

# Quantitative Proteomics Analysis

***Wenching Calvin Chan, Ph.D.***\*<sup>1</sup>

<sup>1</sup>Center for Research Informatics, University of Chicago, Chicago, IL 60637

\*wchan10@bsd.uchicago.edu

**24 January 2019**

## Abstract

- TBC

## Package

Quantitative Proteomics Analysis

## Contents

1	Disclaimer . . . . .	2
2	Background . . . . .	3
2.1	Chemistry . . . . .	3
2.2	Biochemistry . . . . .	4
3	Introduction . . . . .	5
3.1	Quantitative Approaches . . . . .	6
4	Computation . . . . .	19
4.1	Organization . . . . .	19
4.2	Data Format . . . . .	19
4.3	Data Repository . . . . .	20
4.4	Database Search . . . . .	23
5	Miscellaneous . . . . .	23
6	Session information . . . . .	24
	Reference . . . . .	25

## **1 Disclaimer**

---

The data were downloaded from NCBI's SRA or from EBI's ArrayExpress. See the following disclaimers for copyright / license policies.

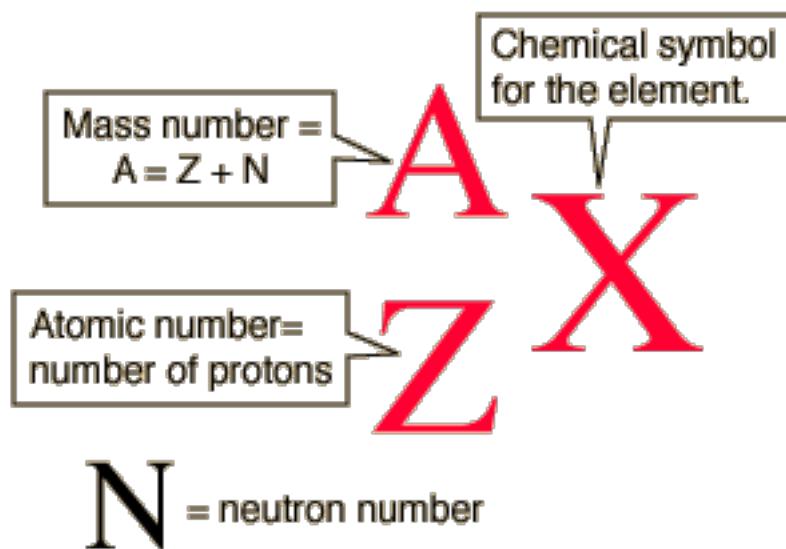
- \* [International Nucleotide Sequence Database Collaboration Policy](#)
- \* [GEO Disclaimer](#)

[Top](#)

## 2 Background

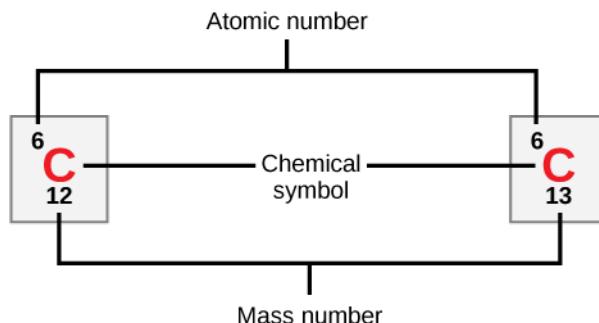
### 2.1 Chemistry

- The **unified atomic mass unit** or **dalton** (symbol:  $u$ , or  $Da$ ) is a standard unit of mass that quantifies mass on an atomic or molecular scale (atomic mass). One unified atomic mass unit is approximately the mass of one nucleon (either a single proton or neutron) and is numerically equivalent to 1 g/mol.
- Nuclear Notation



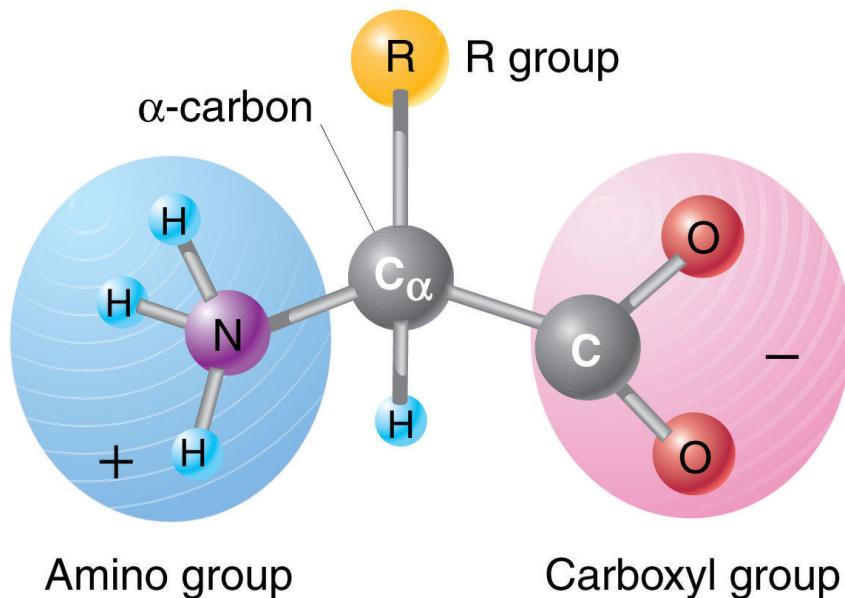
**Figure 1: Nuclear Notation.** Standard nuclear notation shows the chemical symbol, the mass number and the atomic number of the isotope.

- Isotopes



**Figure 2: Isotopes** The different isotopes of a given element have the same atomic number but different mass numbers since they have different numbers of neutrons. E.g.,  $^{12}\text{C}$  and  $^{13}\text{C}$

## 2.2 Biochemistry



© 2010 Pearson Education, Inc.

**Figure 3: Stereoisometry of Amino Acids.** The four bonds of the central or alpha carbon ( $C_\alpha$ ) of an amino acid are directed towards the four corners of a tetrahedron. With respect to the carboxyl ( $COO^-$ ) and amino ( $NH_3^+$ ) groups, there are two possible arrangements of the H and **Radical group**. These arrangement are literally mirror images of each other, and are called stereoisomers (AKA enantiomers). Stereoisomers are designated D (dextro-rotatory) or L (levo-rotatory) according to the direction in which the crystalline forms rotate polarized light, to the right and left, respectively. Naturally-occurring amino acids are exclusively of the L form.

## Quantitative Proteomics Analysis

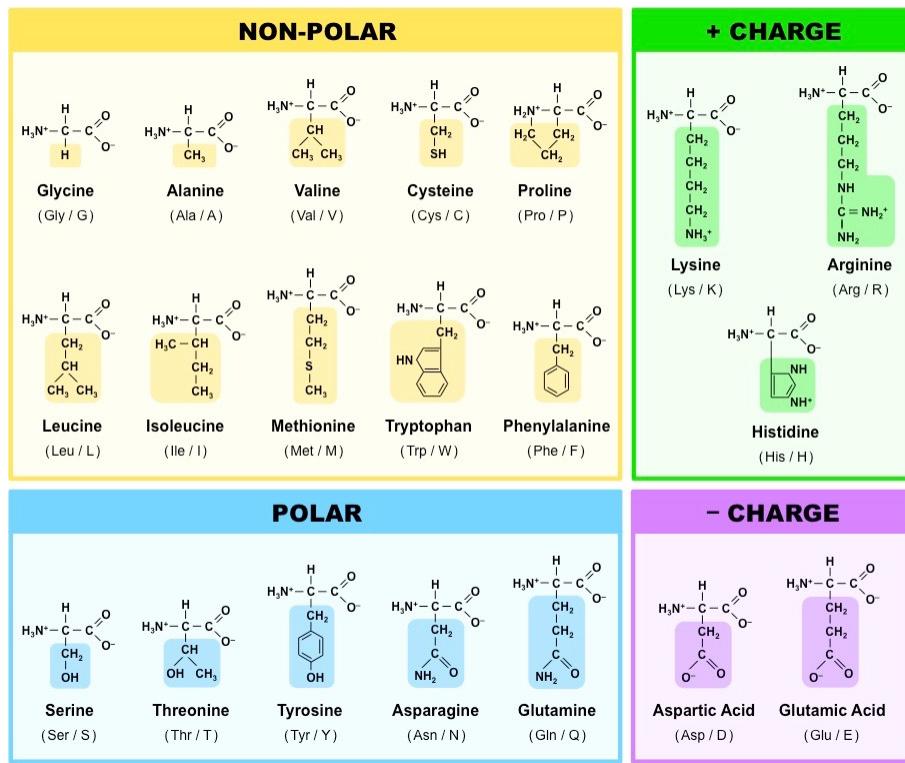


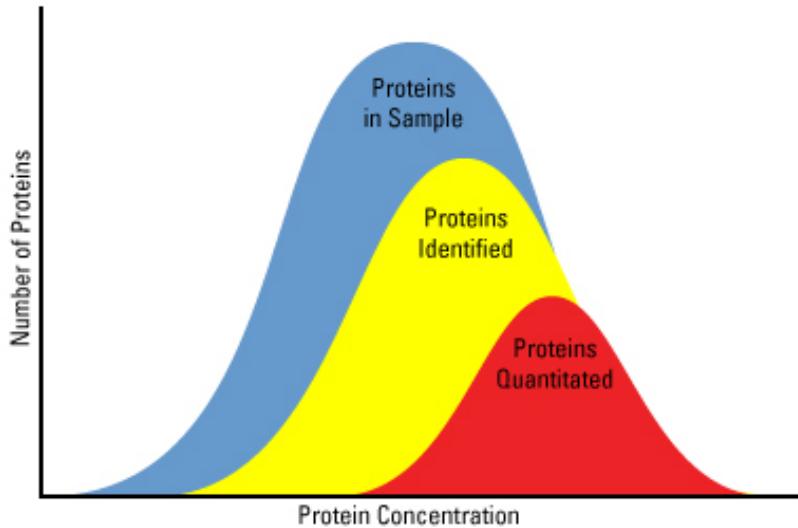
Figure 4: Amino Acid Chart.

