent biochemical or physical interaction offer a solution and provide a convenient way to integrate publicly available interaction data with experimental results. Protein-protein interaction (PPI) data now account for more than 100,000 entries in databases such as BioGRID. These interactions are derived from high-throughput assays such as Yeast two-hybrid (Y2H) and affinity purification (AP-MS), with the former contributing a majority of the pairwise interactions in

these databases. PPI or physical binding interactions are represented as undirected edges in protein networks (Figure 5). Biochemical interaction data such as enzyme-substrate relationships found in curated databases such as HPRD can also be integrated into protein networks as directed edges, where an arrow indicates reactivity between enzyme and substrate. Directed edges may also be used to distinguish reversible post-translational modification or activation versus inhibition of a substrate (Figure 5). Although PPI data outnumber biochemical interaction data by two orders of magnitude, directed edges are nonetheless more appropriate for interrogation of activity-based profiling data (*e.g.*, phosphorylation or other enzyme-substrate relationships).

Network diagrams are built upon detected proteins as opposed to generic curated pathway maps, thus providing a context-specific and high-resolution representation of experimental proteomics data. Furthermore, network representations lend themselves to in-depth analysis based on decades of graph theory research that can be directly applied to protein networks. For example, Breitkreutz *et al.* use the graph theoretical concept of characteristic path length to illustrate the robustness of the kinase-kinase physical interaction network in yeast [138]. Networks are also widely used to visualize the local and global dynamics of entire datasets to elucidate condition-dependent recruitment of binding partners or relative activity of pathway components (Figure 5) [139].

Unlike pathway enrichment analyses, which are robust to missing proteins, high-resolution networks can become unstable if key network proteins are not present within the experimental data. In these cases many detected proteins become orphans in the network with no edges connecting them to other network members. To account for this problem, Huang and Fraenkel applied another graph theory algorithm called the Prize-Collecting Steiner Tree (PCST) to connect detected proteins by selectively introducing hidden nodes (undetected proteins) in the network [140]. The authors were thus able to identify hidden signaling components in the yeast pheromone response pathway by integrating LC-MS/MS and publicly available interaction data.

2.7: Prediction from quantitative MS/MS data

Although the PCST algorithm was used to predict new signaling pathway members, the aforementioned clustering and data integration methods are mostly used to reveal biologically significant and recurring patterns in mass spectrometry data. However, with continued development of proteomic methods for consistently and simultaneously quantifying proteins, computational scientists have been able to apply machine learning methods to build predictive models. Kumar *et al.* used a machine learning method called partial least squares regression (PLSR) to predict HER2-mediated changes in migration and proliferation from phosphoproteomic data [141]. As a side effect of constructing a predictive model that could use new data to predict cell behavior, the authors identified a small set of 9 phosphorylation sites that maintained high fidelity with respect to predictive power. Woolf *et al.* used a graphical machine learning method called Bayesian Networks to construct a model of 28 signaling proteins governing embryonic stem cell fate decisions [142]. The model was trained on 49 quantitative phosphorylation measurements per protein. Output of the machine learning method is a causal network where nodes are variables, which include phosphorylation levels of proteins and phenotypic measurements such as differentiation rate, and directed edges represent causal relationships between variables. This model can be used to infer or predict new values of a subset of variables, such as the differentiation rate given a certain level of STAT3 phosphorylation. Similar to PLSR, constructing a Bayesian network model has the added benefit of identifying causal relationships between proteins based on the network structure. Woolf *et al.* not only identified known signaling cascades such as the

classical RAF-MEK-ERK cascade but also proposed a new prediction that MEK3/6 causally affects MEK1/2 activity.

2.8: Integrated computational analysis of quantitative phosphoproteomic data

In this section we will discuss how several of the aforementioned tools and methods were used to interrogate phosphorylation dynamics during early differentiation of human embryonic stem cells (hESCs) [143]. Van Hoof *et al.* identified 1399 phosphoproteins with 3067 phosphosites, of which 1091 were regulated across three different time points. The regulated phosphoproteins belonged to several pathways including BMP, PI3K, WNT and JNK signaling pathways, as determined by enrichment analysis described above. To better understand temporal behavior, the regulated phosphopeptides were grouped using k-means clustering. Clustering revealed that the most dramatic phosphorylation changes occurred during the first hour of differentiation. Interestingly, different phosphosites on certain hyperphosphorylated proteins, such as tumor suppressor p53-binding protein 1, exhibited different temporal profiles and therefore segregated into different clusters, suggesting that certain key proteins act as platforms for integrating signals from several kinases. In order to gain higher resolution, the authors used the NetworKIN algorithm [144] to predict kinases upstream of the regulated phosphosites and thus generated the first kinase-substrate database for hESCs. CDK1/2 was predicted to be the most active upstream kinase accounting for 26% of the hESC phosphosites. Focusing on the regulated phosphosites on kinases, the authors noted that the phosphonetwork, connecting upstream predicted kinases to experimentally determined phosphorylated kinases, expanded over time during differentiation.

Although important in generating hypotheses about upstream phosphorylation cascades, kinase-substrate prediction algorithms face several challenges as noted by the authors of the NetworKIN algorithm [144], which combines context specificity from the STRING database [145] with phosphosite motif scanning algorithms such as NetPhosK [146] and Scansite [147]. First, the algorithm is prone to errors from the phosphopeptide data, incorrect kinase family assignment to phosphorylation motifs, and errors in the probabilistic context association network. Second, predictive algorithms have poor coverage of the kinome (<22%) and seem to focus on pleiotropic kinase families, which are potentially responsible for most phosphosites, and consequently ignore the more specific and selectively expressed kinases. Given that modern phosphoproteomic data analyses rely heavily on prediction of upstream kinases, greater availability of accurate kinase-substrate data is required not only to increase our knowledge of the global kinase-substrate network but also to improve prediction algorithms.

3.1: Activity-Based Protein Profiling (ABPP)

Although traditional shotgun proteomics experiments generate large catalogs of proteins [148], and can also provide relative quantification data [68], they do not directly interrogate the catalytic activity of any protein class. This information is an essential piece of the systems biology puzzle as proteins are the major elements of cellular catalysis. As an example, consider proteases, which are synthesized as inactive zymogens. An increased amount of protease in this form does not translate into increased activity until specific processing or co-factor binding events activate them [149]. In addition many kinases are regulated by phosphorylation within the activation loop, an event that can yield an increase in catalytic activity of up to four orders of magnitude [150]. So although protein profiling is important to identify relevant enzymes, mere changes in protein quantity do not necessarily correlate with overall activity. Moreover, many enzymes are expressed at relatively low levels making them difficult to detect in global profiling experiments. The emergent field of

ABPP helps to bridge this gap, allowing for system-wide detection of active enzymes in lysates, cells, or even live animals.

