Protein mass spectrometry and proteomics

L. Dayon

Spring Semester 2025

Course outline

- 1. Introduction

Introduction to protein analysis and proteomics; Reminders in mass spectrometry; Why proteomics and mass spectrometry?; Ionization sources, analysers, and detectors used in proteomics; Latest generation of mass spectrometers used in proteomics

- 2. Proteomic strategy and workflows

Bottom-up versus top-down strategies; Data-dependent acquisition (DDA) and data-independent acquisition (DIA) approaches; Sample preparation

- 3. Separations techniques in proteomics

Gel electrophoresis; Isoelectric focusing; Liquid chromatography (RP, IEX)

- 4. Quantitative proteomic workflows

Label-free methods; Labelling-based techniques; Other quantitative techniques

- 5. Proteomic bioinformatics

Databases; Identification of protein; Quantification of proteins; Bioinformatics tools; Practical examples

- 6. Applications to biology and clinical research

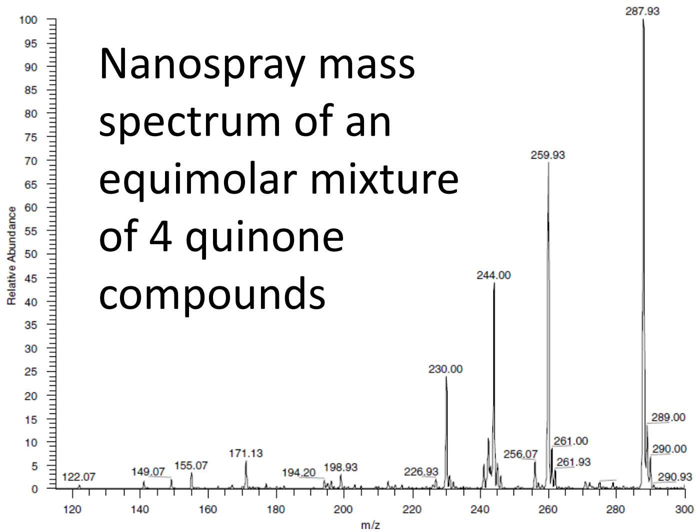
What strategy?; Experimental design; Biomarker discovery; Industrialized and population proteomics; Forensics; Targeted mass spectrometry-based approaches; Other biological applications of mass spectrometry; Advanced innovations (single-cells, 4D proteomics, multi-omics) and emerging technologies; Limitations and ethical consideration; Lab visit

Course outline

- 4. Quantitative proteomic workflows

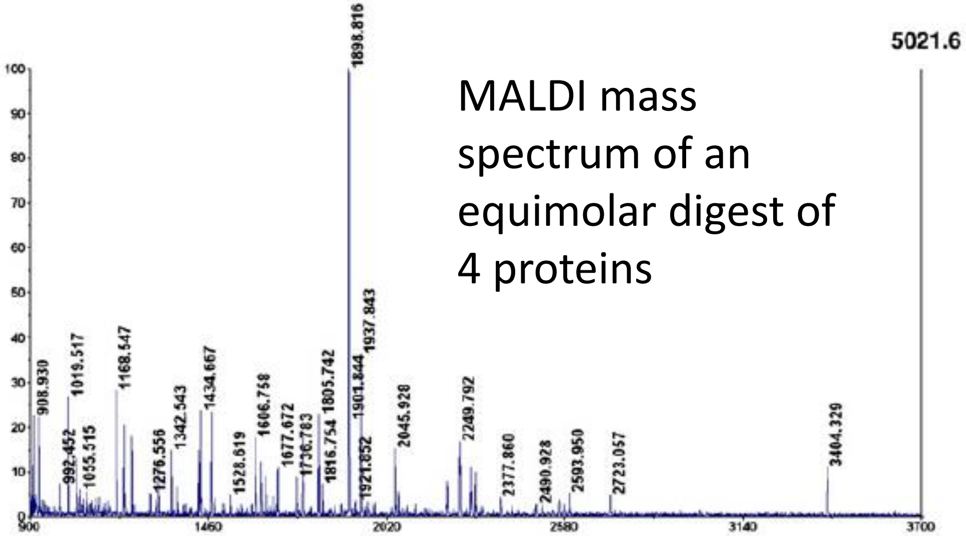
Label-free methods; Labelling-based techniques; Other quantitative techniques

Mass spectrometry is not inherently quantitative



Nanospray mass spectrum of an equimolar mixture of 4 quinone compounds

doi:10.1016/j.jelechem.2004.03.029



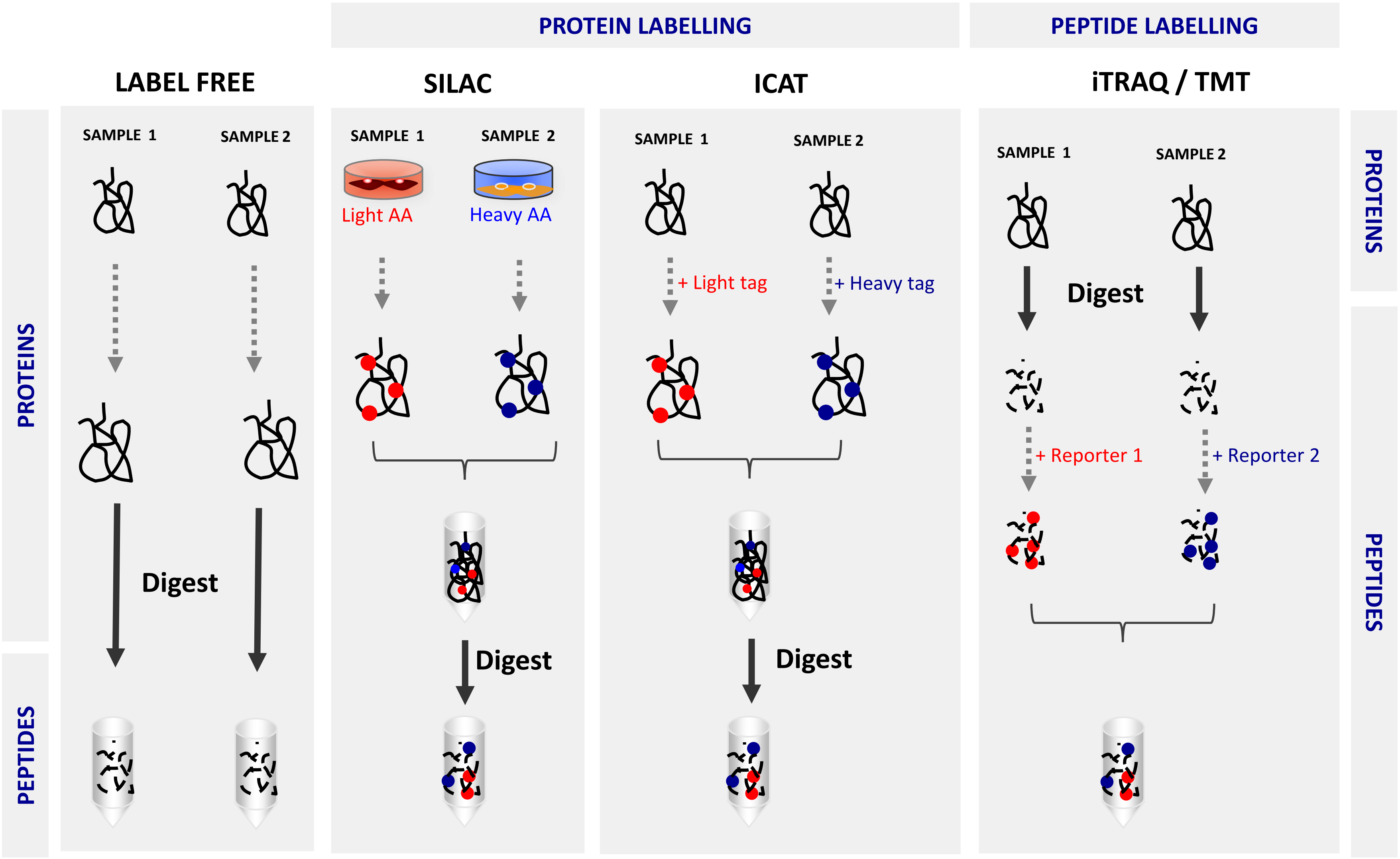
MALDI mass spectrum of an equimolar digest of 4 proteins

doi:10.1016/j.jprot.2008.11.009

But we need to compare proteomes quantitatively and therefore, methodologies have been implemented to turn MS-based proteomics quantitative

