

Published in final edited form as:

Curr Opin Biotechnol. 2007 August ; 18(4): 378–384.

Advances in proteomic workflows for systems biology

Johan Malmström^{#1}, Hookeun Lee^{#1}, and Ruedi Aebersold^{#1,#2,*}

^{#1}*Institute for Molecular Systems Biology, ETH Zürich, Switzerland* ^{#2}*Faculty of Science University of Zurich, Switzerland and Institute for Systems Biology, Seattle, WA*

Summary and recent advances

Mass spectrometry, specifically the analysis of complex peptide mixtures by liquid chromatography and tandem mass spectrometry (shotgun proteomics) has been at the center of proteomics research for the last decade. To overcome some of the fundamental limitations of the approach, including its limited sensitivity and high degree of redundancy, new proteomics workflows are being developed. Among these, targeting methods in which specific peptides are selectively isolated, identified and quantified are particularly promising. Here we summarize recent incremental advances in shotgun proteomics methods and outline emerging targeted workflows. The development of the target driven approaches with their ability to detect and quantify identical, non-redundant sets of proteins in multiple repeat analyses will be critically important for the application of proteomics to biomarker discovery and validation, and to systems biology research.

Incremental improvements of non targeted mass spectrometry based proteomics

For the last few years shotgun tandem mass spectrometry has been the most popular and widely used method in proteomics. In this method, complex protein mixtures are digested to peptides, usually using trypsin as the protease, and the resulting peptides are fractionated by one, two or three dimensional separation and analyzed by tandem mass spectrometry [1,2]. Optionally, stable isotope signatures are introduced into proteins or peptides to allow quantitatively accurate comparisons of samples [3–6]. Over the last years incremental improvements have increased the reproducibility of peptide separation, the speed and accuracy of data collection and the confidence of inferring the sequence of peptides and proteins from the fragment ion spectra. Figure 1 illustrates the general workflow and indicates significant recent technical advances that are further described in the following sections.

Advances in sample preparation

To improve on the resolution achievable by the classical two-dimensional (cation exchange/reversed-phase) chromatography peptide separation, iso-electric focusing techniques in gels and in solution have been described [7–10]. Since the *pI* of peptides can be accurately calculated from the amino acid sequence of a peptide, the *pI* information obtained by such experiments has also proven beneficial for the correct assignment of fragment ion spectra to peptide sequences (see below). It can be expected that with the development of instruments supporting robust [7,10] and preferentially multiplexed peptide IEF separations [8] these methods will gain in importance in proteomics research. The development of highly reproducible capillary

*Correspondence: Ruedi Aebersold. Email: aebersold@imsb.biol.ethz.ch

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

chromatography methods using particle or monolithic columns [11,12] and columns manufactured on chips [13] have also greatly increased the data quality, especially in cases in which multiple repeat analyses are being performed in a single study.

The development of strategies for quantitative analysis by stable isotope labeling has accelerated over the last few years. Multiple chemical, enzymatic or biosynthetic labeling strategies have been described that support the collection of quantitative protein data on a large scale [3–6]. In addition to measuring changes in protein abundance in protein samples the same quantitative methods or variations thereof are increasingly being used to distinguish groups of proteins of interest from background proteins, e.g. for the analysis of protein complexes [14–16] or organelles [17–20]

Advances in data acquisition

The quality of the spectra acquired and the pace of data acquisition in shotgun proteomics experiments determine the depth at which a proteome can be analyzed and the level of confidence with which fragment ion spectra can be assigned to peptide sequences. Over the last few years different types of mass spectrometers providing high mass resolution and accuracy have been developed. These include the TOF-TOF [21] Q-TOF [22] FT-ICR mass spectrometers [23–25] and orbitrap mass spectrometers [26] and the performance of these instruments in proteomics research was recently reviewed by Domon and Aebersold 2006 [27]. The impact of these high performance mass spectrometers was documented in a series of manuscripts in which a variety of samples, including the urinary and tear fluid proteome, were analyzed in great depth [28–30]. and in which these results were discussed [30]. Furthermore, it has been documented that the ionization methods electrospray ionization (ESI) and matrix assisted laser desorption (MALDI) ionize different but overlapping segments of the observable proteome [31–33]. Therefore, the combination of different types of mass analyzers and ion sources have contributed to more extensive proteome coverage compared to that achievable by a single method.