Top

## 3 Introduction

Since the development of the first modern mass spectrometer in 1918 (Dempster 1918), the technique has advanced steadily over time and found its' way into a range of applications from forensic toxicology to cancer diagnostics. **Quantitative proteomics** is a powerful approach used for both discovery and targeted proteomic analyses to understand global proteomic dynamics in a cell, tissue or organism.

## Quantitative Proteomics Analysis



**Figure 5: Protein abundance and sample complexity.** Protein abundance and sample complexity are significant factors that affect the availability of proteins for mass spectrometric quantitation. (Bantscheff et al. 2007)

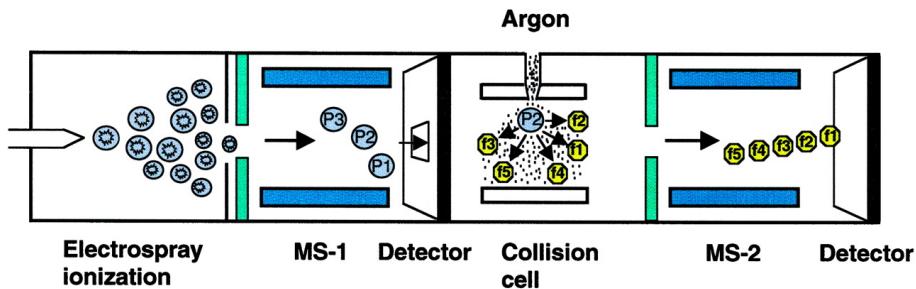
Most quantitative proteomic analyses entail the isotopic labeling of proteins or peptides in the experimental groups, which can then be differentiated by mass spectrometry. Relative quantitation methods (**Stable isotope labeling with amino acids in cell culture (SILAC)**, **Isotope-coded affinity tag (ICAT)**, **Isotope-coded protein label (ICPL)** and isobaric tags (e.g., **iTRAQ**)) are used to compare protein or peptide abundance between samples, while **spiking unlabeled samples with known concentrations of isotopically-labeled synthetic peptides** can yield absolute quantitation of target peptides via selected reaction monitoring (SRM).

**Label-free strategies** are also available for both relative and absolute quantitation. Although these strategies are more complex than mere protein identification, quantitative proteomics is critical for our understanding of global protein expression and modifications underlying the molecular mechanisms of biological processes and disease states.

### 3.1 Quantitative Approaches

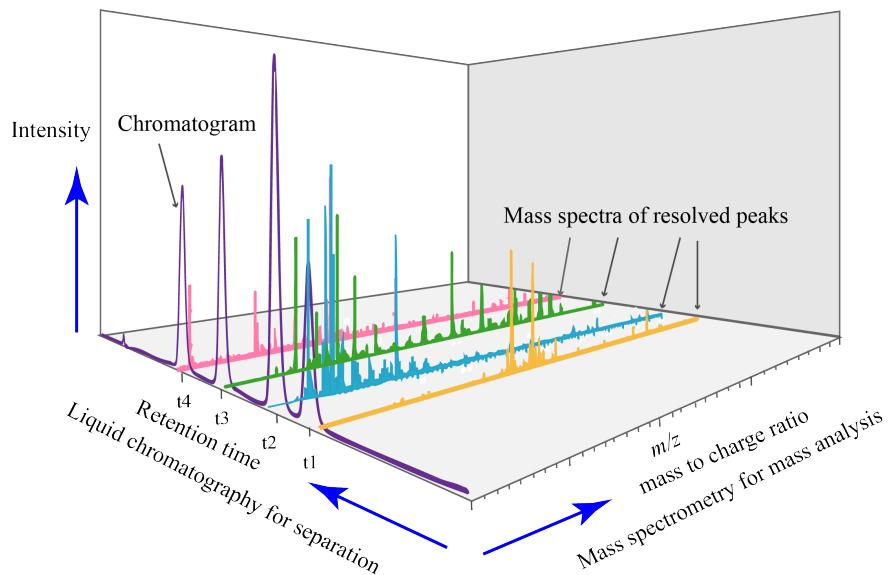
- Mass spectrometry vs. Tandem MS (MS/MS)
  - Intensity based: (MS1)
  - Spectral Counting: (MS2 = MS/MS)

## Quantitative Proteomics Analysis



**Figure 6: Tandem mass spectrometry.** Tandem mass spectrometry, also known as MS/MS or MS2, involves multiple steps of mass spectrometry selection, with some form of fragmentation occurring in between the stages. In a tandem mass spectrometer, ions are formed in the ion source and separated by **mass-to-charge ratio** in the first stage of mass spectrometry (**MS1**). Ions of a particular mass-to-charge ratio (precursor ions) are selected and fragment ions (product ions) are created by **collision-induced dissociation**, ion-molecule reaction, photodissociation, or other process. The resulting ions are then separated and detected in a second stage of mass spectrometry (**MS2**).

- Liquid chromatography – (tandem) mass spectrometry (LC-MS(/MS))



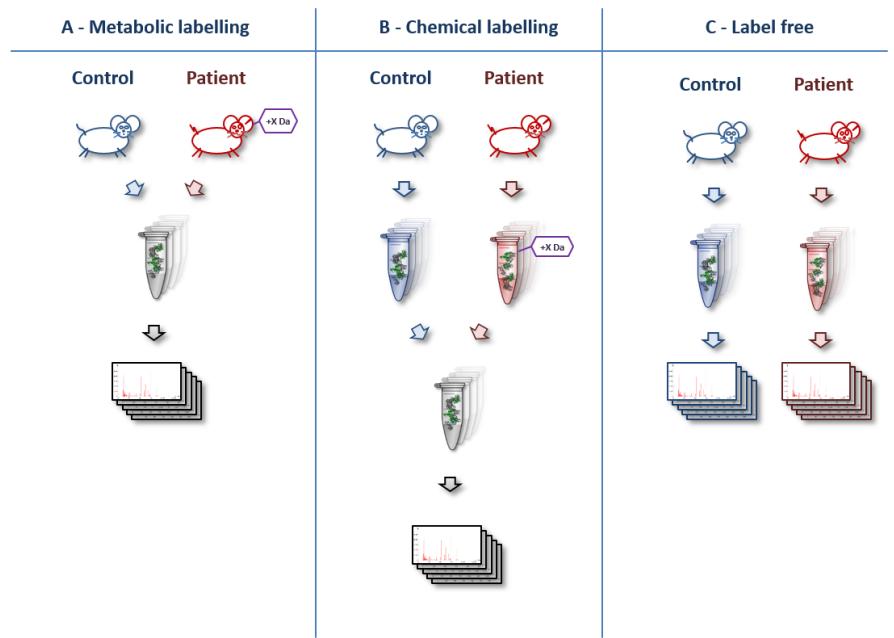
**Figure 7: LC-MS Spectrum.** N.B., Most scans (Retention time-axis) don't have any signal, once starting to receive signal, continues scans should have signals around specific  $m/z$  ratio.