ABPP relies on chemical probes that have two essential components (Figure 6A). The first is a reactive group that targets a specific enzyme class. Probes have been developed for many types of enzymes including kinases [151], serine hydrolases [152], cysteine proteases [153], phosphatases [154], glycosidases [155], and several others [156]. In addition to a reactive group, the probe may have structural components that improve selectivity for a particular enzyme class. Patricelli *et al.* designed ADP/ATP analogues that target the ATP binding pocket of kinases and position an activated carbonyl near the epsilon amino group of conserved lysine residues in these proteins [151]. Second, the probe must facilitate detection of its target through a reporter, such as a fluorescent tag or an affinity handle that allows isolation (*e.g.*, biotin). Note that the choice of reporter dictates the type of workflow used to assess protein activity; for example, fluorescent tags are typically used for gel-based assays, whereas affinity handles can facilitate isolation of proteins or peptides for LC/MS-based interrogation. Many of the early studies utilized probes directly attached to reporter tags through a linker region [152, 153]. These types of probes tend to be larger in molecular weight, are not easily internalized by cells or tissue, and are typically used after lysis of cells. This approach has the disadvantage of removing proteins from their native environment and may often lead to loss of activity, and hence poor recovery depending on the probe used. To circumvent this limitation, Speers *et al.* developed click chemistry-based ABPP, or CC-ABPP (Figure 6B) [157]. Click chemistry is a philosophy of synthesis [158] that advocates building target molecules through highly efficient and favorable reactions such as Huisgen's 1,3 dipolar cycloaddition that couples azide and alkyne moieties [159]. This reaction has an interesting history, and has been optimized through copper catalysis [160, 161] as well as ligands that stabilize copper in the correct oxidation state [162]. In CC-ABPP, the probe and reporter are synthesized as discrete functional molecules that are linked through the bio-orthogonal azide/alkyne cycloaddition. In a typical experiment, probes are synthesized with reactive and binding groups that target the enzyme of interest, along with an alkyne or azide moiety. This molecule, without the reporter, is more compact, and (if properly designed) can be added to live cells or fed to animals. After reaction of the probe with target proteins, cells or tissues are lysed, and the reporter is selectively attached to the tagged protein through the click chemistry reaction. One drawback of the copper catalyzed CC-ABPP is that reporter incorporation must occur *ex vivo*, due to the toxicity of copper. Bertozzi and colleagues have developed a copper free click reaction based on strained, fluorinated cyclooctynes that allows reporters to be attached to labeled molecules in living organisms [163, 164]. Although these experiments were used to label cell surface glycans that carried azide functionalities and were not activity-based, they nonetheless demonstrated the potential to apply click chemistry for dynamic imaging of targets in living organisms [165].

Once target proteins are labeled, they are typically quantified using gel-(fluorescence scanning or avidin blotting) or LC/MS- (for affinity tagging) based assays. Gel assays allow for a higher-throughput assessment of probe binding in general, and are also useful to survey differences in protein activity between cell states. LC-MS/MS experiments are utilized to determine the identity of tagged proteins, where two levels of information can be obtained. First, tryptic peptides from isolated proteins can be analyzed (Figure 7A). Second, peptides covalently conjugated to the probe, which represent the site of activity, can also be identified (Figure 7B). Although it is possible to obtain both types of information by digesting isolated proteins and subjecting their peptides to LC-MS/MS analysis, it may be difficult to identify the probe-conjugated fragments amongst the other, and far more abundant, tryptic peptides. An alternative strategy (Figure 7C) is to separate tryptic fragments from probe conjugates, and analyze these subsets in separate LC-MS/MS analyses. For example, Weerapana *et al.*

[166] have developed a strategy where the reporter contains a TEV protease cleavage site as well as a biotin affinity handle. After proteins are labeled, the reporter module is attached in the click reaction, and labeled proteins are isolated by streptavidin. The proteins are digested on-resin, and the tryptic fragments are analyzed by LC-MS/MS. Next, a TEV digestion is employed to release only the probe-labeled peptides, which are analyzed separately.

Variations of the methodology described above have been applied with myriad chemical probes to assess enzyme activities in both cellular systems and animal models. These studies have met with a great deal of success and have uncovered enzymes associated with human disease and allowed for the development of selective inhibitors [167–171].

4.1: Concluding Remarks

Mass spectrometry-based proteomics has emerged as the method of choice for large-scale protein analyses because of technological advancements on three fronts: 1) improvements in mass spectrometer acquisition speed and data quality (e.g., mass resolution, mass accuracy, and detection limit), 2) experimental approaches that improve dynamic range through biochemical enrichment and/or chromatographic fractionation, and 3) bioinformatics approaches that reveal quantitative trends and relevant biological pathways in complex datasets. Some of the more prominent developments have been discussed in this review in the context of bottom-up proteomic analysis. Despite these advances, the wide dynamic range of protein expression along with the number and variable stoichiometry of post-translational modifications will likely keep comprehensive proteome characterization well beyond our reach. In addition, the stochastic nature of discovery or “shotgun” proteomics methods often limits the reproducibility of high-throughput studies; this scenario is particularly problematic for protein biomarker discovery efforts, where putative markers would preferably be subject to thorough analytical validation prior to advancement through the pre-clinical pipeline. Fortunately the field is making rapid progress on all fronts, from standardization of pre-analytical sample treatment procedures (e.g., storage and distribution) to the use of multiple reaction monitoring mass spectrometry (MRM-MS) for targeted and reproducible biomarker studies that span multiple laboratories [172–175]. When used in combination with the rapidly maturing enrichment/separation techniques described above, these proteomics methods may become a standard for pre-clinical biomarker discovery and validation. Activity-based protein profiling is a more recent development that will facilitate analysis of important enzymatically active protein subclasses. These functional data represent another important step in the application of proteomics to the field of personalized medicine, where treatments will be tailored based on patient-specific biomarkers or quantitative elucidation of associated disease pathways.

Acknowledgments

Generous support for this work was provided by the Dana-Farber Cancer Institute and the National Institutes of Health, NHGRI (P50HG004233), and NINDS (P01NS047572).