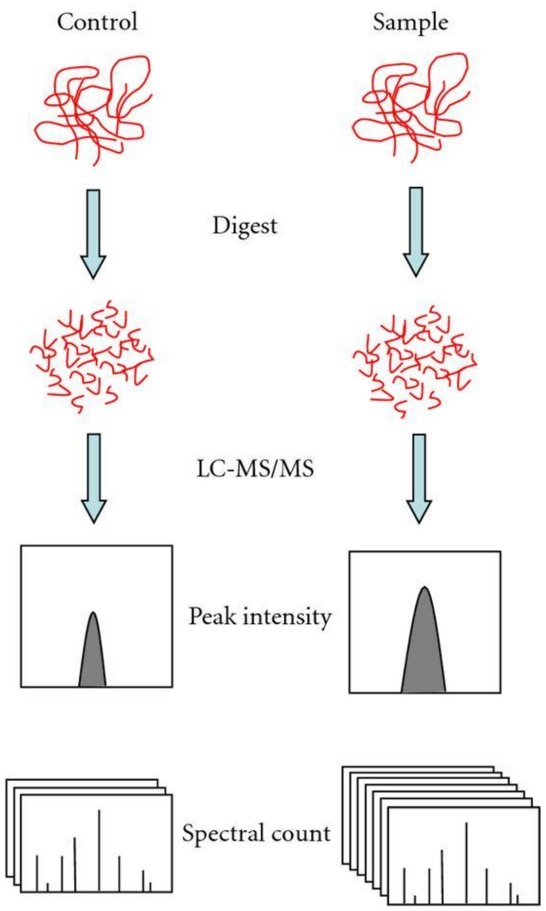
doi:10.1038/nchembio736

Protein relative quantitation with mass spectrometry



M. Abonnenc and M. Mayr ***2012****, Inflammation and Atherosclerosis, 249-266, DOI: 10.1007/978-3-7091-0338-8_13*

4.1. Label-free methods



http://dx.doi.org/10.1155/2010/840518

- Spectrum Counting, which counts and compares the number of fragment spectra identifying peptides of a given protein

- Total Ion Count (TIC), which considers peak intensities from MS/MS spectra combined with counting of the spectra

- Precursor Ion Intensity, which measures and compares the mass spectrometric signal intensity of peptide precursor ions belonging to a particular protein

Spectrum counting

spectral counting: quantification based on spectral counting and concomitant identification (MS/MS)
dynamic range with spectral counting: 3 orders of magnitude

Total spectra

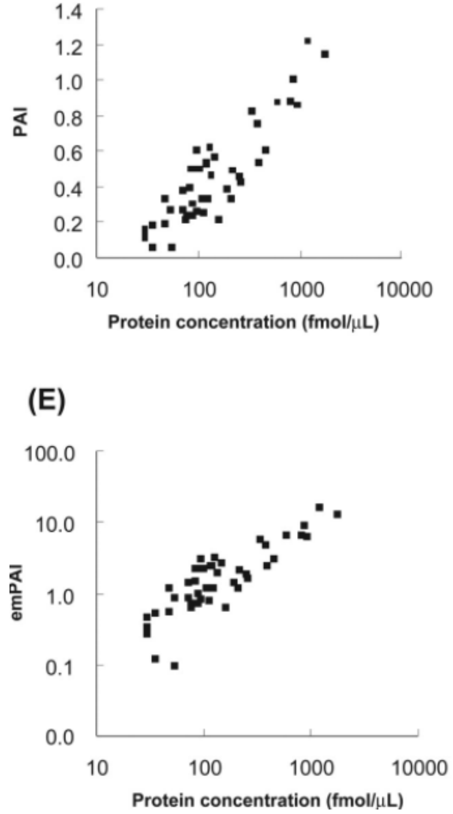
This method uses the sum of all the spectra associated with a specific protein within a sample which includes also those spectra that are shared with other proteins and is referred to as the Total Spectrum Count

emPAI
Spectrum Counting methods can also be used in the determination of absolute abundance of proteins. The Protein Abundance Index (PAI) is defined as the number of observed peptides divided by the number of all possible tryptic peptides from a particular protein, that are within the mass range of the employed mass spectrometer:

$$PAI = \frac{N_{observed}}{N_{observable}}$$

where N_{observed} is the number of experimentally observed peptides and $N_{\text{observable}}$ is the calculated number of observable peptides for each protein.
In a subsequent refinement PAI was transformed into an exponential form called emPAI and defined as follows:

$$emPAI = 10^{PAI} - 1$$



DOI: 10.1074/mcp.M500061-MCP200

Precursor ion intensity

extracted ion chromatogram (XIC): quantification based on peptide-ion intensity (MS) and subsequent identification (MS/MS)
dynamic range for intensities: 6 orders of magnitude

Average Precursor Intensity
This method takes the geometric mean of the peptide intensity values for a given protein

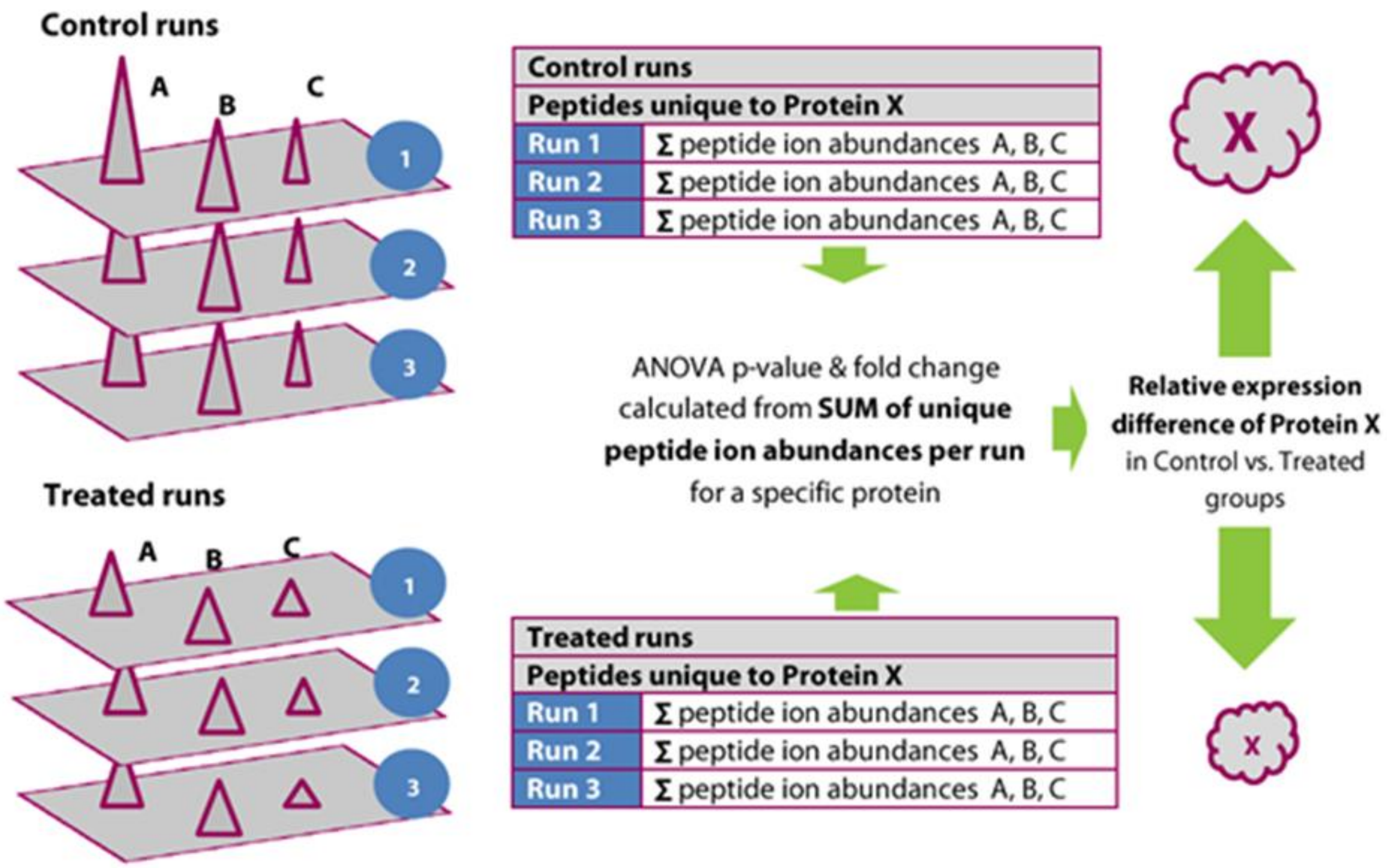
Total Precursor Intensity
The sum of all distinct intensity values for a protein