## Quantitative Proteomics Analysis

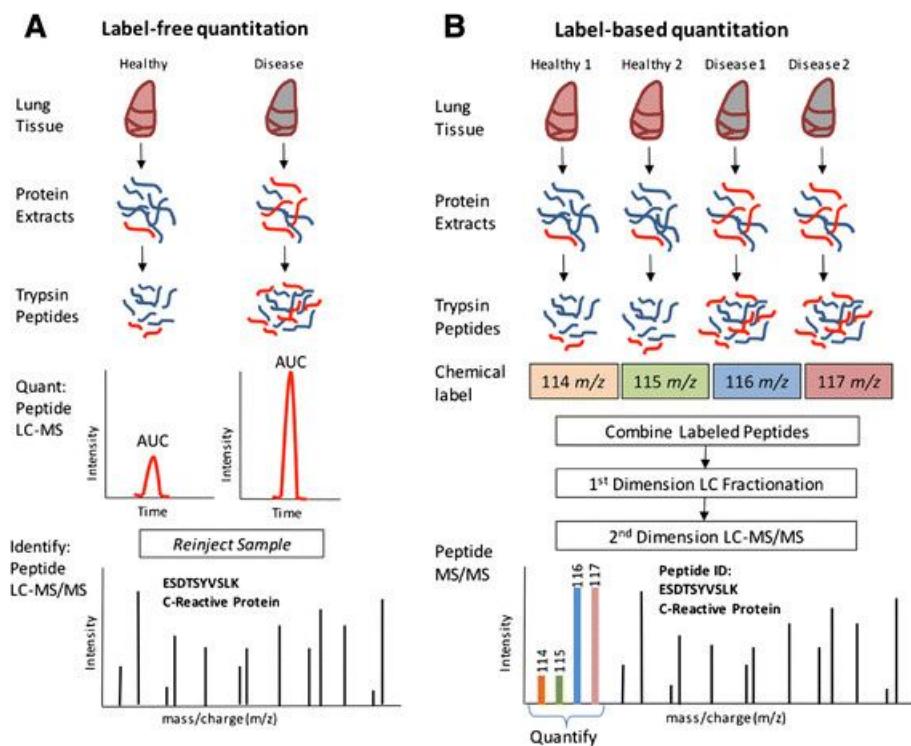
### 3.1.1 Labeling versus Label-Free

- labeling approaches
  - Metabolic: **SILAC** (Ong et al. 2002)
  - Enzymatic: 16O to 18O
  - Chemical
    - **ICAT** (Gygi et al. 1999): 1H to 2H/D
    - **iTRAQ** (Ross et al. 2004)/**TMT** (Andrew Thompson et al. 2003) (followed by Spectral based MS/MS): 12C6 to 13C6, 14N7 to 15N7, 16O8 to 18O8
  - Cons: Most label-based quantification approaches have potential limitations:
    - **complex sample preparation,**
    - **the requirement for increased sample concentration,**
    - and **incomplete labeling.** (Patel et al. 2009)
- label-free (unlabeled) approaches (Panchaud et al. 2008)
  - Nonlabeled techniques which have been (first) developed include peptide match score summation (**PMSS**) (Allet et al. 2004) and spectrum sampling (**SpS**) (Liu, Sadygov, and Yates 2004), both of which can be combined with statistical evaluation to detect differentially expressed proteins (Colinge et al. 2005). Another approach utilizes a protein abundance indices (**PAIs**) (Rappsilber et al. 2002), which can be converted to exponentially modified PAI (**emPAI**) for absolute protein quantification (Ishihama et al. 2005).
  - PMSS (Allet et al. 2004)
    - The method is based on **the assumption that a protein score is a sum of identification scores of its peptides and that a high protein score is correlated with a higher abundance**, thus yielding semi-quantitative information.
  - SpS (Liu, Sadygov, and Yates 2004)
    - A very similar approach to PMSS relies on the counting of spectra identifying a protein.
  - PAIs (Rappsilber et al. 2002)
    - Another method to PMSS/SpS is believed to be more reliable as they are based on observable parameters.
  - emPAI (Ishihama et al. 2005)
    - An improved method of PAI by introducing the observed logarithmic relationship between the number of peptides observed and the protein amount within given sample.

## Quantitative Proteomics Analysis



**Figure 8: A basic overview of various quantitation techniques.** Label-free methods for quantitation are popular in the proteomics community and may be the most straight-forward laboratory technique in the field. However, labelled approaches have some benefits, which make them the method-of-choice for some scientific questions or experimental designs. (“Galaxy Training: Label-Free Versus Labelled - How to Choose Your Quantitat...” 2019)



## Quantitative Proteomics Analysis

**Figure 9: Principles of quantitative proteomics.** **A)** Label-free quantitation performed by peptide peak area under the curve. Proteins are extracted from tissue, proteolytically digested into peptides and analyzed by liquid chromatography (LC)-MS. Analyte intensity versus retention time profiles are generated from which area under the curve (AUC) or summed peak intensities are calculated. Relative peptide amount in healthy versus disease sample is proportional to peak AUC or summed intensities. Targeted peptide identification is typically performed on a subsequent injection. **B)** Label-based quantitation with the iTRAQ (isotope tagging for relative and absolute quantitation) 4plex workflow. Proteins from four individual samples are digested into peptides that are tagged with isobaric stable isotope labeled chemicals. Four chemical tags have 4 unique mass-to-charge ( $m/z$ ) values that are produced during peptide tandem MS (MS/MS) and used for relative quantitation by relative peak intensity. Peptide fragment ions are used for peptide ID and protein inference.. (Bhargava et al. 2014) **N.B., In Figure A, a specific  $m/z$  ratio was set/selected to draw the curve in Quant: Peptide LC-MS.**

**Table 1: Pros and Cons of Label-free and Labelled Quantitation Methods**

Method	Pros	Cons
Label-free	• No labeling required • Simple workflow • Cost-effective	• Low sensitivity • Limited dynamic range • No absolute quantification
iTRAQ	• High sensitivity • Absolute quantification • Multiplexing capability	• Expensive reagents • Complex workflow • Requires multiple injections
SILAC	• High sensitivity • Absolute quantification • Multiplexing capability	• Expensive reagents • Complex workflow • Requires multiple injections
ICAT	• High sensitivity • Absolute quantification • Multiplexing capability	• Expensive reagents • Complex workflow • Requires multiple injections
TMT	• High sensitivity • Absolute quantification • Multiplexing capability	• Expensive reagents • Complex workflow • Requires multiple injections

**Table 2: A Mass Spec Timeline**

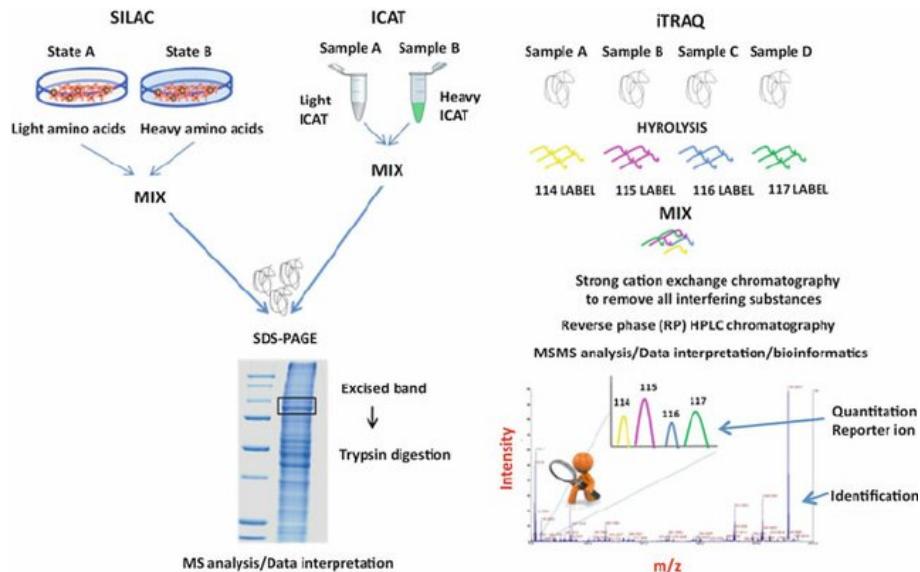
Year	Labeling	MS Level	Label-free	MS Level
1899~1911	1st mass spectrometer			
1918	1st modern mass spectrometer			
1999	ICAT	MS		
2002	SILAC	Both	PAI	MS/MS
2003	TMT	MS/MS		
2004	iTRAQ	MS/MS	PMSS/SpS	MS/MS
2005			emPAI	MS/MS

### 3.1.2 Labeling

- Reagent Kit / Protein labeling / Multiplexing
  - Treatment vs. control followed by MS
    - **Stable isotope labeling with amino acids in cell culture (SILAC)**
      - use cell-culture enrichment with a stable isotope-labeled amino acid, including arginine/Arg/R, lysine/Lys/K, tyrosine/Tyr/Y, and leucine/Leu/L, for *in vivo* incorporation of a mass difference to support relative quantitation.
    - **Isotope-coded affinity tag (ICAT)**
      - $^{1\text{H}}$  to  $^{2\text{H}}$ /D
  - Multiplexing followed by MS/MS: **Isobaric tags** (TMT or iTRAQ) have identical masses and chemical properties that allow heavy and light isotopologues to co-elute together. The tags are then cleaved from the peptides by **collision-induced dissociation (CID)** during MS/MS, which is used for quantification.
    - **Isobaric tag for relative and absolute quantitation (iTRAQ)**
      - a multiplexed set of reagents for quantitative protein analysis that place isobaric mass labels **at the N termini and lysine/Lys/K side chains** of peptides in a digest mixture.
      - is a registered trademark of **SCIEX**

## Quantitative Proteomics Analysis

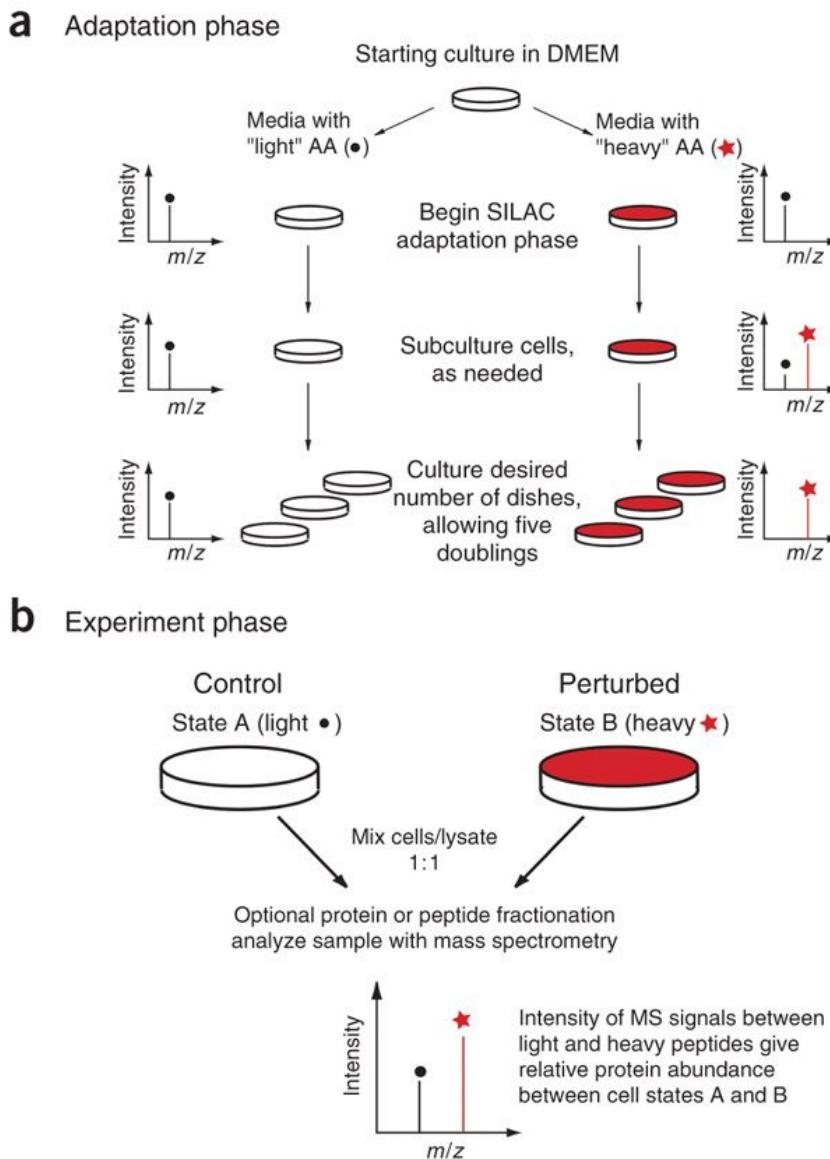
- is available in **4-plex** and **8-plex** formats
- **Tandem Mass Tag (TMT)**
  - is a registered trademark of Proteome Sciences PLC (licensing its Patented TMT Technology to **Thermo Fisher Scientific**)
  - is available in **2-plex**, **6-plex**, and (since recently) **10-plex** formats