References

1. Weston AD, Hood L. Systems biology, proteomics, and the future of health care: toward predictive, preventative, and personalized medicine. *J Proteome Res.* 2004; 3(2):179–96. [PubMed: 15113093]
2. Hu Q, et al. The Orbitrap: a new mass spectrometer. *J Mass Spectrom.* 2005; 40(4):430–443. [PubMed: 15838939]
3. Loboda AV, et al. A tandem quadrupole/time-of-flight mass spectrometer with a matrix-assisted laser desorption/ionization source: design and performance. *Rapid Commun Mass Spectrom.* 2000; 14(12):1047–57. [PubMed: 10861986]

4. Wenzel RJ, et al. Analysis of megadalton ions using cryodetection MALDI time-of-flight mass spectrometry. *Anal Chem.* 2005; 77(14):4329–37. [PubMed: 16013843]
5. Zhou M, Robinson CV. When proteomics meets structural biology. *Trends Biochem Sci.* 2010; 35(9):522–9. [PubMed: 20627589]
6. Schwartz JC, Senko MW, Syka JE. A two-dimensional quadrupole ion trap mass spectrometer. *J Am Soc Mass Spectrom.* 2002; 13(6):659–69. [PubMed: 12056566]
7. March RE. An Introduction to Quadrupole Ion Trap Mass Spectrometry. *J Mass Spectrom.* 1997; 32(4):351–369.
8. Makarov A, et al. Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal Chem.* 2006; 78(7):2113–20. [PubMed: 16579588]
9. Yates JR, et al. Performance of a linear ion trap-Orbitrap hybrid for peptide analysis. *Anal Chem.* 2006; 78(2):493–500. [PubMed: 16408932]
10. Syka JE, et al. Novel linear quadrupole ion trap/FT mass spectrometer: performance characterization and use in the comparative analysis of histone H3 post-translational modifications. *J Proteome Res.* 2004; 3(3):621–6. [PubMed: 15253445]
11. Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem.* 1988; 60(20):2299–301. [PubMed: 3239801]
12. Medzihradszky KF, et al. The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. *Anal Chem.* 2000; 72(3):552–8. [PubMed: 10695141]
13. Laiko VV, Baldwin MA, Burlingame AL. Atmospheric pressure matrix-assisted laser desorption/ionization mass spectrometry. *Anal Chem.* 2000; 72(4):652–7. [PubMed: 10701247]
14. Karas M, et al. Matrix-assisted ultraviolet laser desorption of non-volatile compounds. *Int J Mass Spectrom Ion Proc.* 1987; 78:53–68.
15. Zenobi R, Knochenmuss R. Ion formation in MALDI mass spectrometry. *Mass Spectrom Rev.* 1998; 17(5):337–366.
16. Tanaka K, et al. Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry.* 1988; 2(8):151–153.
17. Fenn JB, et al. Electrospray ionization for mass spectrometry of large biomolecules. *Science.* 1989; 246(4926):64–71. [PubMed: 2675315]
18. Morris HR, et al. High Sensitivity Collisionally-activated Decomposition Tandem Mass Spectrometry on a Novel Quadrupole/Orthogonal-acceleration Time-of-flight Mass Spectrometer. *Rapid Communications in Mass Spectrometry.* 1996; 10(8):889–896. [PubMed: 8777321]
19. Yamashita M, Fenn JB. Electrospray ion source. Another variation on the free-jet theme. *The Journal of Physical Chemistry.* 1984; 88(20):4451–4459.
20. Yamashita M, Fenn JB. Negative ion production with the electrospray ion source. *The Journal of Physical Chemistry.* 1984; 88(20):4671–4675.
21. He F, et al. Theoretical and experimental prospects for protein identification based solely on accurate mass measurement. *J Proteome Res.* 2004; 3(1):61–7. [PubMed: 14998164]
22. Taylor, BN. International Bureau of Weights and Measures. NIST special publication. 2001. Vol. viii. U.S. Dept. of Commerce, Technology Administration; U.S. G.P.O: Gaithersburg, MD; Washington; 2001. The international system of units (SI); p. 68 For sale by the Supt. of Docs
23. Todd JFJ. Recommendations for nomenclature and symbolism for mass spectroscopy (including an appendix of terms used in vacuum technology). (Recommendations 1991). *Pure Appl Chem.* 1991; 63(10):1541–1566.
24. Mamyrin BA. Time-of-flight mass spectrometry (concepts, achievements, and prospects). *Int J Mass Spectrom.* 2001; 206(3):251–266.
25. Grundwürmer JM, et al. High-resolution mass spectrometry in a linear time-of-flight mass spectrometer. *Int J Mass Spectrom Ion Proc.* 1994; 131:139–148.
26. Bienvenut WV, et al. Matrix-assisted laser desorption/ionization-tandem mass spectrometry with high resolution and sensitivity for identification and characterization of proteins. *Proteomics.* 2002; 2(7):868–876. [PubMed: 12124932]

27. Pelander A, et al. Evaluation of a High Resolving Power Time-of-Flight Mass Spectrometer for Drug Analysis in Terms of Resolving Power and Acquisition Rate. *Journal of the American Society for Mass Spectrometry*. 2011; 22(2):379–385. [PubMed: 21472597]
28. Pandhal J, et al. Improving N-glycosylation efficiency in *Escherichia coli* using shotgun proteomics, metabolic network analysis, and selective reaction monitoring. *Biotechnology and Bioengineering*. 2011; 108(4):902–912. [PubMed: 21404263]
29. Gorshkov MV, Zubarev RA. On the accuracy of polypeptide masses measured in a linear ion trap. *Rapid Commun Mass Spectrom*. 2005; 19(24):3755–3758. [PubMed: 16302203]
30. He F, Hendrickson CL, Marshall AG. Baseline mass resolution of peptide isobars: a record for molecular mass resolution. *Anal Chem*. 2001; 73(3):647–50. [PubMed: 11217775]
31. Zubarev RA, Håkansson P, Sundqvist B. Accuracy Requirements for Peptide Characterization by Monoisotopic Molecular Mass Measurements. *Analytical Chemistry*. 1996; 68(22):4060–4063.
32. McLuckey SA, et al. Ion spray liquid chromatography/ion trap mass spectrometry determination of biomolecules. *Analytical Chemistry*. 1991; 63(4):375–383.
33. Olsen JV, et al. Higher-energy C-trap dissociation for peptide modification analysis. *Nat Meth*. 2007; 4(9):709–712.
34. Hunt DF, et al. Sequence analysis of polypeptides by collision activated dissociation on a triple quadrupole mass spectrometer. *Biomed Mass Spectrom (now incorporated into J Mass Spectrom)*. 1981; 8(9):397–408.
35. Syka JE, et al. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A*. 2004; 101(26):9528–33. [PubMed: 15210983]
36. Mikesk LM, et al. The utility of ETD mass spectrometry in proteomic analysis. *Biochim Biophys Acta*. 2006; 1764(12):1811–22. [PubMed: 17118725]
37. McLafferty FW, et al. Electron capture dissociation of gaseous multiply charged ions by Fourier-transform ion cyclotron resonance. *J Am Soc Mass Spectrom*. 2001; 12(3):245–9. [PubMed: 11281599]
38. Zubarev RA, Kelleher NL, McLafferty FW. Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *Journal of the American Chemical Society*. 1998; 120(13):3265–3266.
39. Turecek F, McLafferty FW. Non-ergodic behavior in acetone-enol ion dissociations. *Journal of the American Chemical Society*. 1984; 106(9):2525–2528.
40. Perkins DN, et al. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*. 1999; 20(18):3551–67. [PubMed: 10612281]
41. Ducret A, et al. High throughput protein characterization by automated reverse-phase chromatography/electrospray tandem mass spectrometry. *Protein Sci*. 1998; 7(3):706–19. [PubMed: 9541403]
42. Phenix. Available from: <http://www.genecbio.com/products/phenyx/>
43. Colinge J, et al. OLAV: towards high-throughput tandem mass spectrometry data identification. *Proteomics*. 2003; 3(8):1454–63. [PubMed: 12923771]
44. Craig R, Beavis RC. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics*. 2004; 20(9):1466–7. [PubMed: 14976030]
45. Sadygov RG, Cociorva D, Yates JR 3rd. Large-scale database searching using tandem mass spectra: looking up the answer in the back of the book. *Nat Methods*. 2004; 1(3):195–202. [PubMed: 15789030]
46. Kapp E, Schütz F. Overview of Tandem Mass Spectrometry (MS/MS) Database Search Algorithms. *Curr Protoc Prot Sci*. 2001; Chapter 25(Unit 25.2)
47. Balgley BM, et al. Comparative evaluation of tandem MS search algorithms using a target-decoy search strategy. *Mol Cell Proteomics*. 2007; 6(9):1599–608. [PubMed: 17533222]
48. Nesvizhskii AI, et al. A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem*. 2003; 75(17):4646–58. [PubMed: 14632076]
49. Nesvizhskii AI, Aebersold R. Interpretation of shotgun proteomic data: the protein inference problem. *Mol Cell Proteomics*. 2005; 4(10):1419–40. [PubMed: 16009968]