Top Three Precursor Intensities
The sum of the three highest peptide intensity values for a protein. If fewer than three peptides have intensity values, the intensities that are present are summed

iBAQ
iBAQ (Intensity-Based Absolute Quantification) is a popular approach for absolute quantification of proteins.
It is similar in its approach to the emPAI method. iBAQ and similar algorithms are called intensity-based because they calculate the sum of parent or precursor ion intensities of identified peptides per protein.
iBAQ intensities provide an accurate determination of the relative abundance of all proteins identified in a sample:

iBAQ: Σ intensity/#theoretical peptides

Precursor ion intensity

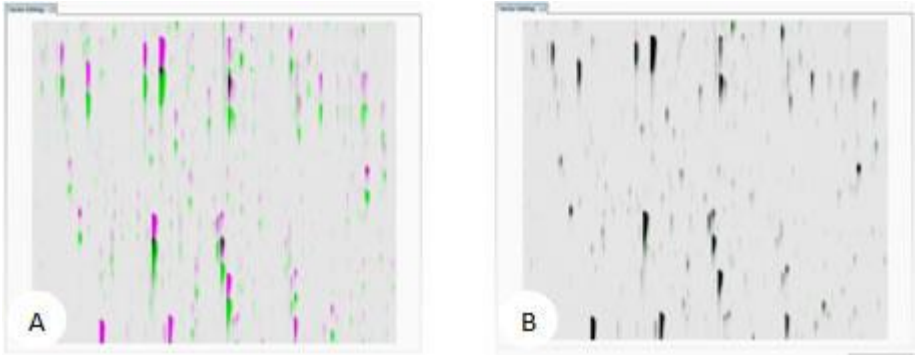


Control runs	
Peptides unique to Protein X	
Run 1	Σ peptide ion abundances A, B, C
Run 2	Σ peptide ion abundances A, B, C
Run 3	Σ peptide ion abundances A, B, C

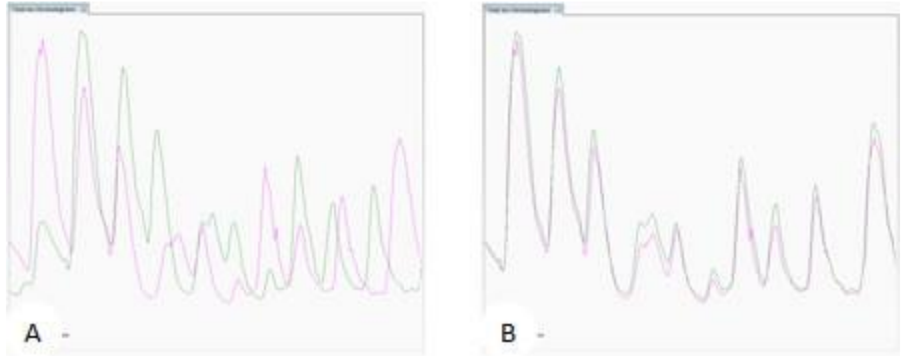
ANOVA p-value & fold change calculated from **SUM of unique peptide ion abundances per run** for a specific protein

Relative expression difference of Protein X in Control vs. Treated groups

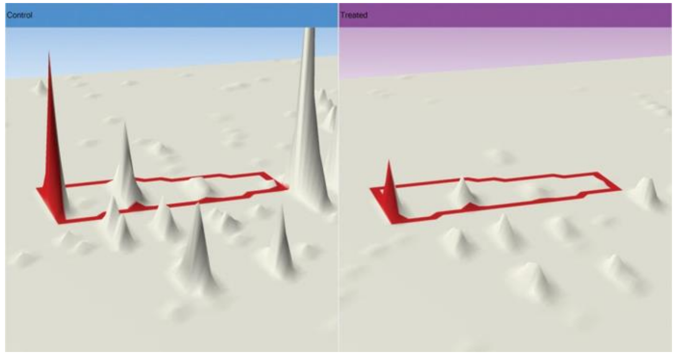
Treated runs	
Peptides unique to Protein X	
Run 1	Σ peptide ion abundances A, B, C
Run 2	Σ peptide ion abundances A, B, C
Run 3	Σ peptide ion abundances A, B, C



LC-MS runs represented as *m/z* vs. RT overlaid. (A) unaligned runs (B) aligned run