**Figure 10: Representative workflows for SILAC, ICAT, and iTRAQ.** The main differences among labeling techniques are (i) **SILAC** and **ICAT** labeling are applied on intact proteins, while **iTRAQ** labeling is performed on peptides, and (ii) in the case of **SILAC** and **ICAT**, peptides are quantified during MS analysis, while in the case of **iTRAQ**, quantitation occurs during fragmentation, i.e., MS/MS analysis. (Salvatore et al. 2015)

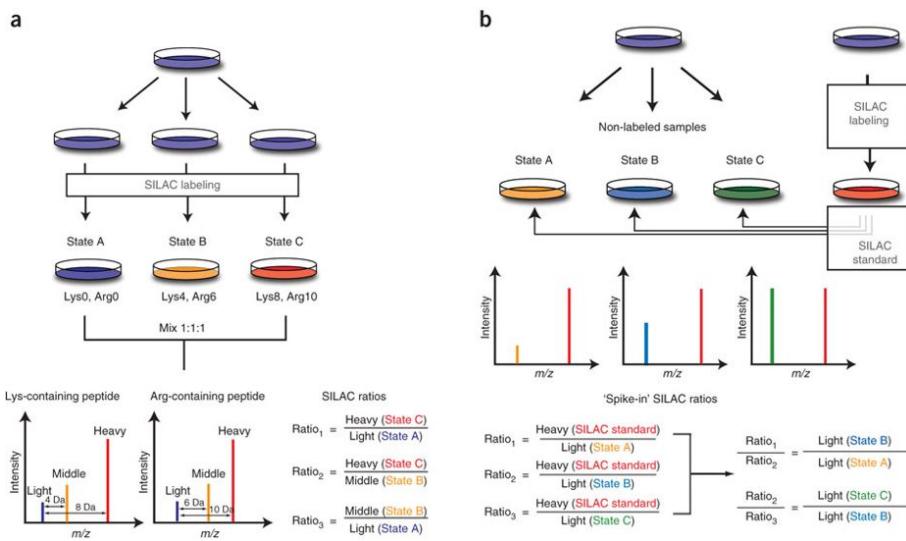
## Quantitative Proteomics Analysis

### 3.1.3 Labeling::SILAC: stable isotope labeling with amino acids in cell culture



**Figure 11: Overview of SILAC protocol.** The SILAC experiment consists of two distinct phases — an adaptation (a) and an experimental (b) phase. (a) During the adaptation phase, cells are grown in light and heavy SILAC media until the heavy cells have fully incorporated the heavy amino acids (red star). This allows the two SILAC cell pools to be fully distinguishable by MS (black dot and red star, indicating light and heavy SILAC peptides, respectively) and can then be mixed and processed as a single sample. The adaptation phase can include the expansion of cells to reach the required number of dishes for the experiment. (b) In the second phase, the two cell populations are differentially treated, inducing changes in the proteome. The sample is mixed, a subproteome can be purified by an enrichment step or other fractionation, digested to peptides as a single pool and analyzed by MS for protein identification and quantification. (Ong and Mann 2006)

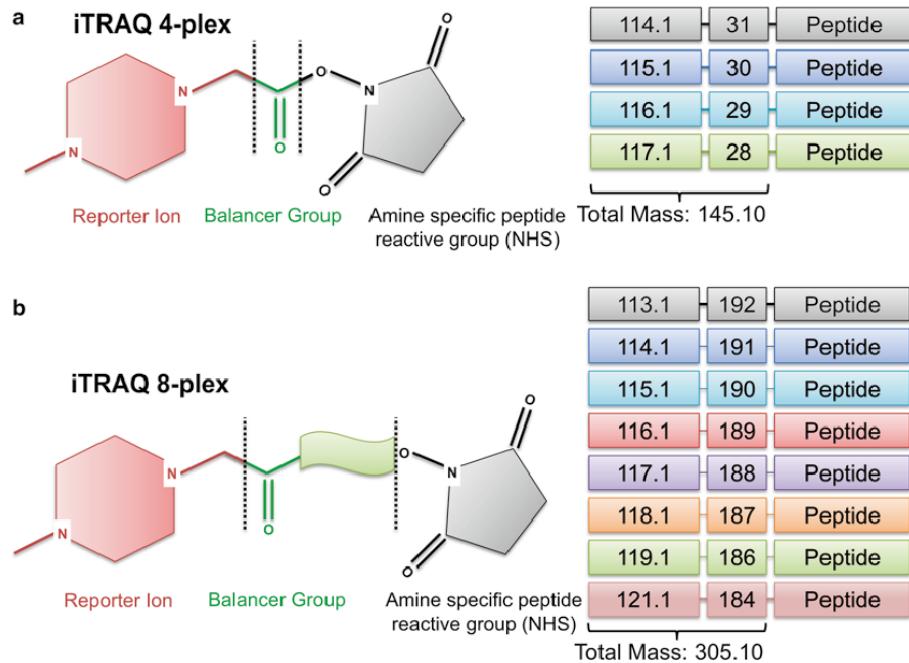
## Quantitative Proteomics Analysis



**Figure 12: The workflow of classical SILAC experiment versus spike-in SILAC standard.** (a) In the classical approach, two or three cell populations are labeled with heavy amino acids, then combined and analyzed together by LC-MS/MS. In the MS spectra, each peptide appears as a doublet or triplet with distinct mass differences. The ratios between the samples are calculated directly by comparing the differences in the intensities of the peaks. (b) With the spike-in SILAC standard, the labeling is separated or 'decoupled' from the biological experiment, which is then carried out under normal cell culture conditions. After the experiment is performed, the non-labeled samples are combined with the SILAC standard and each of these combined samples is analyzed separately by LC-MS/MS. The difference between the experimental samples is calculated as the 'ratio of ratios', where the ratio of one sample relative to the standard is divided by the ratio of the other relative to the standard. (Geiger et al. 2011)

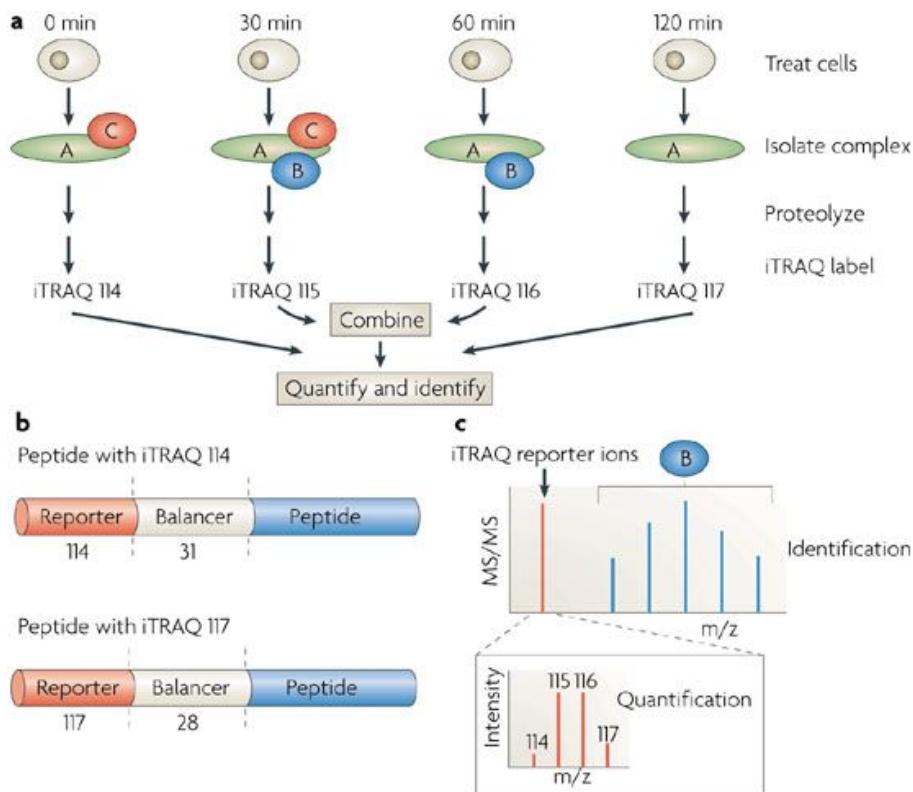
## Quantitative Proteomics Analysis

### 3.1.4 Labeling::iTRAQ: isobaric tag for relative and absolute quantitation



**Figure 13: Chemical structures for iTRAQ ( a ) 4-plex and ( b ) 8-plex isobaric tags.** Balancer + reporter ions add up to 145 Da in 4-plex and 304 Da in 8-plex experiments. In 8-plex, reporter mass of 120 is not present as it will give erroneous quantitation since phenylalanine immonium ion is also observed at a mass of 120 Da. (Aggarwal and Yadav 2016)