50. Carr S. The Need for Guidelines in Publication of Peptide and Protein Identification Data: Working Group On Publication Guidelines For Peptide And Protein Identification Data. *Mol Cell Proteomics*. 2004; 3(6):531–533. [PubMed: 15075378]
51. Mann M, Hendrickson RC, Pandey A. Analysis of proteins and proteomes by mass spectrometry. *Annu Rev Biochem*. 2001; 70:437–73. [PubMed: 11395414]
52. Palagi PM, et al. Proteome informatics I: bioinformatics tools for processing experimental data. *Proteomics*. 2006; 6(20):5435–44. [PubMed: 16991191]
53. Illustrations generated using ACD/Chemsketch Freeware, version 12.01. Advanced Chemistry Development, Inc; Toronto, ON, Canada: 2010. www.acdlabs.com
54. Roepstorff P, Fohlman J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom*. 1984; 11(11):601. [PubMed: 6525415]
55. Johnson RS, et al. Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. *Anal Chem*. 1987; 59(21):2621–5. [PubMed: 3688448]
56. Schlosser A, Lehmann WD. Five-membered ring formation in unimolecular reactions of peptides: a key structural element controlling low-energy collision-induced dissociation of peptides. *J Mass Spectrom*. 2000; 35(12):1382–90. [PubMed: 11180628]
57. Polce MJ, Ren D, Wesdemiotis C. Dissociation of the peptide bond in protonated peptides. *J Mass Spectrom*. 2000; 35(12):1391–8. [PubMed: 11180629]
58. Griffiths, WJ. The Encyclopedia of Mass Spectrometry. In: Gross, ML.; Caprioli, RM., editors. *Biological Applications: Part A: Peptides and Proteins*. 2. Elsevier; Amsterdam; Boston: 2005. p. 71–74.
59. Chi A, et al. Analysis of phosphorylation sites on proteins from *Saccharomyces cerevisiae* by electron transfer dissociation (ETD) mass spectrometry. *Proc Nat Acad Sci USA*. 2007; 104(7):2193–2198. [PubMed: 17287358]
60. Olsen JV, Ong SE, Mann M. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol Cell Proteomics*. 2004; 3(6):608–14. [PubMed: 15034119]
61. Mann M, Wilm M. Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal Chem*. 1994; 66(24):4390–9. [PubMed: 7847635]
62. BLAST: Basic Local Alignment Search Tool. National Center for Biotechnology Information; <http://blast.ncbi.nlm.nih.gov/Blast.cgi>
63. Ong SE, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics*. 2002; 1(5):376–86. [PubMed: 12118079]
64. Mann M. Functional and quantitative proteomics using SILAC. *Nat Rev Mol Cell Biol*. 2006; 7(12):952–8. [PubMed: 17139335]
65. Gerber SA, et al. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci U S A*. 2003; 100(12):6940–5. [PubMed: 12771378]
66. Gygi SP, et al. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol*. 1999; 17(10):994–9. [PubMed: 10504701]
67. Gygi SP, et al. Proteome analysis of low-abundance proteins using multidimensional chromatography and isotope-coded affinity tags. *J Proteome Res*. 2002; 1(1):47–54. [PubMed: 12643526]
68. Ross PL, et al. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics*. 2004; 3(12):1154–69. [PubMed: 15385600]
69. Phanstiel D, et al. Peptide and protein quantification using iTRAQ with electron transfer dissociation. *J Am Soc Mass Spectrom*. 2008; 19(9):1255–62. [PubMed: 18620867]
70. Thompson A, et al. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem*. 2003; 75(8):1895–904. [PubMed: 12713048]
71. Dayon L, et al. Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags. *Anal Chem*. 2008; 80(8):2921–31. [PubMed: 18312001]
72. Wang W, et al. Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. *Anal Chem*. 2003; 75(18):4818–26. [PubMed: 14674459]