Total Ion Chromatogram view. (A) unaligned runs (B) aligned runs



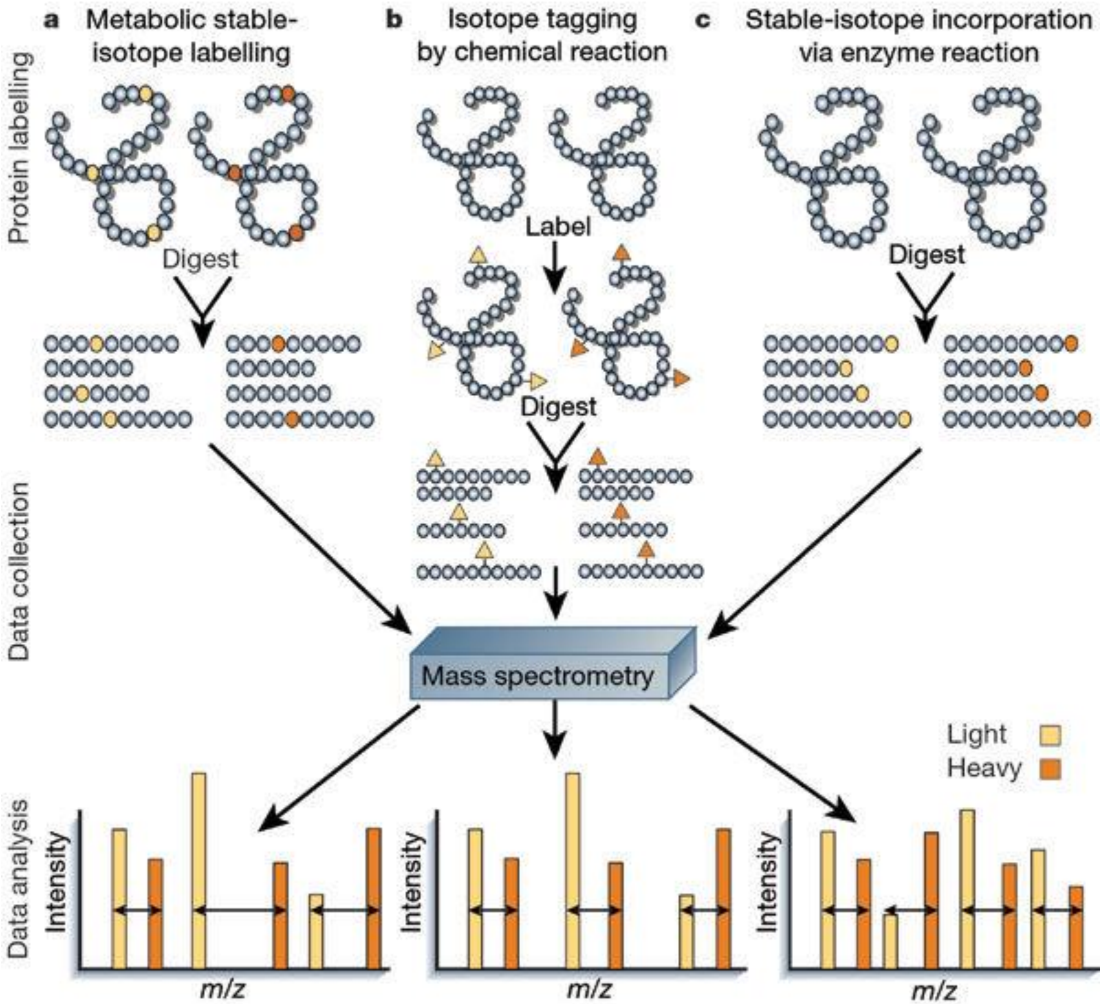
http://www.nonlinear.com

4.2. Labelling-based techniques

Metabolic labelling

Proteolytic labelling

Chemical labelling

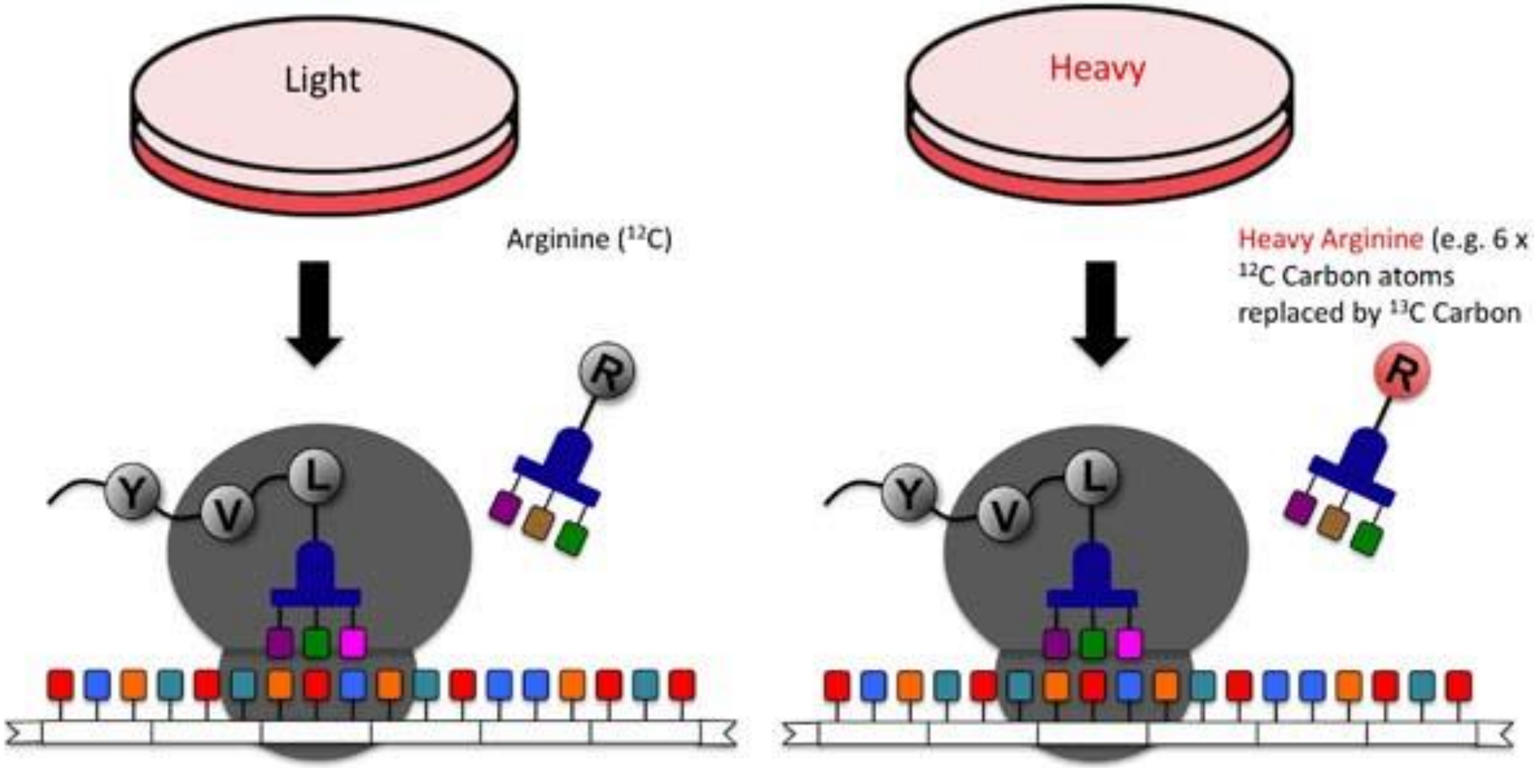


doi:10.1038/nature01511

Metabolic labelling: SILAC

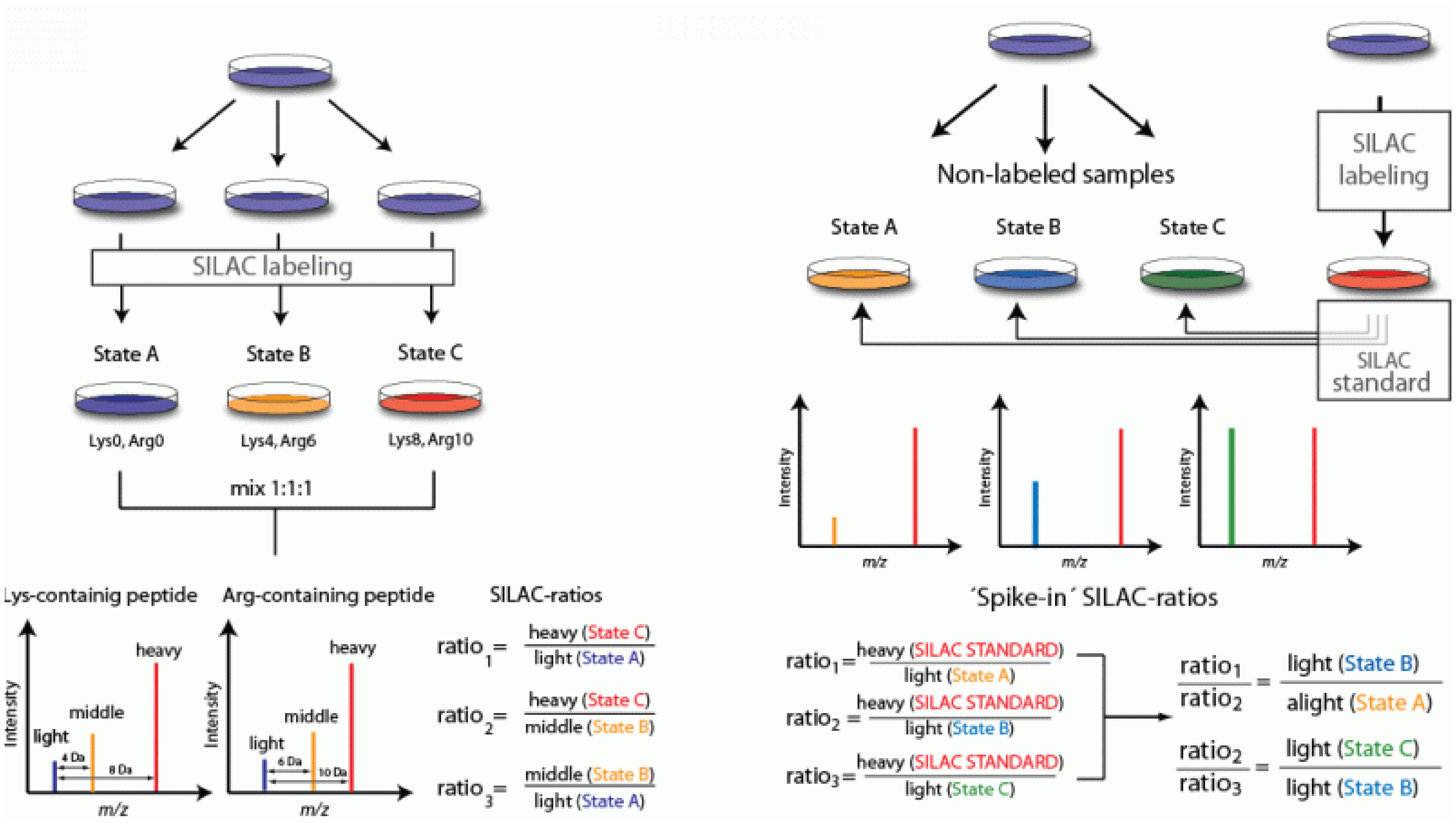
Matthias Mann

Stable Isotope Labeling with Amino Acids in Cell Culture (SILAC)



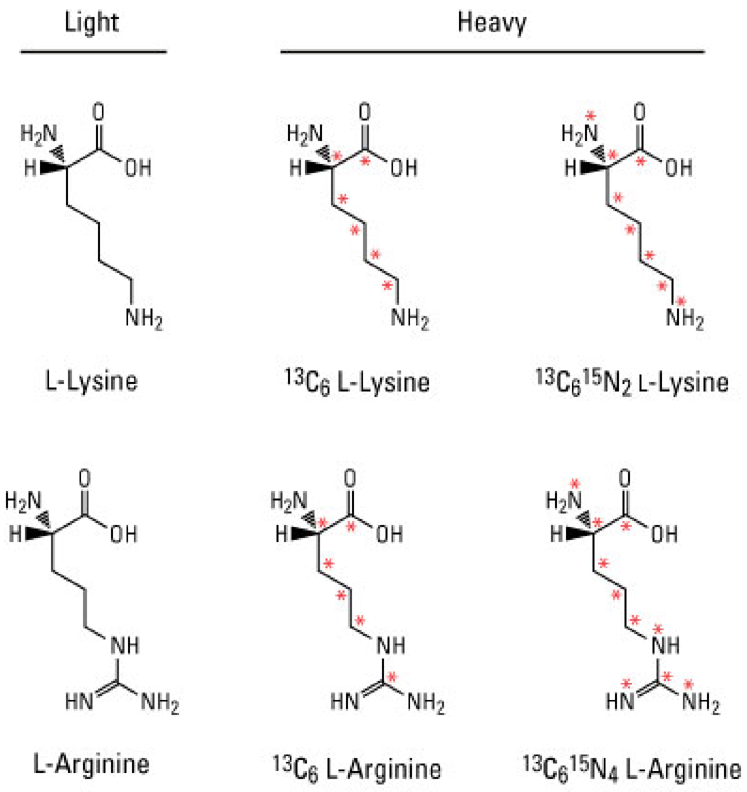
http://www.trinklelab.com/proteomics.html