Advances in data processing and analysis

Sequence database search tools such as Sequest [34] and Mascot [35] that assign a peptide sequence to a fragment ion spectrum have been commercially available and in wide use for more than a decade. In recent years a number of open source database search tools have also become available [36–39]. In addition, tools and strategies have been developed to assess the reliability of each peptide identification and for the assessment of the overall false positive or false negative error rates in large datasets. These include probabilistic models such as PeptideProphet [40] and the searching of reverse or scrambled sequence databases to estimate the false positive rate in an experiment [41–43]. Increasingly, proteomics journals require detailed documentation of the applied search strategies and error rate estimates to assure the quality of published proteomics data [44]. Tools and methods to estimate the false positive rate of peptide identification and tools to infer protein identifications from the list of correctly identified peptides [45–48] are therefore expected to find widespread application.

Collectively the incremental improvements of the shotgun approaches listed above have increased the speed of the data collection and ensured the collection of high confidence data, allowing increasingly more comprehensive analysis to be undertaken. A main goal of proteomics, the identification of complete proteomes has, however, remained unachieved.

Principle limitations of shotgun approaches

Despite the recent advances discussed above and the success of the current state of the art shotgun approach a number of principal limitations have become apparent that seem difficult to overcome with incremental technical improvements. These include the following:

Extreme redundancy of shotgun MS/MS datasets

The detailed analysis of single large datasets or aggregates of datasets generated from substantially similar samples have shown that the most highly expressed proteins are identified multiple (up to thousands) of times at the cost of identifying proteins of interest expressed at lower levels of abundance [49,50]. While the improved separation methods described above alleviate the problem, they may not be, by themselves, sufficient to completely overcome it, due to the extreme complexity of the proteome.

Under sampling

As a consequence of data acquisition redundancy and sample complexity, the peptides present in proteomic samples are usually under sampled, i.e. only a fraction of the peptides detectable by a mass spectrometer are in fact identified. Repeat analyses of the same or similar biological samples therefore show limited overlap of identified proteins making sample to sample comparison difficult.

Sample complexity

The presence of the large number of semi-tryptic or non-tryptic peptides in samples generated for proteomic analyses (Picotti et al submitted) greatly increases the complexity of the samples, further complicating the complete analysis of proteomes.

Saturation

The analysis of very large data sets generated by combining the results of multiple studies on the same organism, e.g. yeast [51], have shown a saturation effect in which the discovery rate of new proteins per fragment ion spectrum added to the data set decreases. As the level of saturation is well below the complete proteome, the discovery of new proteins beyond the level of saturation will require the use of improved or orthogonal separation schema prior to mass spectrometry. Of course, additional fractionation steps increase the time consumed for a single proteomic study.

Based on these fundamental limitations of the shotgun approach, it appears that the in spite of recent incremental improvements the goal of rapid, complete and quantitative proteome analysis will remain elusive.

Targeted driven quantitative proteomics

The sequencing of the human genome and the resulting (almost complete) genome and transcript maps have catalyzed the development of high throughput assays for probing transcript expression patterns, protein:DNA interactions and other genomic features. Similar developments are also emerging in proteomics where the collection of many (thousands) of data sets have led collectively to extensive proteome maps of a number of species and sample types [49]. As in genomics, such maps then form the basis for the development of high throughput proteomic assays in which defined sub proteomes and eventually whole proteomes are quickly and robustly identified and quantified by targeted mass spectrometric analyses [52,53]. In these targeted approaches the mass spectrometer is trained on the analysis of proteotypic peptides [53] i.e. those peptides that are observable by mass spectrometry and uniquely identify a protein, thus largely eliminating the redundancy inherent in shotgun

proteomics. Such targeted approaches are becoming possible due to advances in instrumentation and bioinformatics and the availability of extensive and reliable proteome maps such as PeptideAtlas. An example of the steps required for a targeted proteomics is outlined in Figure 2 and the individual steps are discussed below.