73. Anderle M, et al. Quantifying reproducibility for differential proteomics: noise analysis for protein liquid chromatography-mass spectrometry of human serum. *Bioinformatics*. 2004; 20(18):3575–82. [PubMed: 15284095]
74. Bantscheff M, et al. Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem*. 2007; 389(4):1017–31. [PubMed: 17668192]
75. Breitwieser FP, et al. General statistical modeling of data from protein relative expression isobaric tags. *J Proteome Res*. 2011; 10(6):2758–66. [PubMed: 21526793]
76. auf dem Keller U, et al. A statistics-based platform for quantitative N-terminome analysis and identification of protease cleavage products. *Mol Cell Proteomics*. 2010; 9(5):912–27. [PubMed: 20305283]
77. Zhang Y, et al. A robust error model for iTRAQ quantification reveals divergent signaling between oncogenic FLT3 mutants in acute myeloid leukemia. *Mol Cell Proteomics*. 2010; 9(5):780–90. [PubMed: 20019052]
78. Fenyo D, Qin J, Chait BT. Protein identification using mass spectrometric information. *Electrophoresis*. 1998; 19(6):998–1005. [PubMed: 9638946]
79. Gan CS, et al. Technical, experimental, and biological variations in isobaric tags for relative and absolute quantitation (iTRAQ). *J Proteome Res*. 2007; 6(2):821–7. [PubMed: 17269738]
80. Boehm AM, et al. Precise protein quantification based on peptide quantification using iTRAQ. *BMC Bioinformatics*. 2007; 8:214. [PubMed: 17584939]
81. Oda Y, Nagasu T, Chait BT. Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat Biotechnol*. 2001; 19(4):379–82. [PubMed: 11283599]
82. Goshe MB, et al. Phosphoprotein isotope-coded affinity tags: application to the enrichment and identification of low-abundance phosphoproteins. *Anal Chem*. 2002; 74(3):607–16. [PubMed: 11838682]
83. McLachlin DT, Chait BT. Improved beta-elimination-based affinity purification strategy for enrichment of phosphopeptides. *Anal Chem*. 2003; 75(24):6826–36. [PubMed: 14670042]
84. Zhou H, Watts JD, Aebersold R. A systematic approach to the analysis of protein phosphorylation. *Nat Biotechnol*. 2001; 19(4):375–8. [PubMed: 11283598]
85. Tao WA, et al. Quantitative phosphoproteome analysis using a dendrimer conjugation chemistry and tandem mass spectrometry. *Nature Methods*. 2005; 2(8):591–598. [PubMed: 16094384]
86. Nuwaysir LM, Stults JT. Electrospray ionization mass spectrometry of phosphopeptides isolated by on-line immobilized metal-ion affinity chromatography. *Journal of the American Society for Mass Spectrometry*. 1993; 4(8):662–669.
87. Watts JD, et al. Identification by electrospray ionization mass spectrometry of the sites of tyrosine phosphorylation induced in activated Jurkat T cells on the protein tyrosine kinase ZAP-70. *J Biol Chem*. 1994; 269(47):29520–9. [PubMed: 7961936]
88. Posewitz MC, Tempst P. Immobilized gallium(III) affinity chromatography of phosphopeptides. *Anal Chem*. 1999; 71(14):2883–92. [PubMed: 10424175]
89. Ficarro SB, et al. Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat Biotechnol*. 2002; 20(3):301–5. [PubMed: 11875433]
90. Ficarro S, et al. Phosphoproteome analysis of capacitated human sperm. Evidence of tyrosine phosphorylation of a kinase-anchoring protein 3 and valosin-containing protein/p97 during capacitation. *J Biol Chem*. 2003; 278(13):11579–89. [PubMed: 12509440]
91. Ndassa YM, et al. Improved immobilized metal affinity chromatography for large-scale phosphoproteomics applications. *J Proteome Res*. 2006; 5(10):2789–99. [PubMed: 17022650]
92. Ficarro SB, et al. Magnetic bead processor for rapid evaluation and optimization of parameters for phosphopeptide enrichment. *Anal Chem*. 2009; 81(11):4566–75. [PubMed: 19408940]
93. Pinkse MW, et al. Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal Chem*. 2004; 76(14):3935–43. [PubMed: 15253627]
94. Larsen MR, et al. Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol Cell Proteomics*. 2005; 4(7):873–86. [PubMed: 15858219]

95. Tsai CF, et al. Immobilized metal affinity chromatography revisited: pH/acid control toward high selectivity in phosphoproteomics. *J Proteome Res.* 2008; 7(9):4058–69. [PubMed: 18707149]
96. Sugiyama N, et al. Phosphopeptide enrichment by aliphatic hydroxy acid-modified metal oxide chromatography for nano-LC-MS/MS in proteomics applications. *Mol Cell Proteomics.* 2007; 6(6):1103–9. [PubMed: 17322306]
97. Kweon HK, Hakansson K. Selective zirconium dioxide-based enrichment of phosphorylated peptides for mass spectrometric analysis. *Anal Chem.* 2006; 78(6):1743–9. [PubMed: 16536406]
98. Ficarro SB, et al. Niobium(V) oxide (Nb₂O₅): application to phosphoproteomics. *Anal Chem.* 2008; 80(12):4606–13. [PubMed: 18491922]
99. Leitner A. Phosphopeptide enrichment using metal oxide affinity chromatography. *Trac-Trends in Analytical Chemistry.* 2010; 29(2):177–185.
100. Hung CW, Kubler D, Lehmann WD. pI-based phosphopeptide enrichment combined with nanoESI-MS. *Electrophoresis.* 2007; 28(12):2044–52. [PubMed: 17523138]
101. Beausoleil SA, et al. Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc Natl Acad Sci U S A.* 2004; 101(33):12130–5. [PubMed: 15302935]
102. Dephoure N, et al. A quantitative atlas of mitotic phosphorylation. *Proc Natl Acad Sci U S A.* 2008; 105(31):10762–7. [PubMed: 18669648]
103. Dai J, et al. Fully automatic separation and identification of phosphopeptides by continuous pH-gradient anion exchange online coupled with reversed-phase liquid chromatography mass spectrometry. *J Proteome Res.* 2009; 8(1):133–41. [PubMed: 19053533]
104. Albuquerque CP, et al. A multidimensional chromatography technology for in-depth phosphoproteome analysis. *Mol Cell Proteomics.* 2008; 7(7):1389–96. [PubMed: 18407956]
105. McNulty DE, Annan RS. Hydrophilic interaction chromatography reduces the complexity of the phosphoproteome and improves global phosphopeptide isolation and detection. *Mol Cell Proteomics.* 2008; 7(5):971–80. [PubMed: 18212344]
106. Alpert AJ. Electrostatic repulsion hydrophilic interaction chromatography for isocratic separation of charged solutes and selective isolation of phosphopeptides. *Anal Chem.* 2008; 80(1):62–76. [PubMed: 18027909]
107. Zarei M, et al. Comparison of ERLIC-TiO(2), HILIC-TiO(2), and SCX-TiO(2) for Global Phosphoproteomics Approaches. *J Proteome Res.* 2011
108. Hunter T. The Croonian Lecture 1997. The Phosphorylation of Proteins on Tyrosine: Its Role in Cell Growth and Disease. *Philosophical Transactions: Biological Sciences.* 1998; 353(1368): 583–605. [PubMed: 9602534]
109. Hunter T, Sefton BM. Transforming gene product of Rous sarcoma virus phosphorylates tyrosine. *Proc Natl Acad Sci U S A.* 1980; 77(3):1311–5. [PubMed: 6246487]
110. Salomon AR, et al. Profiling of tyrosine phosphorylation pathways in human cells using mass spectrometry. *Proceedings of the National Academy of Sciences.* 2003; 100(2):443–448.
111. Rush J, et al. Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nature Biotechnology.* 2005; 23(1):94–101.
112. Zhang Y, et al. Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Mol Cell Proteomics.* 2005; 4(9):1240–50. [PubMed: 15951569]
113. Annan RS, et al. A multidimensional electrospray MS-based approach to phosphopeptide mapping. *Anal Chem.* 2001; 73(3):393–404. [PubMed: 11217738]
114. Schlosser A, et al. Analysis of Protein Phosphorylation by a Combination of Elastase Digestion and Neutral Loss Tandem Mass Spectrometry. *Analytical Chemistry.* 2000; 73(2):170–176. [PubMed: 11199962]
115. Song C, et al. Reversed-phase-reversed-phase liquid chromatography approach with high orthogonality for multidimensional separation of phosphopeptides. *Anal Chem.* 2010; 82(1):53–6. [PubMed: 19950968]
116. Ficarro SB, et al. Online nanoflow multi-dimensional fractionation for high efficiency phosphopeptide analysis. *Molecular & Cellular Proteomics.* 2011