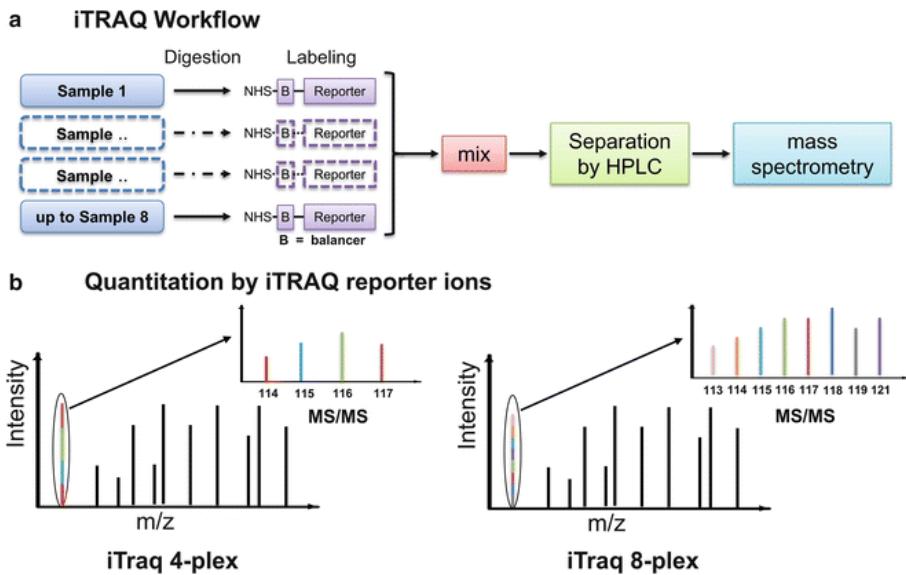
## Quantitative Proteomics Analysis



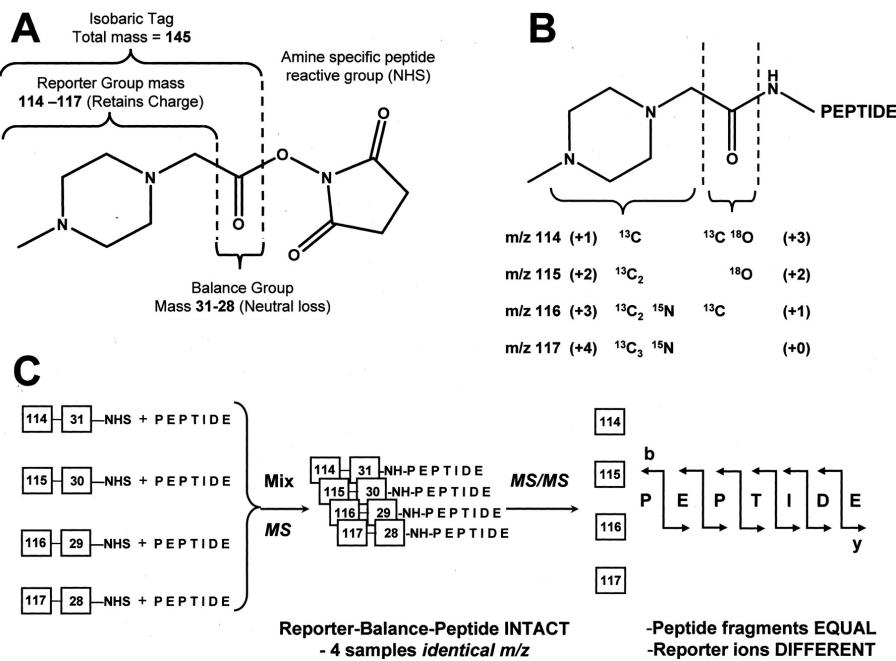
Nature Reviews | Molecular Cell Biology

**Figure 14: Isobaric tags to elucidate complex formation dynamics.** **a** | Desired treatment of cells is followed by isolation of protein complexes and proteolysis. **iSobaric tags (iTRAQ)** are chemically added to the N terminus of every peptide (as well as to lysine  $\epsilon$ -amine groups). Samples from multiple treatment time points are combined and subjected to analysis. **b** | A peptide labelled with the iTRAQ 114 and iTRAQ 117 reagents. iTRAQ is isobaric, such that addition of the 114-Da or 117-Da mass tags alter the mass of a given peptide by the same amount. To maintain a constant mass, the reporter moiety (for example, of mass 114) is separated from the peptide by a balancer group. The reporter and balancer groups fragment in the collision cell of the mass spectrometer during the tandem mass spectrometry (MS/MS) event, and the intensity of the reporter ions is monitored. **c** | Analysis of an iTRAQ experiment. MS/MS analysis of a labelled peptide generates a fragmentation spectrum that yields the sequence of the peptide. The iTRAQ reagent is fragmented in the same step and reporter ions are quantified by magnifying the low mass range (114–117) area. In the example shown, protein B associates with protein A (the bait) after 30 and 60 minutes of stimulation, but not after 120 minutes of treatment.  $m/z$ , mass/charge ratio. (Gingras et al. 2007)

## Quantitative Proteomics Analysis



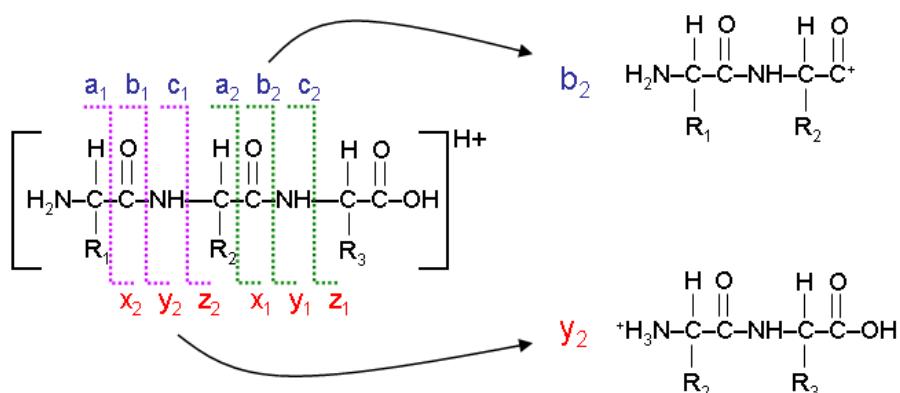
**Figure 15: iTRAQ Workflow.** (a) Basic iTRAQ workflow in which up to eight samples can be separately digested and labeled with iTRAQ tags, mixed together, separated by HPLC (high-performance liquid chromatography), and analyzed by mass spectrometer. (b) During MS/MS, the reporter ions of differential masses are released from peptide to give sample-specific quantitation of a particular peptide (Aggarwal and Yadav 2016)



**Figure 16: Features of Multiplexed Tagging Chemistry.** A, diagram showing the components of the multiplexed isobaric tagging chemistry. The complete molecule consists of a reporter group (based on N-methylpiperazine), a mass balance group (carbonyl), and a peptide-reactive group (NHS ester). The overall mass of reporter and balance components of the molecule are kept constant using differential isotopic enrichment with **13C**, **15N**, and **18O** atoms (B), thus avoiding problems with chromatographic separation seen with enrichment involving deuterium substitution. The number and position of enriched centers in

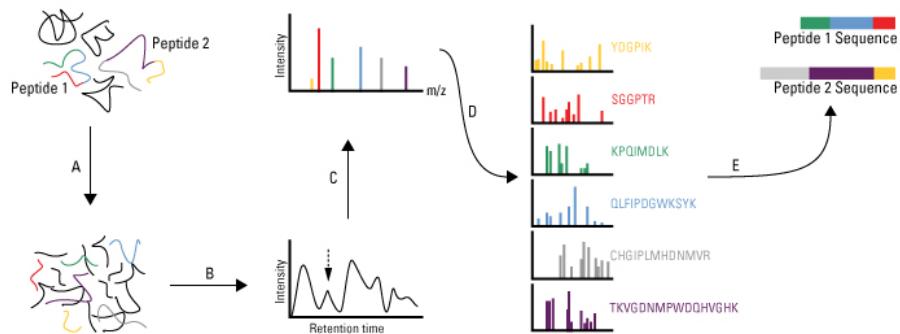
## Quantitative Proteomics Analysis

the ring has no effect on chromatographic or MS behavior. The reporter group ranges in mass from  $m/z$  114.1 to 117.1, while the balance group ranges in mass from 28 to 31 Da, such that the combined mass remains constant (145.1 Da) for each of the four reagents. **B**, when reacted with a peptide, the tag forms an amide linkage to any peptide amine (N-terminal or  $\epsilon$  amino group of lysine/Lys/K). These amide linkages fragment in a similar fashion to backbone peptide bonds when subjected to CID. Following fragmentation of the tag amide bond, however, the balance (carbonyl) moiety is lost (neutral loss), while charge is retained by the reporter group fragment. The numbers in parentheses indicate the number of enriched centers in each section of the molecule. **C**, illustration of the isotopic tagging used to arrive at four isobaric combinations with four different reporter group masses. A mixture of four identical peptides each labeled with one member of the multiplex set appears as a single, unresolved precursor ion in MS (identical  $m/z$ ). Following CID, the four reporter group ions appear as distinct masses (114–117 Da). All other sequence-informative fragment ions ( $b$ -,  $y$ -, etc.) remain isobaric, and their individual ion current signals (signal intensities) are additive. This remains the case even for those tryptic peptides that are labeled at both the N terminus and lysine/Lys/K side chains, and those peptides containing internal lysine/Lys/K residues due to incomplete cleavage with **trypsin (Solution Stable Enzyme for Mass Spectrometry)**. The relative concentration of the peptides is thus deduced from the relative intensities of the corresponding reporter ions. In contrast to **Isotope-coded affinity tag (ICAT)** and similar mass-difference labeling strategies, **quantitation is thus performed at the MS/MS stage rather than in MS**. (Ross et al. 2004)