Protein and peptide selection for targeted approaches

A targeted proteomics experiment starts from a list of proteins of interest. From these proteins the target peptides need to be determined. Since there are large differences in the propensity of peptides to be identified by mass spectrometry, careful selection of the target peptides is necessary. Target peptides should be preferentially observable and uniquely identify one protein or protein isoform and should ideally not contain amino acids which are easily modified such as methionine. Complications in the selection of suitable target peptides are illustrated in Figure 3 where the association of experimentally identified peptides with their parent protein (s) is shown. There are two ways to select proteotypic target peptides, experimental and computational.

Experimental—Upon assembling aggregates of data sets and aligning the peptides with high probability of being correct to the genomic sequence, an atlas of the detectable and expressed proteome could be assembled into a Peptide Atlas [49]. As of today several of these atlases exist from the most common model organisms and more are being produced, and several additional repositories for proteomics data and other databases resources have recently been described (see Table 1 for examples) [51,54–58]. This data can be mined to select proteotypic peptides of the targeted proteins, along with the representative fragment ion spectra and chromatographic coordinates.

Computational—The above described resources are used to derive empirically observed proteotypic peptides for targeted proteomics. However, such resources are at present not available for all species with sequenced genomes nor are they available for all proteins in genomes where peptide collection repositories exist. Therefore computational approaches have been developed to select proteotypic peptides from proteins that have not been previously observed [33,59]. Lu et al described an approach that involves the training of a classifier to estimate the number of unique peptides expected from a given protein and to incorporate information on the repeated sampling of spectra from each protein in a shotgun proteomics experiment [59,60]. Mallick and colleagues empirically identified 16,000 proteotypic peptides for more than 4000 yeast proteins starting from 600,000 peptide identifications [33]. Characteristic properties of these peptides were then used to develop a computational tool that can predict proteotypic peptides for any given organism and instrument platform.

Multiple reaction monitoring (MRM) and data analysis

MRM is a highly selective, highly sensitive tandem mass spectrometry method carried out on triple quadrupole type mass spectrometers that has been used extensively, e.g. for drug metabolism studies. Increasingly, the method is now also being applied to protein analysis and proteomics [61]. This trend has been accelerated by recent instrumentation developments such as the hybrid quadrupole-linear ion trap mass spectrometer [62]. Knowing the mass and structure of the analyte molecule, it is possible to predict the precursor m/z and fragments m/z (MRM transition). Each targeted peptide has a set of accompanying transitions which are then selectively detected in the second stage of the mass spectrometer [61,63]. The emerging targeted proteomics workflow thus consists of the selection of proteins to be assayed, the selection of proteotypic peptides from these proteins, the selection of suitable transitions and their measurement. Using the expected elution times of the target peptides as constraints in the MRM experiment, we have shown that in a single LC-MS/MS experiment in excess of 600 transitions can be measured. Since in this approach the mass spectrometer only detects the

proteins of interest it can be successfully applied even if the proteins of interest are present within a large background of other proteins [64]. The sensitivity of this approach can be further increased if it is combined with the enrichment of the target peptides using affinity reagent in a method referred to as SISCAPA [65].

Conclusion

Recent incremental improvements of shotgun mass spectrometry based proteomics technology have resulted in increased speed and quality of data collection. However, in spite of these advances a number of principle limitations remain. Therefore, new proteomics approaches are being developed that include, as a very promising approach, targeted mass spectrometry, where selected targets are specifically identified and quantified. This emerging targeting workflow consists of the selection of target proteins, the selection of proteotypic peptides that uniquely represent the target protein and the identification of fragment ions that are used to score the abundance level of the target proteins. The ability of the targeting approaches to identify and quantify non-redundant completely overlapping data sets is expected to have a strong impact on biomarker discovery and validation and systems biology research

Acknowledgements

We would like to thank Luis Mendoza for help with the Figures. This work was supported by SNF grant (nr 3100A0-107679/1) and with Federal (US) Funds from the National Heart, Lung and Blood institute, national institute of health, under contract #N01-HV-28179. J.M is supported by a fellowship from the Swedish society for medical research (SSMF).