SILAC



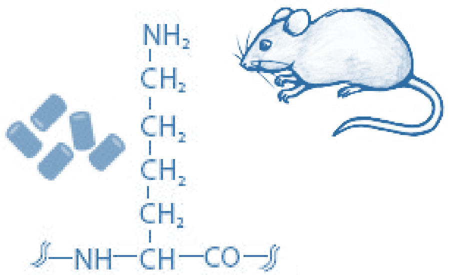
http://www.biochem.mpg.de/221777/SILAC

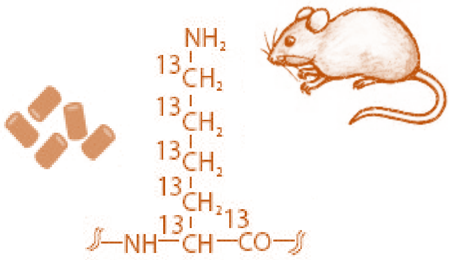
doi:10.1038/nprot.2010.192



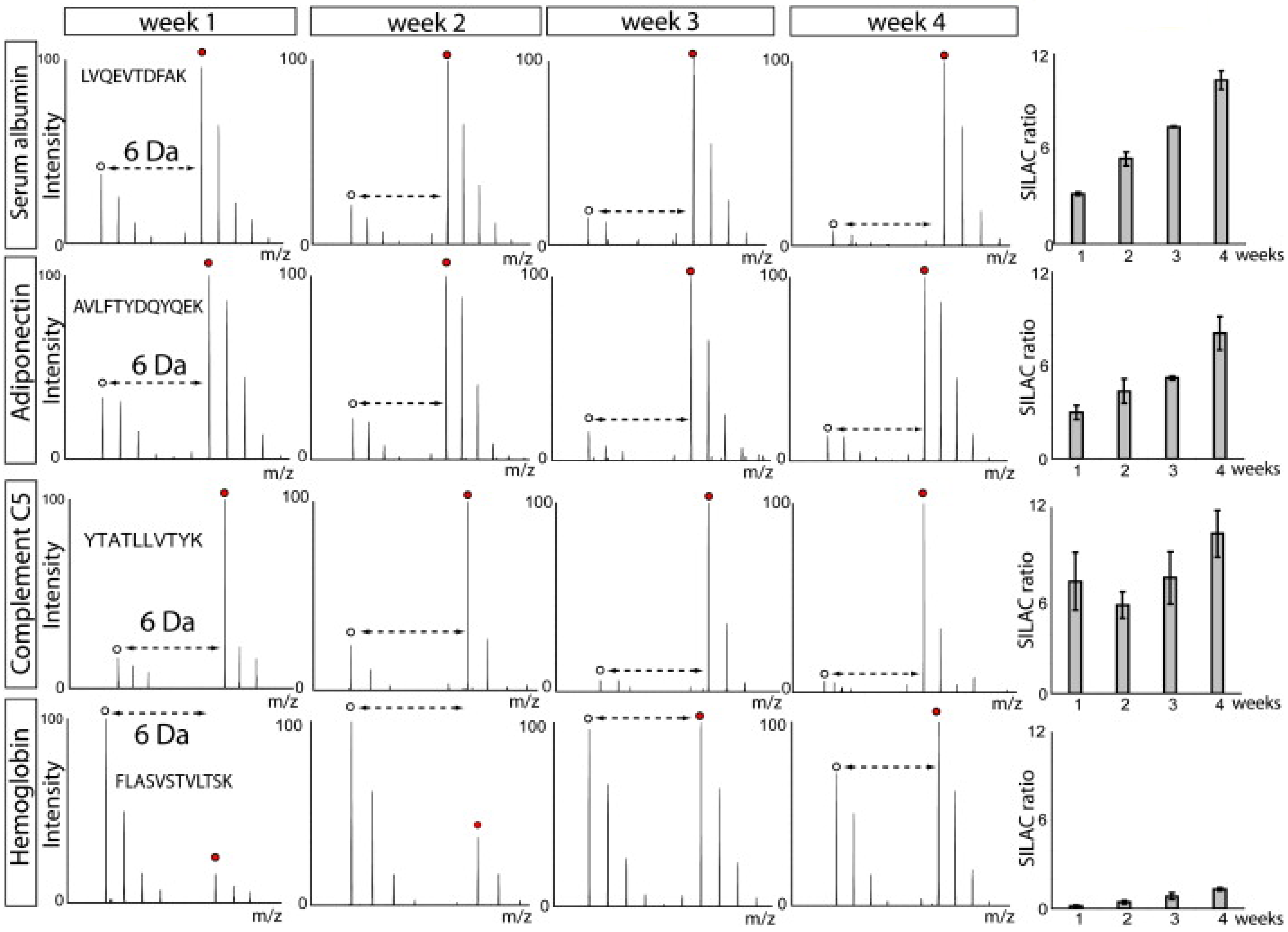
https://www.thermofisher.com

In-vivo SILAC



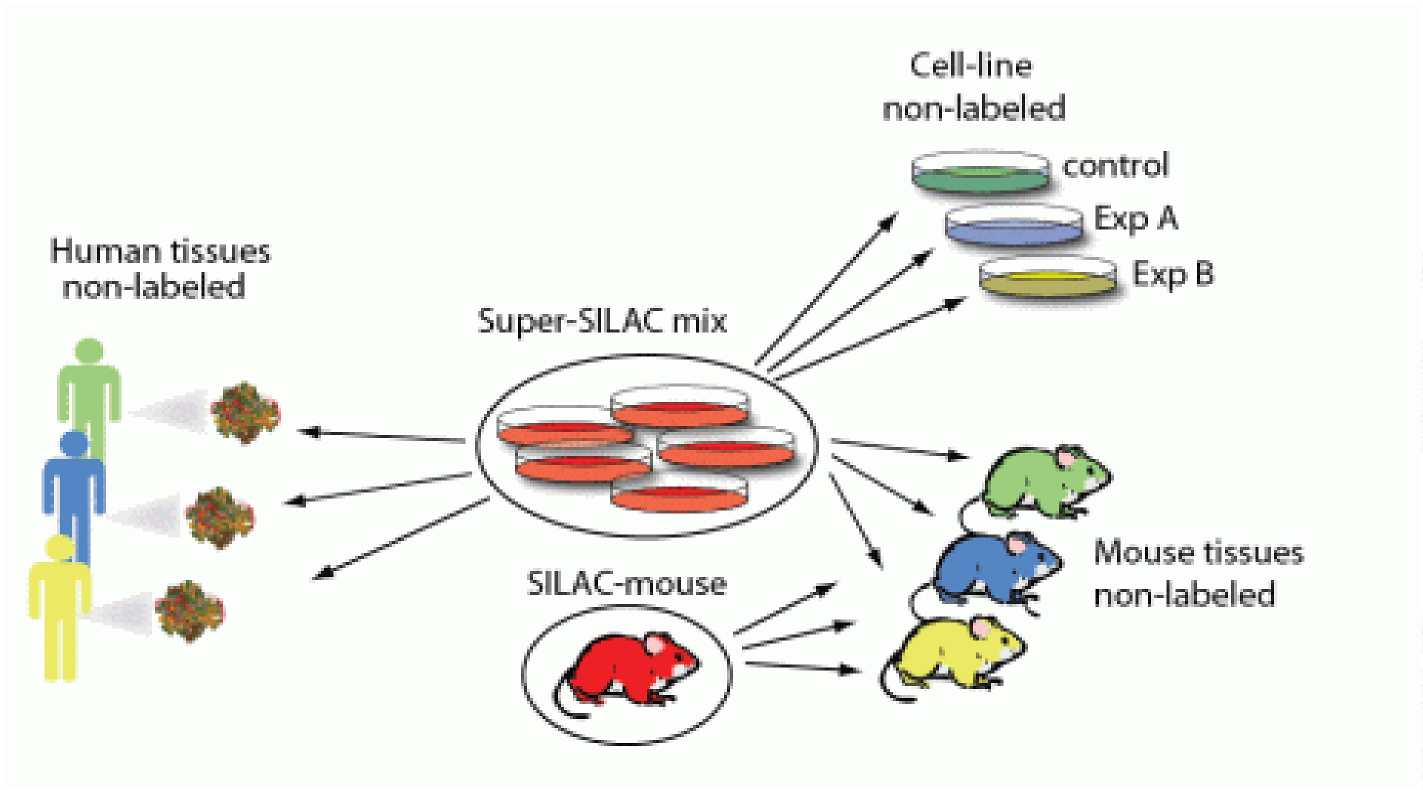


$^{13}C_6$-Lysine

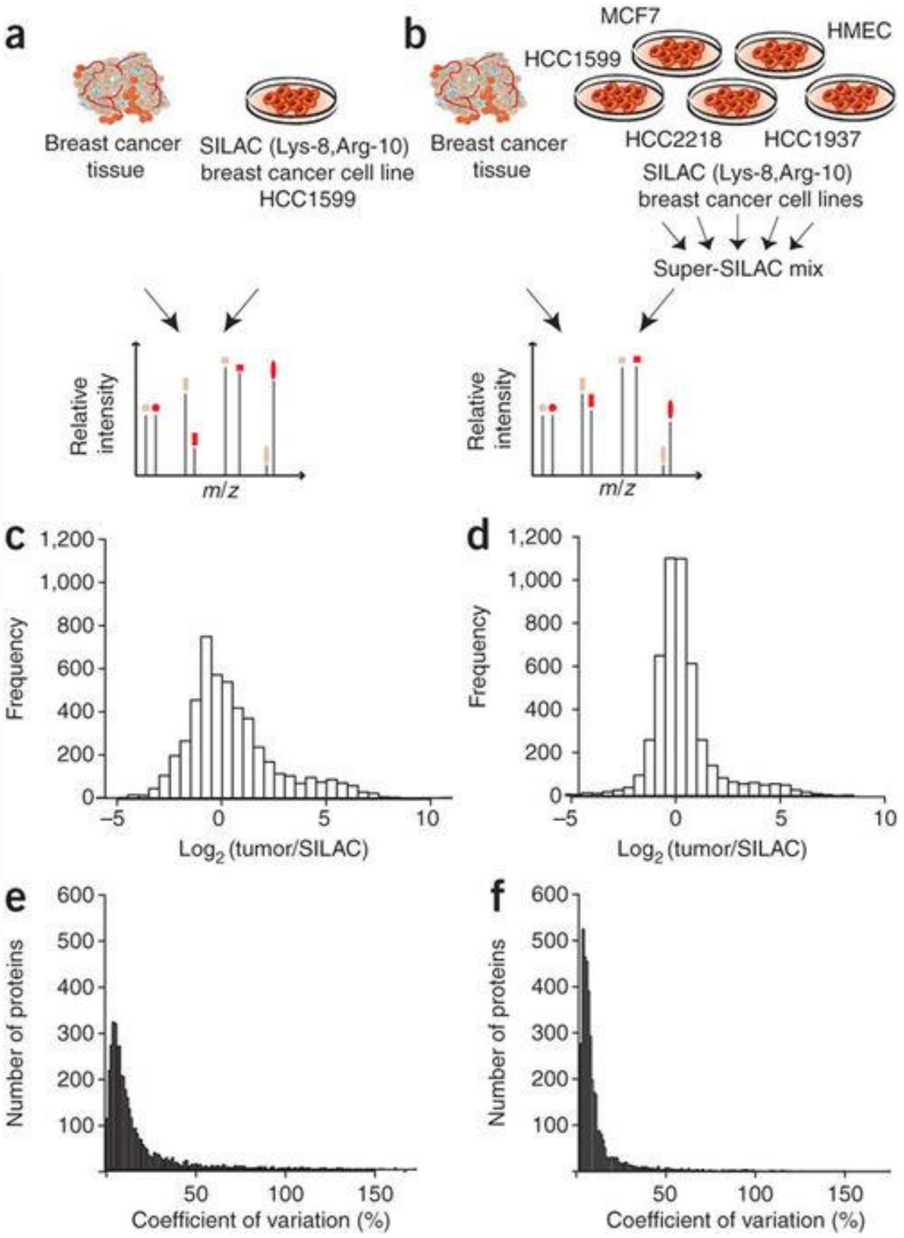


DOI: https://doi.org/10.1016/j.cell.2008.05.033

Super-SILAC

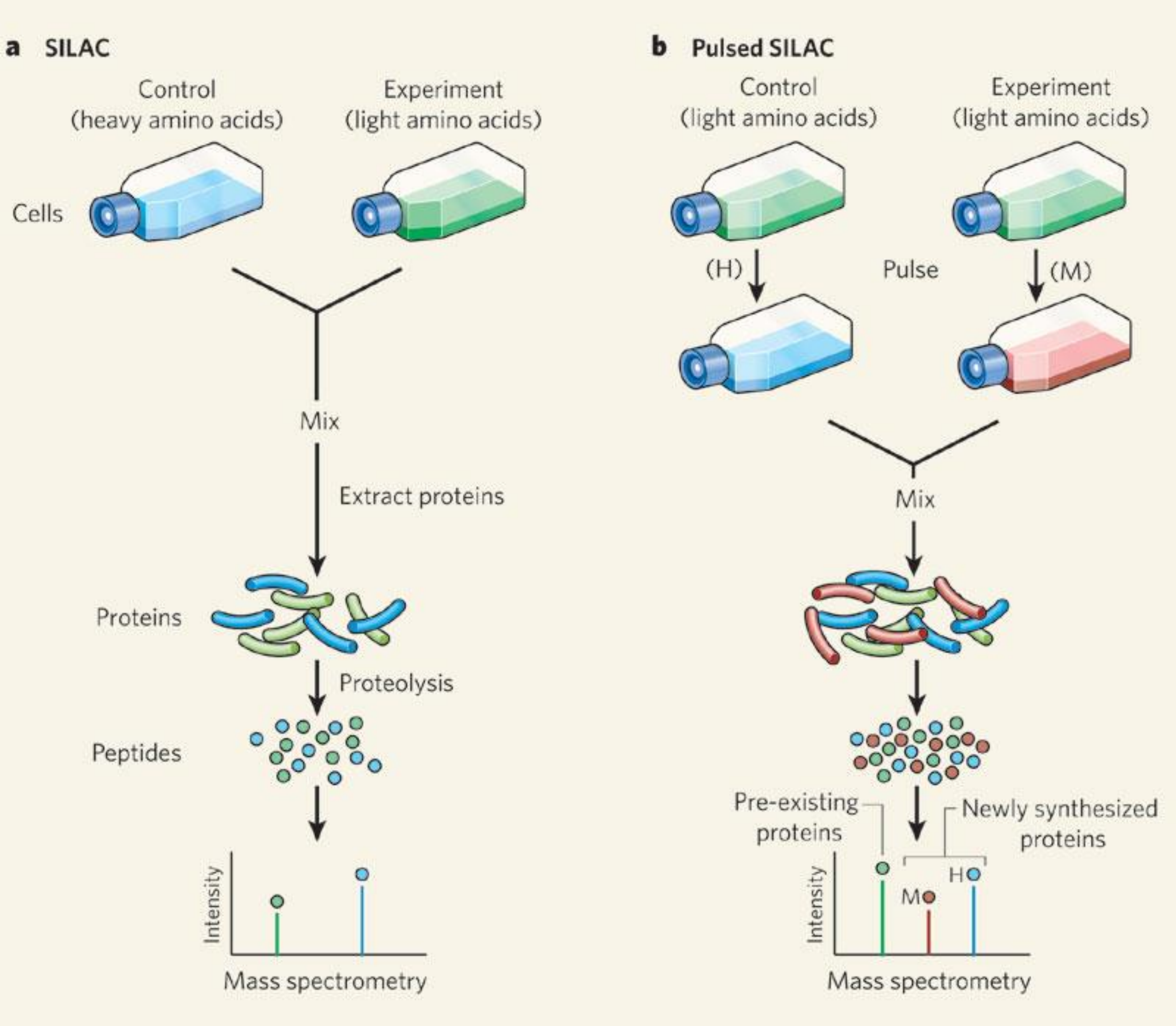


http://www.biochem.mpg.de/221777/SILAC