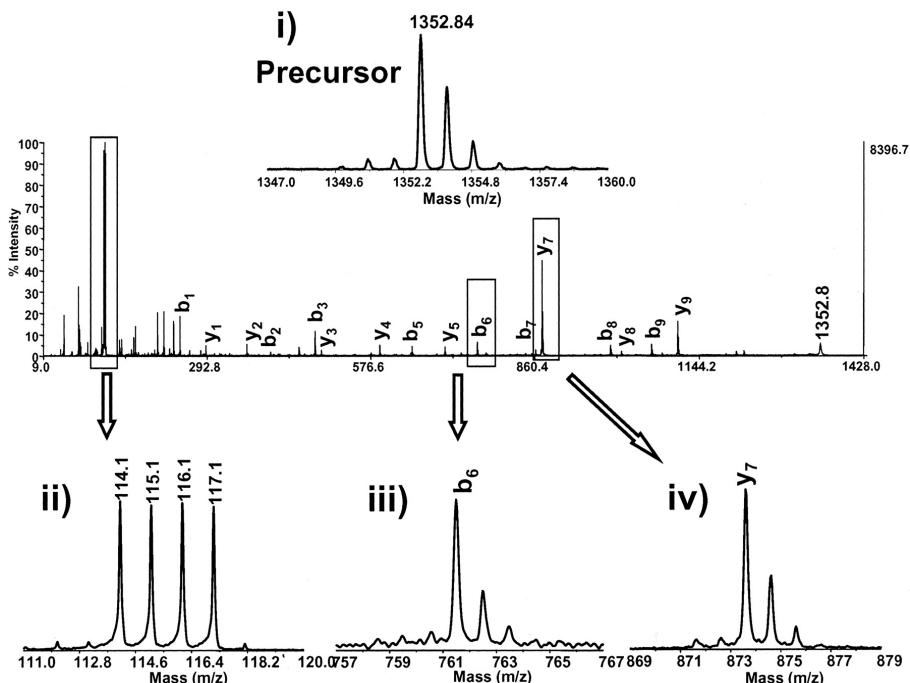


**Figure 17: Peptide Sequence Fragmentation.** Fragmentation of peptides (amino acid chains) typically occurs along the peptide backbone. Each residue of the peptide chain successively fragments off, both in the N->C and C->N direction. The location that the fragmentation occurs, and the nature of the ion remaining results in various ions, a, b, c and x, y, or z ions. The most commonly observed ions are a, b, and y ions. The following diagram illustrates the formation of b and y ions during the fragmentation of a three residue peptide chain. [Resource](#)

## Quantitative Proteomics Analysis



**Figure 18: Overview of proteomic analysis by MS/MS.** Sample proteins are extracted and digested into peptides (A). The sample complexity may then be reduced prior to chemical separation by LC (B). Fractions (indicated by dotted arrow) are then analyzed by MS (C), during which the peptides are ionized and their mass-to-charge ratio ( $m/z$ ) measured to yield a precursor ion spectrum. Selected ions are then fragmented by collision-induced dissociation (CID) and the individual fragment ions measured by MS (D). The fragment ion spectra are then assigned peptide sequences based on database comparison and protein sequences are predicted (E). [Thermo Fisher](#)



**Figure 19: Example MS/MS spectrum of peptide TPHPALTEAK from a protein digest mixture prepared by labeling four separate digests with each of the four isobaric reagents and combining the reaction mixtures in a 1:1:1:1 ratio.** Components of the spectrum illustrated are (i) isotopic distribution of the precursor ( $[M+H]^+$ ,  $m/z$  1352.84), (ii) low mass region showing the signature ions used for quantitation, (iii) isotopic distribution of the b<sub>6</sub> fragment, and (iv) isotopic distribution of the y<sub>7</sub> fragment ion. The peptide is labeled by isobaric tags at both the N terminus and C-terminal lysine/Lys/K side chain. The precursor ion and all the internal fragment ions (e.g. type b- and y-) therefore contain all four members of the tag set, but remain isobaric. The example shown is the spectrum obtained from the

singly charged [M+H]<sup>+</sup> peptide using a 4700 MALDI TOF-TOF analyzer, but the same holds true for any multiply charged peptide analyzed with an ESI-source mass spectrometer. (Ross et al. 2004)

[Top](#)

## 4 Computation

### 4.1 Organization



- HUPO-PSI
  - The **Proteomics Standards Initiative (PSI)** is a part of the **Human Proteome Organisation (HUPO)**
  - The **PSI-MSS** working group defines community data formats and controlled vocabulary terms facilitating data exchange and archiving in the field of proteomics mass spectrometry.



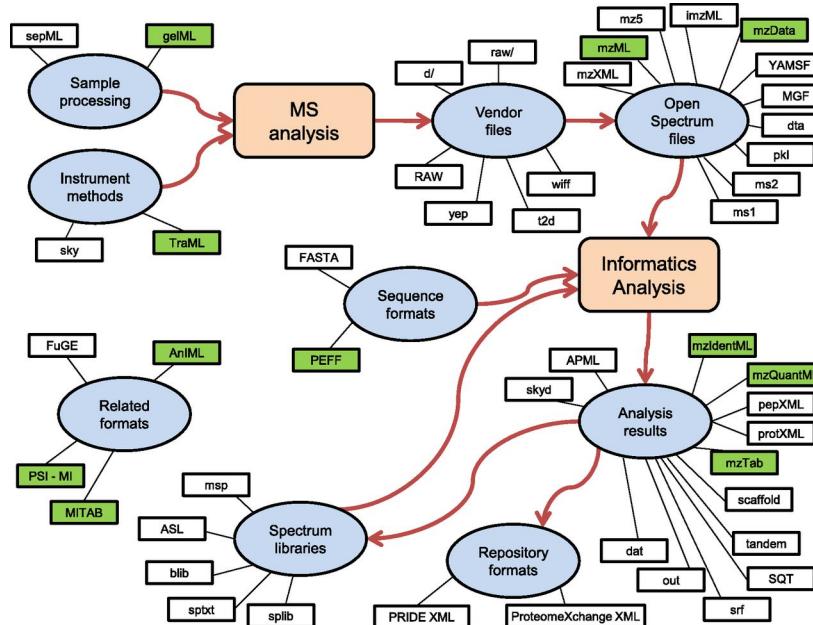
- ISB-SPC
  - **Seattle Proteome Center (SPC)** at the **Institute for Systems Biology (ISB)**

### 4.2 Data Format

- [Wiki](#)
- mzData was the first attempt by the Proteomics Standards Initiative (PSI) from the Human Proteome Organisation (HUPO) to create a standardized format for Mass Spectrometry data. **This format is now deprecated, and replaced by mzML.**

## Quantitative Proteomics Analysis

- [mzXML](#) by ISB-SPC
  - an open data format for storage and exchange of mass spectroscopy data, developed at the Seattle Proteome Center (SPC)/Institute for Systems Biology. mzXML provides a standard container for ms and ms/ms proteomics data and is the foundation of our proteomic pipelines. Raw, proprietary file formats from most vendors can be converted to the open mzXML format.
- [mzML](#) by HUPO-PSI = mzData (by PSI) + mzXML (by SPC)
  - The mzML format, which merges the mzData format and another similar format mzXML.
  - mzML 1.1.0 was released on June 1, 2009 and has been stable since then.