References

1. Delahunty C, Yates JR 3rd. Protein identification using 2D-LC-MS/MS. *Methods* 2005;35:248–255. [PubMed: 15722221]
2. Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature* 2003;422:198–207. [PubMed: 12634793]
3. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 1999;17:994–999. [PubMed: 10504701]
4. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, et al. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* 2004;3:1154–1169. [PubMed: 15385600]
5. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* 2002;1:376–386. [PubMed: 12118079]
6. Schmidt A, Kellermann J, Lottspeich F. A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics* 2005;5:4–15. [PubMed: 15602776]
7. Xie H, Bandhakavi S, Griffin TJ. Evaluating preparative isoelectric focusing of complex peptide mixtures for tandem mass spectrometry-based proteomics: a case study in profiling chromatin-enriched subcellular fractions in *Saccharomyces cerevisiae*. *Anal Chem* 2005;77:3198–3207. [PubMed: 15889909]
- 8*. Heller M, Ye M, Michel PE, Morier P, Stalder D, Junger MA, Aebersold R, Reymond F, Rossier JS. Added value for tandem mass spectrometry shotgun proteomics data validation through isoelectric focusing of peptides. *J Proteome Res* 2005;4:2273–2282. [PubMed: 16335976]
9. Cargile BJ, Bundy JL, Freeman TW, Stephenson JL Jr. Gel based isoelectric focusing of peptides and the utility of isoelectric point in protein identification. *J Proteome Res* 2004;3:112–119. [PubMed: 14998171]
- 10*. Malmstrom J, Lee H, Nesvizhskii AI, Shteynberg D, Mohanty S, Brunner E, Ye M, Weber G, Eckerskorn C, Aebersold R. Optimized peptide separation and identification for mass spectrometry

- based proteomics via free-flow electrophoresis. *J Proteome Res* 2006;5:2241–2249. [PubMed: 16944936]
11. Shen Y, Smith RD, Unger KK, Kumar D, Lubda D. Ultrahigh-throughput proteomics using fast RPLC separations with ESI-MS/MS. *Anal Chem* 2005;77:6692–6701. [PubMed: 16223258]
 12. Premstaller A, Oberacher H, Walcher W, Timperio AM, Zolla L, Chervet JP, Cavusoglu N, van Dorsselaer A, Huber CG. High-performance liquid chromatography-electrospray ionization mass spectrometry using monolithic capillary columns for proteomic studies. *Anal Chem* 2001;73:2390–2396. [PubMed: 11403277]
 13. Yin H, Killeen K, Brennen R, Sobek D, Werlich M, van de Goor T. Microfluidic chip for peptide analysis with an integrated HPLC column, sample enrichment column, and nanoelectrospray tip. *Anal Chem* 2005;77:527–533. [PubMed: 15649049]
 14. Ranish JA, Yi EC, Leslie DM, Purvine SO, Goodlett DR, Eng J, Aebersold R. The study of macromolecular complexes by quantitative proteomics. *Nat Genet* 2003;33:349–355. [PubMed: 12590263]
 15. Selbach M, Mann M. Protein interaction screening by quantitative immunoprecipitation combined with knockdown (QUICK). *Nat Methods* 2006;3:981–983. [PubMed: 17072306]
 16. Blagoev B, Kratchmarova I, Ong SE, Nielsen M, Foster LJ, Mann M. A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling. *Nat Biotechnol* 2003;21:315–318. [PubMed: 12577067]
 17. Andersen JS, Lyon CE, Fox AH, Leung AK, Lam YW, Steen H, Mann M, Lamond AI. Directed proteomic analysis of the human nucleolus. *Curr Biol* 2002;12:1. [PubMed: 11790298]
 18. Borner GH, Harbour M, Hester S, Lilley KS, Robinson MS. Comparative proteomics of clathrin-coated vesicles. *J Cell Biol* 2006;175:571–578. [PubMed: 17116749]
 19. Marelli M, Smith JJ, Jung S, Yi E, Nesvizhskii AI, Christmas RH, Saleem RA, Tam YY, Fagarasanu A, Goodlett DR, et al. Quantitative mass spectrometry reveals a role for the GTPase Rho1p in actin organization on the peroxisome membrane. *J Cell Biol* 2004;167:1099–1112. [PubMed: 15596542]
 - 20*. Dunkley TP, Hester S, Shadforth IP, Runions J, Weimar T, Hanton SL, Griffin JL, Bessant C, Brandizzi F, Hawes C, et al. Mapping the Arabidopsis organelle proteome. *Proc Natl Acad Sci U S A* 2006;103:6518–6523. [PubMed: 16618929]
 21. Medzihradsky KF, Campbell JM, Baldwin MA, Falick AM, Juhasz P, Vestal ML, Burlingame AL. The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. *Anal Chem* 2000;72:552. [PubMed: 10695141]
 22. Morris HR, Paxton T, Dell A, Langhorne J, Berg M, Bordoli RS, Hoyes J, Bateman RH. High sensitivity collisionally-activated decomposition tandem mass spectrometry on a novel quadrupole/orthogonal-acceleration time-of-flight mass spectrometer. *Rapid Commun Mass Spectrom* 1996;10:889–896. [PubMed: 8777321]
 23. Marshall AG, Hendrickson CL, Jackson GS. Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom Rev* 1998;17:1. [PubMed: 9768511]
 24. Martin SE, Shabanowitz J, Hunt DF, Marto JA. Subfemtomole MS and MS/MS peptide sequence analysis using nano-HPLC micro-ESI fourier transform ion cyclotron resonance mass spectrometry. *Anal Chem* 2000;72:4266–4274. [PubMed: 11008759]
 25. Syka JE, Marto JA, Bai DL, Horning S, Senko MW, Schwartz JC, Ueberheide B, Garcia B, Busby S, Muratore T, 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;3:621–626. [PubMed: 15253445]
 26. Makarov A. Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal Chem* 2000;72:1156–1162. [PubMed: 10740853]
 - 27*. Domon B, Aebersold R. Mass spectrometry and protein analysis. *Science* 2006;312:212–217. [PubMed: 16614208]
 - 28*. Adachi J, Kumar C, Zhang Y, Olsen JV, Mann M. The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins. *Genome Biol* 2006;7:R80. [PubMed: 16948836]