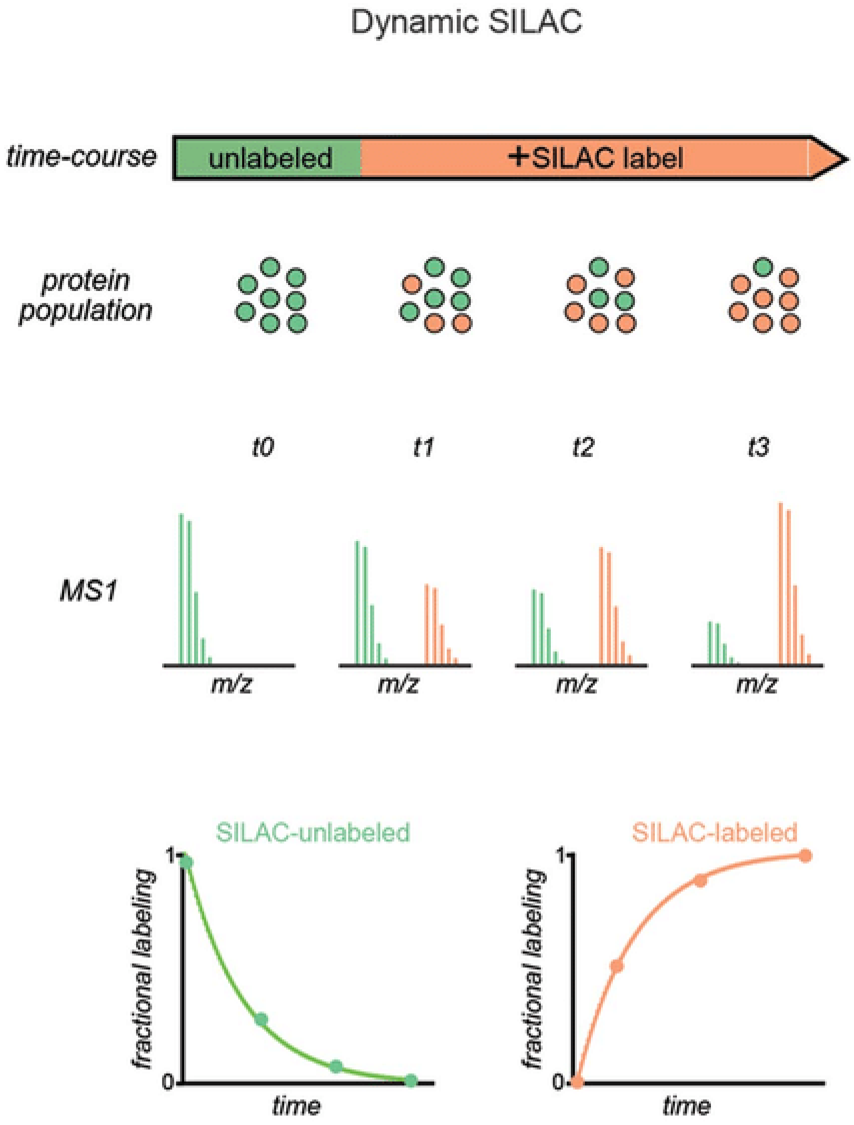


doi:10.1038/nmeth.1446

Pulsed and dynamic SILAC

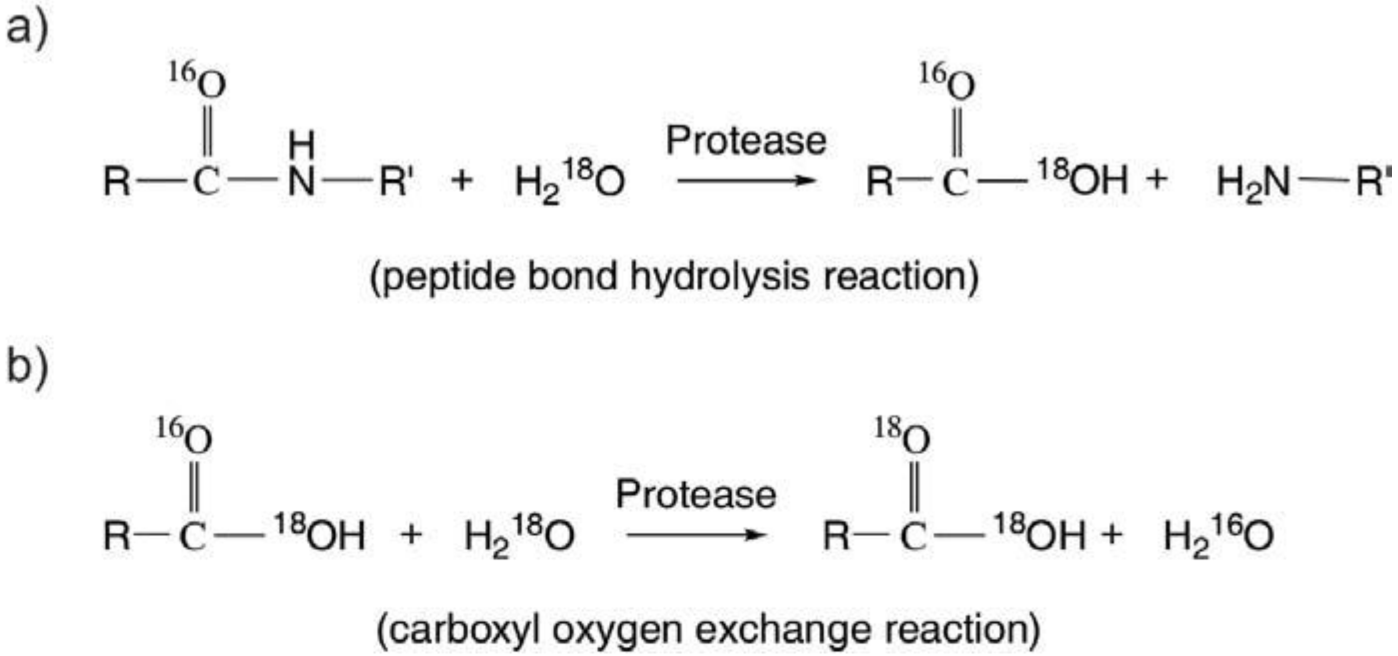


DOI: 10.1038/455044a



DOI: 10.1074/mcp.M116.063230

Proteolytic labelling



a)

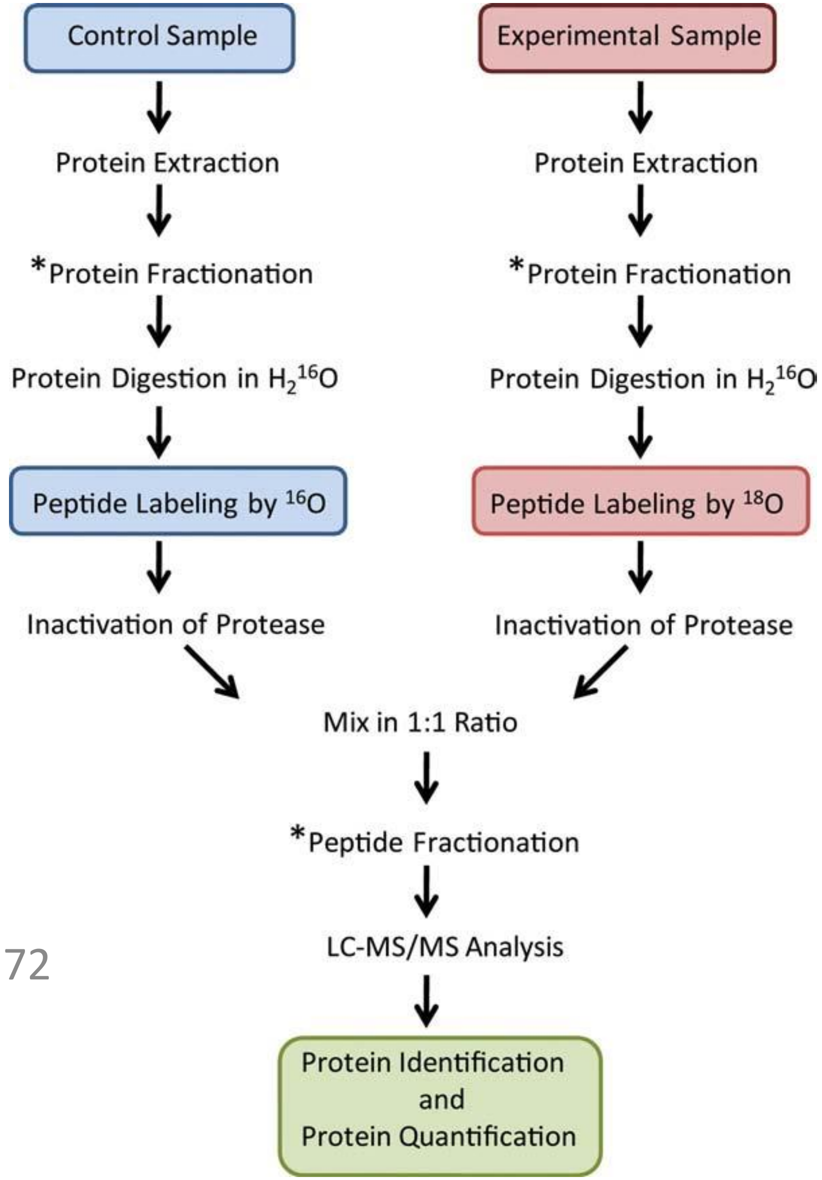
$$\mathrm{R{-}\overset{\overset{^{16}O}{\|}}{C}{-}\overset{H}{N}{-}R' + H_2{}^{18}O \xrightarrow{Protease} R{-}\overset{\overset{^{16}O}{\|}}{C}{-}{}^{18}OH + H_2N{-}R'}$$

(peptide bond hydrolysis reaction)

b)