117. Payne SH, et al. Phosphorylation-specific MS/MS scoring for rapid and accurate phosphoproteome analysis. *Journal of Proteome Research*. 2008; 7(8):3373–3381. [PubMed: 18563926]
118. Beausoleil SA, et al. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nature Biotechnology*. 2006; 24(10):1285–1292.
119. Savitski MM, et al. Confident Phosphorylation Site Localization Using the Mascot Delta Score. *Molecular & Cellular Proteomics*. 2011; 10(2):M110.003830. [PubMed: 21057138]
120. Choe L, et al. 8-plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment for Alzheimer's disease. *Proteomics*. 2007; 7(20):3651–60. [PubMed: 17880003]
121. Phanstiel D, et al. Peptide quantification using 8-plex isobaric tags and electron transfer dissociation tandem mass spectrometry. *Anal Chem*. 2009; 81(4):1693–8. [PubMed: 19154110]
122. Wolf-Yadlin A, et al. Effects of HER2 overexpression on cell signaling networks governing proliferation and migration. *Mol Syst Biol*. 2006; 2:54. [PubMed: 17016520]
123. Tang LY, et al. Quantitative phosphoproteome profiling of Wnt3a-mediated signaling network: indicating the involvement of ribonucleoside-diphosphate reductase M2 subunit phosphorylation at residue serine 20 in canonical Wnt signal transduction. *Mol Cell Proteomics*. 2007; 6(11):1952–67. [PubMed: 17693683]
124. Ashburner M, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000; 25(1):25–9. [PubMed: 10802651]
125. Kanehisa M, et al. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res*. 2006; 34(Database issue):D354–7. [PubMed: 16381885]
126. BioCarta Pathways. Available from: <http://www.biocarta.com/genes/allPathways.asp>
127. Stark C, et al. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res*. 2006; 34(Database issue):D535–9. [PubMed: 16381927]
128. Keshava Prasad TS, et al. Human Protein Reference Database--2009 update. *Nucleic Acids Res*. 2009; 37(Database issue):D767–72. [PubMed: 18988627]
129. Zhou F, et al. Online nanoflow RP-RP-MS reveals dynamics of multicomponent Ku complex in response to DNA damage. *J Proteome Res*. 2010; 9(12):6242–55. [PubMed: 20873769]
130. Ficarro SB, et al. Improved electrospray ionization efficiency compensates for diminished chromatographic resolution and enables proteomics analysis of tyrosine signaling in embryonic stem cells. *Anal Chem*. 2009; 81(9):3440–7. [PubMed: 19331382]
131. Askenazi M, et al. Pathway Palette: a rich internet application for peptide-, protein- and network-oriented analysis of MS data. *Proteomics*. 2010; 10(9):1880–5. [PubMed: 20198642]
132. Dennis G Jr, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol*. 2003; 4(5):3.
133. Chang JT, Nevins JR. GATHER: a systems approach to interpreting genomic signatures. *Bioinformatics*. 2006; 22(23):2926–33. [PubMed: 17000751]
134. Dumont JE, Pecasse F, Maenhaut C. Crosstalk and specificity in signalling. Are we crosstalking ourselves into general confusion? *Cell Signal*. 2001; 13(7):457–63. [PubMed: 11516620]
135. Force T, et al. Molecular scaffolds regulate bidirectional crosstalk between Wnt and classical seven-transmembrane-domain receptor signaling pathways. *Sci STKE*. 2007; 2007(397):pe41. [PubMed: 17666710]
136. Javelaud D, Mauviel A. Crosstalk mechanisms between the mitogen-activated protein kinase pathways and Smad signaling downstream of TGF-beta: implications for carcinogenesis. *Oncogene*. 2005; 24(37):5742–50. [PubMed: 16123807]
137. Matozaki T, Nakanishi H, Takai Y. Small G-protein networks: their crosstalk and signal cascades. *Cell Signal*. 2000; 12(8):515–24. [PubMed: 11027944]
138. Breitkreutz A, et al. A global protein kinase and phosphatase interaction network in yeast. *Science*. 2010; 328(5981):1043–6. [PubMed: 20489023]
139. Overgaard AJ, et al. Quantitative iTRAQ-Based Proteomic Identification of Candidate Biomarkers for Diabetic Nephropathy in Plasma of Type 1 Diabetic Patients. *Clin Proteomics*. 2010; 6(4):105–114. [PubMed: 21124997]