- 29*. de Souza GA, Godoy LM, Mann M. Identification of 491 proteins in the tear fluid proteome reveals a large number of proteases and protease inhibitors. *Genome Biol* 2006;7:R72. [PubMed: 16901338]
30. Schmidt A, Aebersold R. High-accuracy proteome maps of human body fluids. *Genome Biol* 2006;7:242. [PubMed: 17140426]
31. Bodnar WM, Blackburn RK, Krise JM, Moseley MA. Exploiting the complementary nature of LC/MALDI/MS/MS and LC/ESI/MS/MS for increased proteome coverage. *J Am Soc Mass Spectrom* 2003;14:971–979. [PubMed: 12954165]
32. Elias JE, Haas W, Faherty BK, Gygi SP. Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nat Methods* 2005;2:667–675. [PubMed: 16118637]
- 33**. Mallick P, Schirle M, Chen SS, Flory MR, Lee H, Martin D, Ranish J, Raught B, Schmitt R, Werner T, et al. Computational prediction of proteotypic peptides for quantitative proteomics. *Nat Biotechnol* 2007;25:125–131. [PubMed: 17195840]
34. Eng J, McCormack AL, Yates JR. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* 1994;5:976–989.
35. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 1999;20:3551–3567. [PubMed: 10612281]
36. Craig R, Beavis RC. A method for reducing the time required to match protein sequences with tandem mass spectra. *Rapid Commun Mass Spectrom* 2003;17:2310–2316. [PubMed: 14558131]
37. Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM, Yang X, Shi W, Bryant SH. Open mass spectrometry search algorithm. *J Proteome Res* 2004;3:958–964. [PubMed: 15473683]
38. Zhang N, Li XJ, Ye M, Pan S, Schwikowski B, Aebersold R. ProBIDtree: an automated software program capable of identifying multiple peptides from a single collision-induced dissociation spectrum collected by a tandem mass spectrometer. *Proteomics* 2005;5:4096–4106. [PubMed: 16196091]
39. Zhang N, Aebersold R, Schwikowski B. ProBID: a probabilistic algorithm to identify peptides through sequence database searching using tandem mass spectral data. *Proteomics* 2002;2:1406–1412. [PubMed: 12422357]
40. Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* 2002;74:5383–5392. [PubMed: 12403597]
41. Moore RE, Young MK, Lee TD. Qscore: an algorithm for evaluating SEQUEST database search results. *J Am Soc Mass Spectrom* 2002;13:378–386. [PubMed: 11951976]
- 42*. Haas W, Faherty BK, Gerber SA, Elias JE, Beausoleil SA, Bakalarski CE, Li X, Villen J, Gygi SP. Optimization and use of peptide mass measurement accuracy in shotgun proteomics. *Mol Cell Proteomics* 2006;5:1326–1337. [PubMed: 16635985]
- 43*. Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* 2007;4:207–214. [PubMed: 17327847]
44. Carr S, Aebersold R, Baldwin M, Burlingame A, Clauser K, Nesvizhskii A. 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;3:531–533. [PubMed: 15075378]
45. Nesvizhskii AI, Aebersold R. Interpretation of shotgun proteomic data: the protein inference problem. *Mol Cell Proteomics* 2005;4:1419–1440. [PubMed: 16009968]
46. Nesvizhskii AI, Aebersold R. Analysis, statistical validation and dissemination of large-scale proteomics datasets generated by tandem MS. *Drug Discov Today* 2004;9:173–181. [PubMed: 14960397]
47. Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* 2003;75:4646–4658. [PubMed: 14632076]
- 48*. Keller A, Eng J, Zhang N, Aebersold R, Li XJ. A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Molecular Systems Biology* 2005;1:msb4100024-E4100021–msb4100024-E4100028.