$$\mathrm{R{-}\overset{\overset{^{16}O}{\|}}{C}{-}{}^{18}OH + H_2{}^{18}O \xrightarrow{Protease} R{-}\overset{\overset{^{18}O}{\|}}{C}{-}{}^{18}OH + H_2{}^{16}O}$$

(carboxyl oxygen exchange reaction)

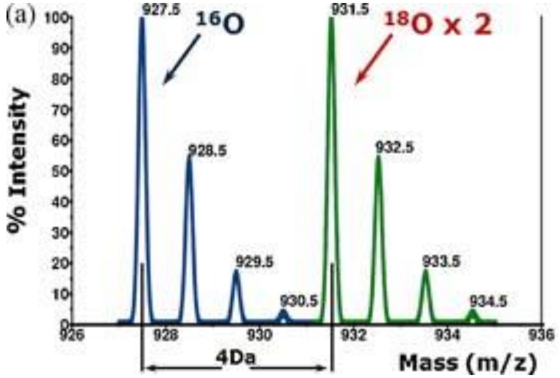


Control Sample → Protein Extraction → *Protein Fractionation → Protein Digestion in $H_2{}^{16}O$ → Peptide Labeling by ^{16}O → Inactivation of Protease

Experimental Sample → Protein Extraction → *Protein Fractionation → Protein Digestion in $H_2{}^{16}O$ → Peptide Labeling by ^{18}O → Inactivation of Protease