140. Huang SS, Fraenkel E. Integrating proteomic, transcriptional, and interactome data reveals hidden components of signaling and regulatory networks. *Sci Signal*. 2009; 2(81):ra40. [PubMed: 19638617]
141. Kumar N, et al. Modeling HER2 effects on cell behavior from mass spectrometry phosphotyrosine data. *PLoS Comput Biol*. 2007; 3(1):e4. [PubMed: 17206861]
142. Woolf PJ, et al. Bayesian analysis of signaling networks governing embryonic stem cell fate decisions. *Bioinformatics*. 2005; 21(6):741–53. [PubMed: 15479714]
143. Van Hoof D, et al. Phosphorylation dynamics during early differentiation of human embryonic stem cells. *Cell Stem Cell*. 2009; 5(2):214–26. [PubMed: 19664995]
144. Linding R, et al. Systematic discovery of in vivo phosphorylation networks. *Cell*. 2007; 129(7):1415–26. [PubMed: 17570479]
145. Szklarczyk D, et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res*. 2011; 39(Database issue):D561–8. [PubMed: 21045058]
146. Blom N, et al. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics*. 2004; 4(6):1633–49. [PubMed: 15174133]
147. Obenaus JC, Cantley LC, Yaffe MB. Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res*. 2003; 31(13):3635–41. [PubMed: 12824383]
148. Washburn MP, Wolters D, Yates JR. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology*. 2001; 19(3):242.
149. Stennicke HR, Salvesen GS. Caspases - controlling intracellular signals by protease zymogen activation. *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology*. 2000; 1477(1–2):299–306.
150. Adams JA. Activation loop phosphorylation and catalysis in protein kinases: Is there functional evidence for the autoinhibitor model? *Biochemistry*. 2003; 42(3):601–607. [PubMed: 12534271]
151. Patricelli MP, et al. Functional Interrogation of the Kinome Using Nucleotide Acyl Phosphates. *Biochemistry*. 2006; 46(2):350–358. [PubMed: 17209545]
152. Liu Y, Patricelli MP, Cravatt BF. Activity-based protein profiling: The serine hydrolases. *Proceedings of the National Academy of Sciences of the United States of America*. 1999; 96(26):14694–14699. [PubMed: 10611275]
153. Greenbaum D, et al. Epoxide electrophiles as activity-dependent cysteine protease profiling and discovery tools. *Chemistry & Biology*. 2000; 7(8):569–581. [PubMed: 11048948]
154. Shreder KR, et al. Design and synthesis of AX7574: A microcystin-derived, fluorescent probe for serine/threonine phosphatases. *Bioconjugate Chemistry*. 2004; 15(4):790–798. [PubMed: 15264866]
155. Vocadlo DJ, Bertozzi CR. A strategy for functional proteomic analysis of glycosidase activity from cell lysates. *Angewandte Chemie-International Edition*. 2004; 43(40):5338–5342.
156. Cravatt BF, Wright AT, Kozarich JW. Activity-based protein profiling: From enzyme chemistry. *Annual Review of Biochemistry*. 2008; 77:383–414.
157. Speers AE, Cravatt BF. Profiling Enzyme Activities In Vivo Using Click Chemistry Methods. *Chemistry & Biology*. 2004; 11(4):535–546. [PubMed: 15123248]
158. Kolb HC, Finn MG, Sharpless KB. Click chemistry: Diverse chemical function from a few good reactions. *Angewandte Chemie-International Edition*. 2001; 40(11):2004–2021.
159. Huisgen R. 1,3-Dipolar Cycloadditions. Past and Future. *Angewandte Chemie International Edition in English*. 1963; 2(10):565–598.
160. Rostovtsev VV, et al. A Stepwise Huisgen Cycloaddition Process: Copper(I)-Catalyzed Regioselective “Ligation” of Azides and Terminal Alkynes. *Angewandte Chemie International Edition*. 2002; 41(14):2596–2599.
161. Tornøe CW, Christensen C, Meldal M. Peptidotriazoles on solid phase: 1,2,3 -triazoles by regioselective copper(I)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *Journal of Organic Chemistry*. 2002; 67(9):3057–3064. [PubMed: 11975567]

162. Chan TR, et al. Polytriazoles as Copper(I)-Stabilizing Ligands in Catalysis. *Organic Letters*. 2004; 6(17):2853–2855. [PubMed: 15330631]
163. Baskin JM, et al. Copper-free click chemistry for dynamic in vivo imaging. *Proceedings of the National Academy of Sciences*. 2007; 104(43):16793–16797.
164. Chang PV, et al. Copper-free click chemistry in living animals. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107(5):1821–1826. [PubMed: 20080615]
165. Laughlin ST, et al. In vivo imaging of membrane-associated glycans in developing zebrafish. *Science*. 2008; 320(5876):664–667. [PubMed: 18451302]
166. Weerapana E, Speers AE, Cravatt BF. Tandem orthogonal proteolysis-activity-based protein profiling (TOP-ABPP)—a general method for mapping sites of probe modification in proteomes. *Nat Protocols*. 2007; 2(6):1414–1425.
167. Nomura DK, Dix MM, Cravatt BF. Activity-based protein profiling for biochemical pathway discovery in cancer. *Nature Reviews Cancer*. 2010; 10(9):630–638.
168. Shields DJ, et al. RBBP9: A tumor-associated serine hydrolase activity required for pancreatic neoplasia. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107(5):2189–2194. [PubMed: 20080647]
169. Jessani N, et al. Enzyme activity profiles of the secreted and membrane proteome that depict cancer cell invasiveness. *Proceedings of the National Academy of Sciences of the United States of America*. 2002; 99(16):10335–10340. [PubMed: 12149457]
170. Jessani N, et al. Carcinoma and stromal enzyme activity profiles associated with breast tumor growth in vivo. *Proceedings of the National Academy of Sciences of the United States of America*. 2004; 101(38):13756–13761. [PubMed: 15356343]
171. Nomura DK, et al. Monoacylglycerol Lipase Regulates a Fatty Acid Network that Promotes Cancer Pathogenesis. *Cell*. 2010; 140(1):49–61. [PubMed: 20079333]
172. Boja E, et al. Evolution of Clinical Proteomics and its Role in Medicine. *Journal of Proteome Research*. 2010; 10(1):66–84. [PubMed: 20715861]
173. Sturgeon C. Perspectives in Clinical Proteomics Conference: translating clinical proteomics into clinical practice. *Expert Review of Proteomics*. 2010; 7(4):469–471. [PubMed: 20653503]
174. Apweiler R, et al. Approaching clinical proteomics: current state and future fields of application in fluid proteomics. *Clinical Chemistry and Laboratory Medicine*. 2009; 47(6):724–744. [PubMed: 19527139]
175. Addona TA, et al. Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat Biotechnol*. 2009; 27(7):633–41. [PubMed: 19561596]

Figure 1. LTQ-Orbitrap Mass Spectra for De Novo Sequence Assignment

(A) Example MS spectrum with an m/z window of 300 – 2000 Da. The inset shows the expansion of a 1.5 Da m/z window around the precursor at 785.84 Da. Isotope ions of the precursor are separated by 0.5 Da, indicating a 2+ charge state. (B) MS/MS spectrum for the precursor at m/z = 785.84. The relative intensities of all b- and y-type fragment ions are normalized to the most abundant fragment at m/z = 480.