- 49**. Desiere F, Deutsch EW, Nesvizhskii AI, Mallick P, King NL, Eng JK, Aderem A, Boyle R, Brunner E, Donohoe S, et al. Integration with the human genome of peptide sequences obtained by high-throughput mass spectrometry. *Genome Biol* 2005;6:R9. [PubMed: 15642101]
- 50**. Omenn GS, States DJ, Adamski M, Blackwell TW, Menon R, Hermjakob H, Apweiler R, Haab BB, Simpson RJ, Eddes JS, et al. Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *Proteomics* 2005;5:3226–3245. [PubMed: 16104056]
51. King NL, Deutsch EW, Ranish JA, Nesvizhskii AI, Eddes JS, Mallick P, Eng J, Desiere F, Flory M, Martin DB, et al. Analysis of the *Saccharomyces cerevisiae* proteome with PeptideAtlas. *Genome Biol* 2006;7:R106. [PubMed: 17101051]
52. Aebersold R. Constellations in a cellular universe. *Nature* 2003;422:115–116. [PubMed: 12634755]
53. Kuster B, Schirle M, Mallick P, Aebersold R. Innovation: Scoring proteomes with proteotypic peptide probes. *Nat Rev Mol Cell Biol*. 2005
54. Craig R, Cortens JP, Beavis RC. Open source system for analyzing, validating, and storing protein identification data. *J Proteome Res* 2004;3:1234–1242. [PubMed: 15595733]
- 55*. Marzolf B, Deutsch EW, Moss P, Campbell D, Johnson MH, Galitski T. SBEAMS-Microarray: database software supporting genomic expression analyses for systems biology. *BMC Bioinformatics* 2006;7:286. [PubMed: 16756676]
56. Jones P, Cote RG, Martens L, Quinn AF, Taylor CF, Derache W, Hermjakob H, Apweiler R. PRIDE: a public repository of protein and peptide identifications for the proteomics community. *Nucleic Acids Res* 2006;34:D659–663. [PubMed: 16381953]
57. Deutsch EW, Eng JK, Zhang H, King NL, Nesvizhskii AI, Lin B, Lee H, Yi EC, Ossola R, Aebersold R. Human Plasma PeptideAtlas. *Proteomics* 2005;5:3497–3500. [PubMed: 16052627]
58. Brunner E, Ahrens CH, Mohanty S, Beatschmann H, Loevnich S, Potthars F, Deutsch EW, deLichtenberg U, Rinner O, Lee H, et al. A high quality map of the *Drosophila melanogaster* proteome. Accepted in *Nature Biotechnology*. 2007
- 59**. Lu P, Vogel C, Wang R, Yao X, Marcotte EM. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol* 2007;25:117–124. [PubMed: 17187058]
- 60*. Bergeron JJ, Hallett M. Peptides you can count on. *Nat Biotechnol* 2007;25:61–62. [PubMed: 17211399]
61. Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci U S A* 2003;100:6940–6945. [PubMed: 12771378]
62. Hager JW, Yves Le Blanc JC. Product ion scanning using a Q-q-Q linear ion trap (Q TRAP) mass spectrometer. *Rapid Commun Mass Spectrom* 2003;17:1056–1064. [PubMed: 12720286]
- 63**. Mayya V, Rezual K, Wu L, Fong MB, Han DK. Absolute quantification of multisite phosphorylation by selective reaction monitoring mass spectrometry: determination of inhibitory phosphorylation status of cyclin-dependent kinases. *Mol Cell Proteomics* 2006;5:1146–1157. [PubMed: 16546994]
- 64**. Anderson L, Hunter CL. Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Mol Cell Proteomics* 2006;5:573–588. [PubMed: 16332733]
65. Anderson NL, Anderson NG, Haines LR, Hardie DB, Olafson RW, Pearson TW. Mass spectrometric quantitation of peptides and proteins using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA). *J Proteome Res* 2004;3:235–244. [PubMed: 15113099]
66. Desiere F, Deutsch EW, King NL, Nesvizhskii AI, Mallick P, Eng J, Chen S, Eddes J, Loevenich SN, Aebersold R. The PeptideAtlas project. *Nucleic Acids Res* 2006;34:D655–658. [PubMed: 16381952]