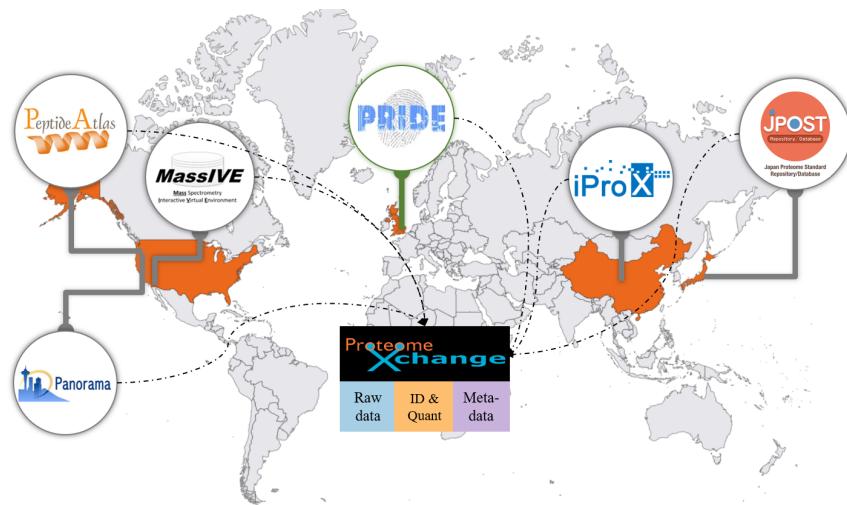


**Figure 20: Overview graph of the mass spectrometry proteomics formats.** The overall workflow of MS proteomics is depicted by the large shapes and the arrows connecting them. Ovals represent the major data types within the workflow. The small rectangles represent the individual file formats associated by an edge to their general data type. Shaded formats are officially approved or soon-to-be-approved standards. Different formats associated with the same data type are not necessarily redundant or equivalent. (Deutsch 2012)

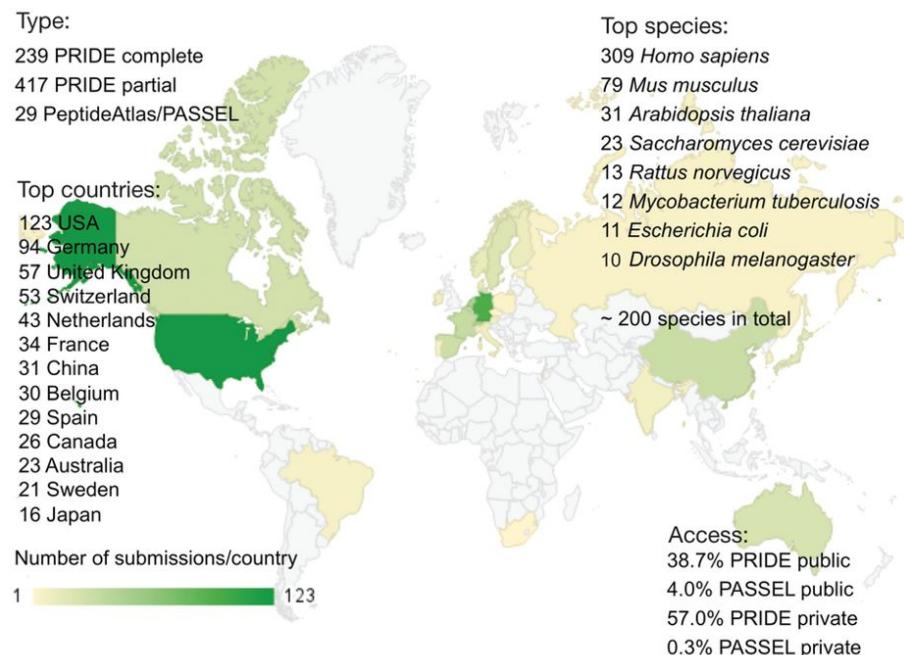
## 4.3 Data Repository

- [ProteomeXchange Consortium](#)
  - The **ProteomeXchange Consortium** was established to provide globally coordinated standard data submission and dissemination pipelines involving the main proteomics repositories, and to encourage open data policies in the field.
  - Publications
    - ProteomeXchange provides globally coordinated proteomics data submission and dissemination (J. A. Vizcaíno et al. 2014)
    - The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition (Deutsch et al. 2017)
  - Members & Submission Summary

## Quantitative Proteomics Analysis



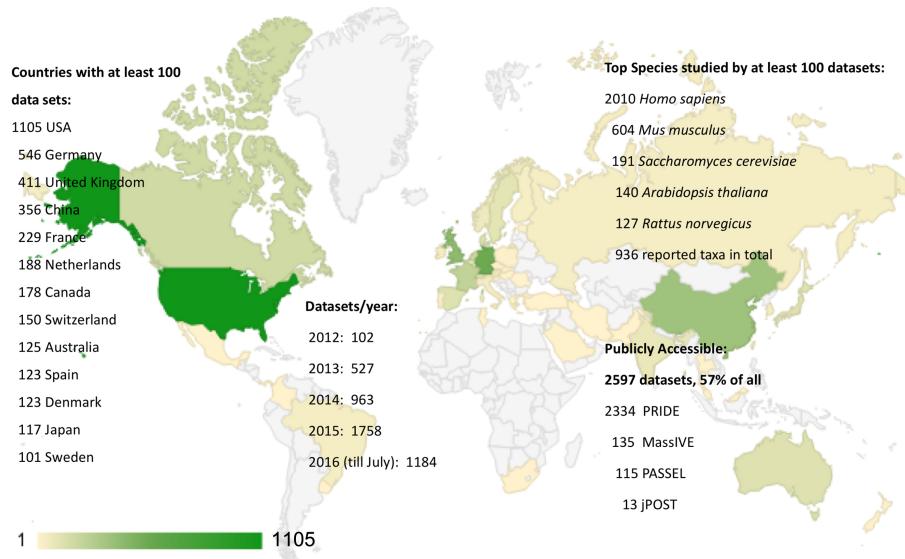
**Figure 21: Members in ProteomeXchange** The current members of the Consortium are: **PRIDE** (EMBL-EBI, Cambridge, UK), **PeptideAtlas** (ISB, Seattle, WA, USA) (both of them are the founding members), **MassIVE** (UCSD, San Diego, CA, USA), **jPOST** (various institutions, Japan), **iProx** (National Center for Protein Sciences, Beijing, China) and **Panorama Public** (University of Washington, Seattle, WA, USA)".



**Figure 22: Summary of the main metrics of ProteomeXchange submissions (as of February 2014).** ProteomeXchange started to accept regular submissions in June 2012. As of the beginning of February 2014, 685 ProteomeXchange data sets have been submitted (consisting of 656 tandem MS and 29 SRM data sets; Fig. 2), a total of ~ 32 Tb of data. The largest submission so far (data sets PXD000320–PXD000324) comprised 5 Tb of data.

## Quantitative Proteomics Analysis

### ProteomeXchange: 4534 datasets up until end of July 2016



**Figure 23: Summary of the main metrics of ProteomeXchange submitted data sets (by the end of July 2016).** By the end of July 2016, a total of 4534 PX data sets had been submitted to any of the PX resources. In terms of individual resources, around 4067 data sets (representing 89.7% of all the data sets), had been submitted to PRIDE, followed by MassIVE (339 data sets), PASSEL (115 data sets) and jPOST (13 data sets, just joined PX at the beginning of July 2016). Data sets come from 50 countries, demonstrating the global reach of the consortium. The most represented countries are USA (1105 data sets), Germany (546), United Kingdom (411), China (356) and France (229). Since 2012, the number of submitted data sets has increased substantially every year, ranging from 102 (2012) to 1758 (2015).

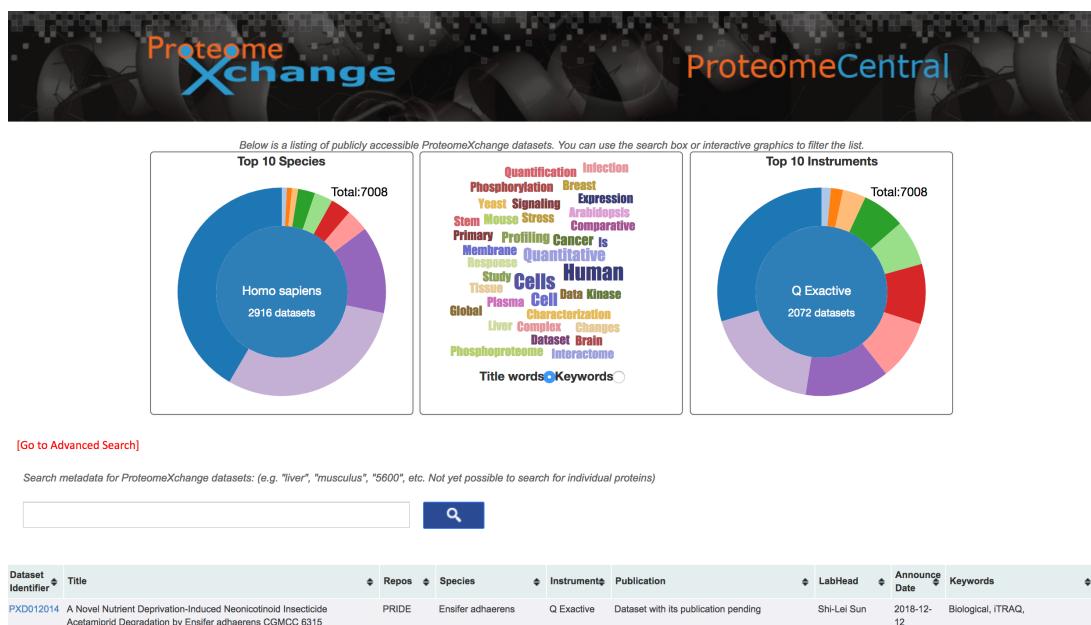


Figure 24: A listing of publicly accessible ProteomeXchange datasets.