Figure 2. Peptide Fragmentation Products and Mechanism

(A) Molecular structure of a peptide, with amino acid side chains represented as R_1 , R_2 , etc. The fragmentation sites for common ion pairs (a/x, b/y and c/z) are indicated with dashed red lines. The b/y pair is the most commonly observed in CAD mass spectrometry. (B) Mechanism for generation of complimentary b- and y- type ions resulting from fragmentation of a doubly-charged tryptic peptide. Protonation of a backbone amide nitrogen leads to lysis of the amide bond and formation of a cyclized oxazolone ion structure, along with the corresponding y-type ion. The oxazolone ring on the b ion can undergo electron rearrangement, leading to loss of a carbonyl group and formation of an a-type fragment ion. *The illustrations in this figure were generated using ACD/Chemsketch Freeware, version 12.01, Advanced Chemistry Development, Inc., Toronto, ON, Canada, www.acdlabs.com, 2010.

Figure 3. Fragment ion Assignment in an MS/MS Spectrum

The MS/MS spectrum for the 785.84 Da, 2+ precursor in Fig. 1B is shown with the b- and y-type ions assigned. In addition, the b – H₂O ion series is also labeled. The fragment ion masses for each series are shown in the data table above the spectrum. The b and b – H₂O ion series always contain the N-terminus and hence progress from the N- to C-terminus. In contrast the y-type ions contain the C-terminus and progress in the opposite direction. For clarity the b₆ and b₉ ion labels have been omitted from the spectrum due to their proximity to the y₅ and y₈ ions, respectively, though they are present and their masses are shown in the data table.

Figure 4. Enrichment of Biological Annotations

An example enrichment analysis displays signaling pathways from the KEGG database whose members are significantly enriched within an experimental dataset. Top: Enrichment analysis involves quantifying the overlap between experimentally determined protein lists (green circles) and protein sets for a particular biological theme extracted from databases such as KEGG (blue circles). The significance of enrichment, computed as a p-value based on a Fisher's Exact Test, accounts for the degree of overlap between an empirically-determined protein list and the protein set corresponding to a particular KEGG pathway, in addition to the size of the input protein list, KEGG pathway protein set, and the total proteome. Bottom: In this example, KEGG signaling pathways that score as "enriched" (p-value < 0.05) are displayed as a bar graph.

Figure 5. Networks

Networks are often used to visualize data where nodes represent experimentally determined proteins and edges represent biochemical relationships extracted from publicly accessible databases. (A) The most common approach is to connect proteins using undirected protein-protein interaction edges (in gray) and to (B) color the nodes based on quantitative data (*e.g.*, “upregulated or downregulated”). In this example, red and blue nodes indicate increased and reduced levels of phosphorylation, respectively. (C) Directed networks are often more appropriate, especially when representing enzymatic and other transient relationships. Here, edges terminated with purple arrows and brown circles represent kinase- and phosphatase- substrate relationships, respectively. Networks lend themselves to

application of graph theoretical methods such as the (D) Steiner Tree algorithm, which introduces undetected nodes (yellow) in order to connect all experimental nodes using a minimal number of edges. Integration of multiple edge types and graph theory algorithms can provide a more comprehensive description of signaling pathways.

Figure 6. Activity-based Protein Profiling – ABPP

(A) Essential components of an activity-based probe include a reactive group (RG) that targets a particular enzyme class as well as a reporter group (Rp) that allows for isolation and/or detection of labeled molecules. (B) The copper-catalyzed click chemistry reaction between an azide and an alkyne can be used to couple reactive probes to reporter functionalities in lysates after labeling *in vivo*.

Figure 7. Mass Spectrometry-based Strategies to Identify Proteins Captured by Activity Probes Affinity tags (*e.g.*, biotin) are typically required to facilitate enrichment of proteins targeted by activity-based probes. Depending on the analytical strategy, MS interrogation can be used to (A) sequence proteolytic peptides from labeled proteins, (B) detect probe modified peptides, or (C) detect both in separate LC-MS/MS analyses.

Naturally Occurring Amino Acids

Table 1

Each entry lists full name, as well as one- and three-letter abbreviations. The molecular formulae are based on the peptide backbone elemental composition, C_2H_2NO , in all cases except for Proline (P – Pro), which is one proton deficient, and has the formula C_2HNO . The monoisotopic mass is calculated using the mass of the most abundant isotope for each atom in the molecular formula of an amino acid and the average mass is calculated using the abundance-weighted average of each atom in the molecular formula. Immonium ions, whose masses are shown in the far-right column, have a backbone molecular formula of CH_3N (or CH_2N for Proline) due to the gain of a proton and loss of carbonyl, for a net loss of 27 Da. References for the mass data are listed below:

Amino Acid	Symbols	Molecular Formula	Monoisotopic Mass*	Average Mass#	Immonium Ion Mass†
Glycine	G - Gly	C_2H_3NO	57.02	57.05	30
Alanine	A - Ala	C_3H_5NO	71.04	71.08	44
Serine	S - Ser	$C_3H_5NO_2$	87.03	87.08	60
Proline	P - Pro	C_5H_7NO	97.05	97.12	70
Valine	V - Val	C_5H_9NO	99.07	99.13	72
Threonine	T - Thr	$C_4H_7NO_2$	101.05	101.10	74
Cysteine	C - Cys	C_3H_5NOS	103.01	103.14	76
Leucine	L - Leu	$C_6H_{11}NO$	113.08	113.16	86
Isoleucine	I - Ile	$C_6H_{11}NO$	113.08	113.16	86
Asparagine	N - Asn	$C_4H_6N_2O_2$	114.04	114.10	87
Aspartic Acid	D - Asp	$C_4H_5NO_3$	115.03	115.09	88
Glutamine	Q - Gln	$C_5H_8N_2O_2$	128.06	128.13	101
Lysine	K - Lys	$C_6H_{12}N_2O$	128.09	128.17	101
Glutamic Acid	E - Glu	$C_5H_7NO_3$	129.04	129.11	102
Methionine	M - Met	C_5H_9NOS	131.04	131.20	104
Histidine	H - His	$C_6H_7N_3O$	137.06	137.14	110
Phenylalanine	F - Phe	C_9H_9NO	147.07	147.17	120
Arginine	R - Arg	$C_6H_{12}N_4O$	156.10	156.19	129
Tyrosine	Y - Tyr	$C_9H_9NO_2$	163.06	163.17	136
Tryptophan	W - Trp	$C_{11}H_{10}N_2O$	186.08	186.21	159

* Monoisotopic masses obtained from *The AME2003 atomic mass evaluation (II). Tables, graphs, and references*, Nucl. Phys. A, Vol. 729, pp. 337–676 (2003).

Average masses of the atoms obtained from *Atomic Weights of the Elements 2007 (IUPAC Technical Report)*, Pure Appl. Chem., Vol. 81, No. 11, pp. 2131–2156 (2009).

‡ Integer values shown, adapted from *The Encyclopedia of Mass Spectrometry, Volume 2: Biological Applications: Part A: Peptides and Proteins*, eds. Richard M. Caprioli and Michael L. Gross (2005), pp. 173–174.