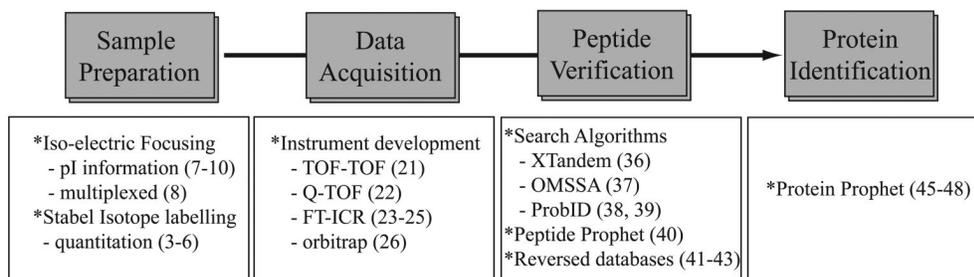
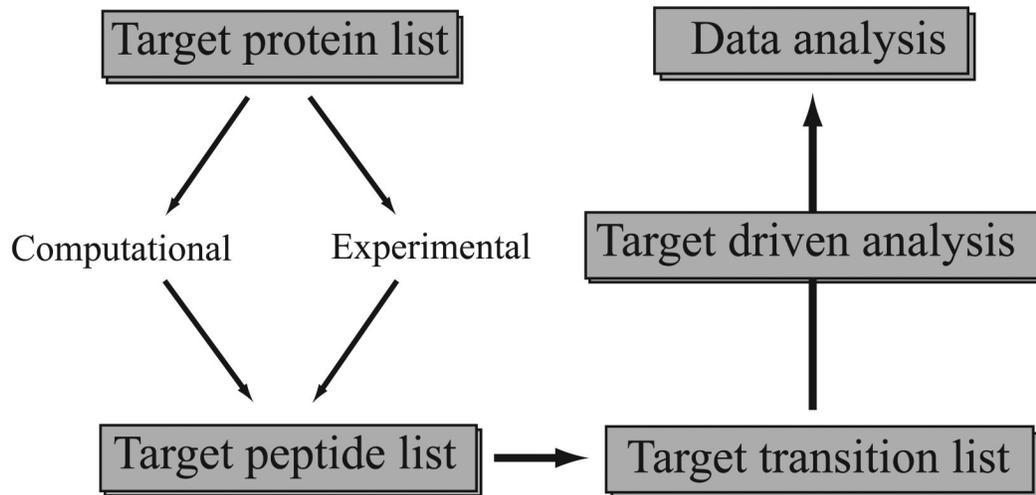