## 4.4 Database Search

- Mascot (Eng, McCormack, and Yates 1994)
- SEQUEST now comet (Perkins et al. 1999)
- MaxQuant/Andromeda (J. Cox et al. 2011)

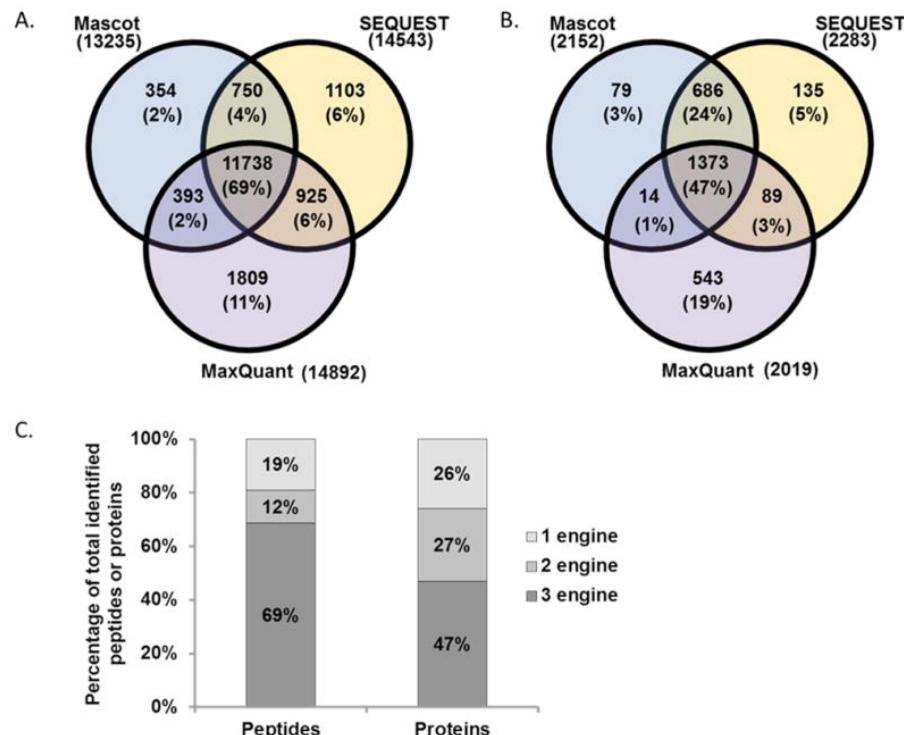


Figure 25: Search engine comparison. Venn diagrams comparing **A**) peptide identifications and **B**) protein identifications. **C**) Bar graph illustrating the redundancy of peptides and proteins identified by one, two, and three search engines. As a result, **Peptide overlap among the search engines was greater than that of proteins**.

Top

## 5 Miscellaneous

- Resource
  - Protein Quantitation by Mass Spectrometry
  - Quantitative Proteomics @ Thermo Fisher
- Papers
  - Using R and Bioconductor for proteomics data analysis (Gatto and Christoforou 2014)

## Quantitative Proteomics Analysis

- Visualization of proteomics data using R and Bioconductor ("Visualization of Proteomics Data Using R and Bioconductor - Gatto - 2015 - PROTEOMICS - Wiley Online Library" 2018)
- Quantitative mass spectrometry in proteomics: a critical review (Bantscheff et al. 2007)
- Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present (Bantscheff et al. 2012)
- Dissecting the iTRAQ Data Analysis (Aggarwal and Yadav 2016)

[Top](#)

## 6 Session information

```
## R version 3.5.1 (2018-07-02)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS High Sierra 10.13.6
##
## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] grid      stats     graphics  grDevices utils     datasets  methods
## [8] base
##
## other attached packages:
## [1] captioner_2.2.3.9000 dplyr_0.7.99.9000   png_0.1-7
## [4] kableExtra_0.9.0    knitr_1.20       BiocStyle_2.10.0
## [7]
## loaded via a namespace (and not attached):
## [1] Rcpp_1.0.0          pillar_1.3.0      compiler_3.5.1
## [4] BiocManager_1.30.4  tools_3.5.1      digest_0.6.18
## [7] evaluate_0.12       tibble_1.4.2      viridisLite_0.3.0
## [10] pkgconfig_2.0.2     rlang_0.3.0.1    rstudioapi_0.8
## [13] yaml_2.2.0         xfun_0.4        stringr_1.3.1
## [16] httr_1.3.1         xml2_1.2.0      hms_0.4.2
## [19] tidyselect_0.2.5    rprojroot_1.3-2  glue_1.3.0
## [22] R6_2.3.0           rmarkdown_1.10   bookdown_0.8
## [25] purrr_0.2.5        readr_1.2.1      magrittr_1.5.0.9000
## [28] backports_1.1.2    scales_1.0.0     htmltools_0.3.6
## [31] assertthat_0.2.0    rvest_0.3.2      colorspace_1.3-2
## [34] stringi_1.2.4      munsell_0.5.0    crayon_1.3.4
```

[Top](#)

## Reference

- Aggarwal, Suruchi, and Amit Kumar Yadav. 2016. "Dissecting the iTRAQ Data Analysis." In *Statistical Analysis in Proteomics*, edited by Klaus Jung, 277–91. Methods in Molecular Biology. New York, NY: Springer New York. doi:[10.1007/978-1-4939-3106-4\\_18](https://doi.org/10.1007/978-1-4939-3106-4_18).
- Allet, Nadia, Nicolas Barrillat, Thierry Baussant, Celia Boiteau, Paolo Botti, Lydie Bougueret, Nicolas Budin, et al. 2004. "In Vitro and in Silico Processes to Identify Differentially Expressed Proteins." *Proteomics* 4 (8): 2333–51. doi:[10.1002/pmic.200300840](https://doi.org/10.1002/pmic.200300840).
- Andrew Thompson, †, ‡ Jürgen Schäfer, ‡ Karsten Kuhn, ‡ Stefan Kienle, ‡ Josef Schwarz, † Günter Schmidt, ‡ Thomas Neumann, and ‡ and Christian Hamon\*. 2003. "Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS." Research-article. March 1. doi:[10.1021/ac0262560](https://doi.org/10.1021/ac0262560).
- Bantscheff, Marcus, Simone Lemeer, Mikhail M. Savitski, and Bernhard Kuster. 2012. "Quantitative Mass Spectrometry in Proteomics: Critical Review Update from 2007 to the Present." *Analytical and Bioanalytical Chemistry* 404 (4): 939–65. doi:[10.1007/s00216-012-6203-4](https://doi.org/10.1007/s00216-012-6203-4).
- Bantscheff, Marcus, Markus Schirle, Gavain Sweetman, Jens Rick, and Bernhard Kuster. 2007. "Quantitative Mass Spectrometry in Proteomics: A Critical Review." *Analytical and Bioanalytical Chemistry* 389 (4): 1017–31. doi:[10.1007/s00216-007-1486-6](https://doi.org/10.1007/s00216-007-1486-6).
- Bhargava, Maneesh, LeeAnn Higgins, Christine H. Wendt, and David H. Ingbar. 2014. "Application of Clinical Proteomics in Acute Respiratory Distress Syndrome." *Clinical and Translational Medicine* 3 (1): 34. doi:[10.1186/s40169-014-0034-1](https://doi.org/10.1186/s40169-014-0034-1).
- Colinge, Jacques, Diego Chiappe, Sophie Lagache, Marc Moniatte, and Lydie Bougueret. 2005. "Differential Proteomics via Probabilistic Peptide Identification Scores." *Analytical Chemistry* 77 (2): 596–606. doi:[10.1021/ac0488513](https://doi.org/10.1021/ac0488513).
- Cox, Jürgen, Nadin Neuhauser, Annette Michalski, Richard A. Scheltema, Jesper V. Olsen, and Matthias Mann. 2011. "Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment." *Journal of Proteome Research* 10 (4): 1794–1805. doi:[10.1021/pr101065j](https://doi.org/10.1021/pr101065j).
- Dempster, A. J. 1918. "A New Method of Positive Ray Analysis." *Physical Review* 11 (4): 316–25. doi:[10.1103/PhysRev.11.316](https://doi.org/10.1103/PhysRev.11.316).
- Deutsch, Eric W. 2012. "File Formats Commonly Used in Mass Spectrometry Proteomics." *Molecular & Cellular Proteomics* 11 (12): 1612–21. doi:[10.1074/mcp.R112.019695](https://doi.org/10.1074/mcp.R112.019695).
- Deutsch, Eric W., Attila Csordas, Zhi Sun, Andrew Jarnuzak, Yasset Perez-Riverol, Tobias Ternent, David S. Campbell, et al. 2017. "The ProteomeXchange Consortium in 2017: Supporting the Cultural Change in Proteomics Public Data Deposition." *Nucleic Acids Research* 45 (D1): D1100–D1106. doi:[10.1093/nar/gkw936](https://doi.org/10.1093/nar/gkw936).
- Eng, J. K., A. L. McCormack, and J. R. Yates. 1994. "An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database." *Journal of the American Society for Mass Spectrometry* 5 (11): 976–89. doi:[10.1016/1044-0305\(94\)80016-2](https://doi.org/10.1016/1044-0305(94)80016-2).
- "Galaxy Training: Label-Free Versus Labelled - How to Choose Your Quantitat..." 2019. Accessed January 14. <https://galaxyproject.github.io/training-material/topics/proteomics/tutorials/labelfree-vs-labelled/tutorial.html>.
- Gatto, Laurent, and Andy Christoforou. 2014. "Using R and Bioconductor for Proteomics Data Analysis." *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, Computational proteomics in the post-identification era, 1844 (1): 42–51. doi:[10.1016/j.bbapap.2013.04.032](https://doi.org/10.1016/j.bbapap.2013.04.032).
- Geiger, Tamar, Jacek R. Wisniewski, Juergen Cox, Sara Zanivan, Marcus Kruger, Yasushi Ishihama, and Matthias Mann. 2011. "Use of Stable Isotope Labeling by Amino Acids in Cell Culture as a Spike-in Standard in Quantitative Proteomics." *Nature Protocols* 6 (2): 147–57. doi:[10.1038/nprot.2010.192](https://doi.org/10.1038/nprot.2010.192).
- Gingras, Anne-Claude, Matthias Gstaiger, Brian Raught, and Ruedi Aebersold. 2007. "Analysis of Protein Complexes Using Mass Spectrometry". *Nature Reviews Molecular Cell Biology* 8