Figure 1. Shotgun proteomic and incremental improvements

In shotgun based mass spectrometry approaches the first step is to prepare the sample of interest by if necessary enriching for proteins of interest and introducing stable isotope labels, digest the proteins into peptides, to separate the sample into even parts suitable for mass spectrometry analysis and to perform necessary clean up steps. The second step involves the mass spectrometry analysis of the peptide fractions and the last two steps involve the data analysis, including peptide verification and protein identification. In each of these steps have there been incremental improvements listed in the white boxes.

**Figure 2. Target driven analysis**

Targeted driven approaches starts with a list of proteins. Once a list of candidate proteins have been assembled, either from the literature, from specific biological knowledge or from other sources, the next step is to select a subset of proteotypic peptides which can be used for scoring the abundance level of the proteins. From the peptide list specific fragments m/z (MRM transition) which can be used as product ions in the second stage of the mass spectrometer in the target driven analysis

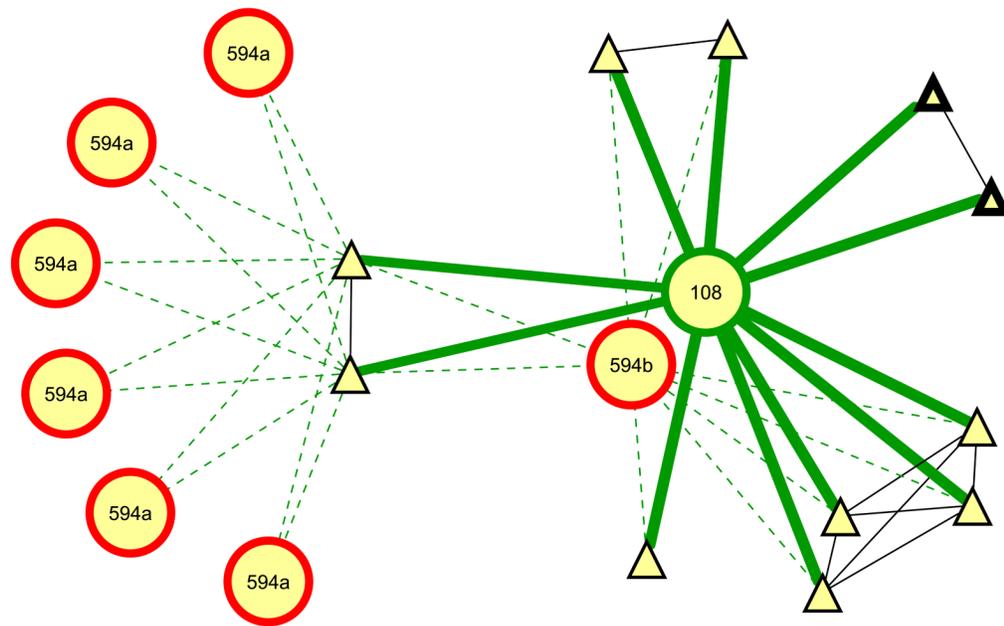


Figure 3. Example of peptide to protein ambiguity

Peptide nodes are represented by small triangles; those with thick borders map only to a single protein and are candidate peptides for a targeted approach. Protein nodes are represented by large circles and their border color is mapped to the assigned ProteinProphet probability, with weight=0.0 represented by dashed lines. Sequence-identical peptides (e.g. spectra that were identified to different charge states or modified versions of the same peptide sequence) are joined by thin black edges. Entry #587f is identified by 21 peptides (8 unique sequences) with high probabilities, and entry #163 is identified by one additional non-shared peptide. All peptide weights are thus set to 0.0 for entry #587f, resulting in protein probabilities of 0.0 and 1.0, respectively.

Table 1

Resource	Url	Reference
Peptide atlas	http://www.peptideatlas.org/	[66]
GPM	http://www.thegpm.org/	[54]
SBEAMS	http://www.sbeams.org/	[55]
PRIDE	http://www.ebi.ac.uk/pride/	[56]