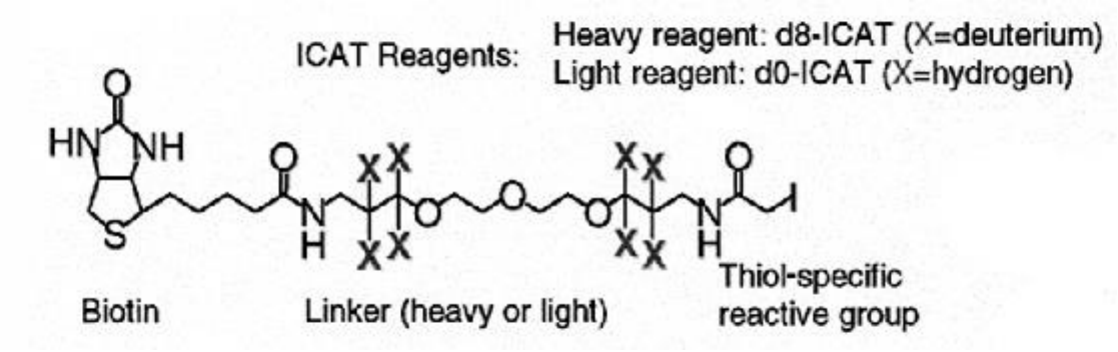
Mix in 1:1 Ratio → *Peptide Fractionation → LC-MS/MS Analysis → Protein Identification and Protein Quantification

DOI: 10.2174/157016411794697372

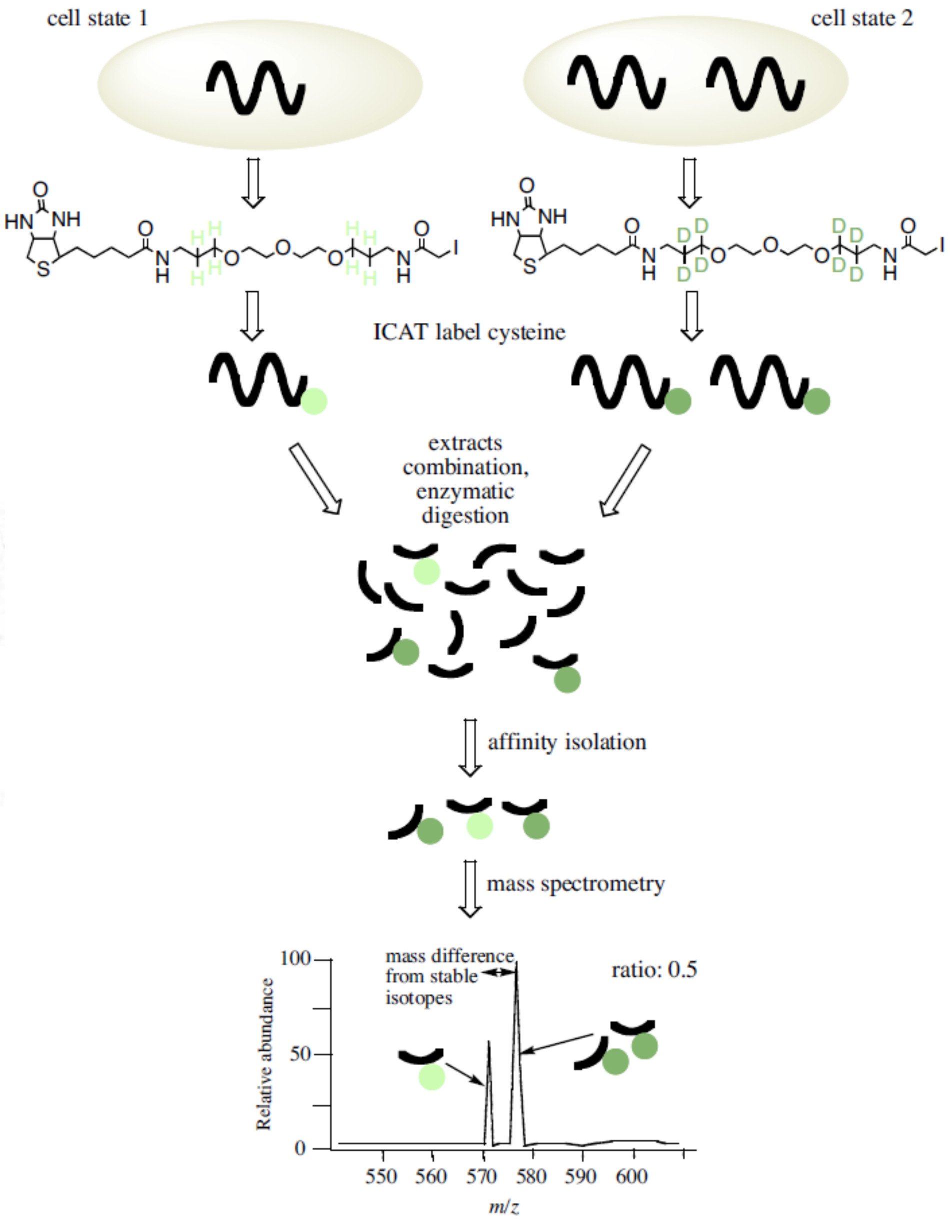


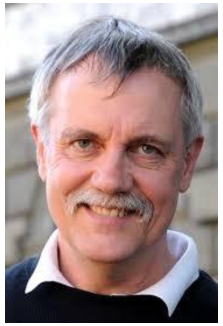
DOI: 10.1002/rcm.4816

Chemical labelling



DOI: 10.1038/13690





Ruedi Aebersold

Steven Gygi

https://infoscience.epfl.ch/record/83512/files/EPFL_TH3525.pdf

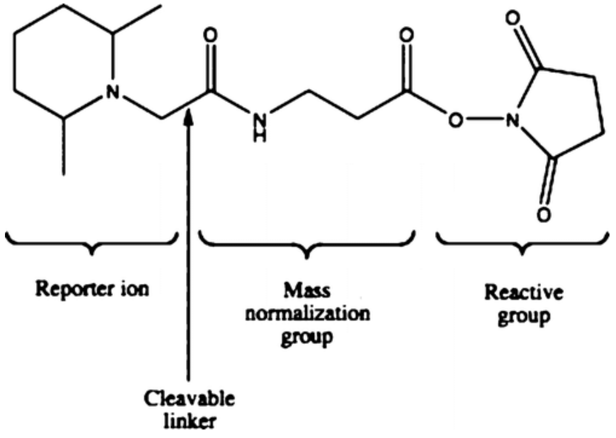
Chemical labelling: Isobaric tagging

Tandem mass tags (TMTs)

Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS

Andrew Thompson,[†] Jürgen Schäfer,[‡] Karsten Kuhn,[‡] Stefan Kienle,[‡] Josef Schwarz,[‡]
Günter Schmidt,[†] Thomas Neumann,[‡] and Christian Hamon*,[‡]

DOI: 10.1021/ac0262560



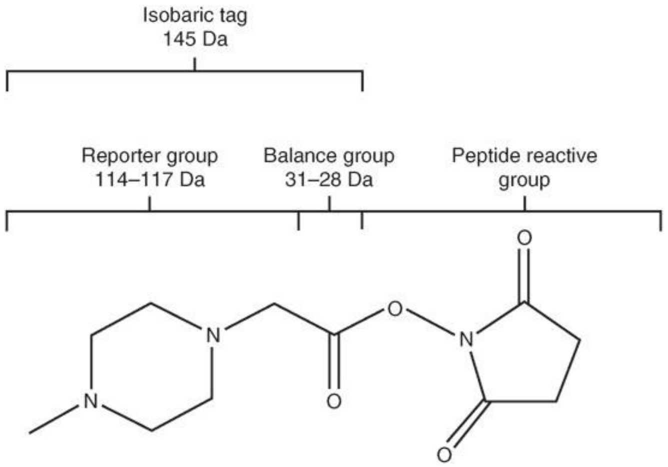
DOI: 10.1021/ac702422x

Isobaric tags for relative and absolute quantitation (iTRAQ)

Multiplexed Protein Quantitation in *Saccharomyces cerevisiae* Using Amine-reactive Isobaric Tagging Reagents*

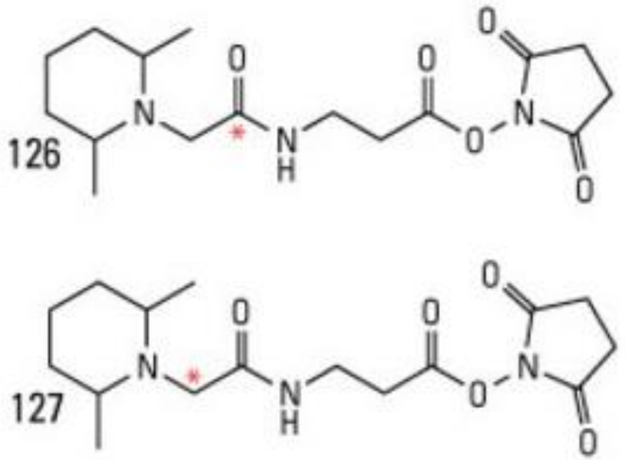
Philip L. Ross‡, Yulin N. Huang‡, Jason N. Marchese‡, Brian Williamson‡,
Kenneth Parker‡, Stephen Hattan‡, Nikita Khainovski‡, Sasi Pillai‡, Subhakar Dey‡,
Scott Daniels‡, Subhasish Purkayastha‡, Peter Juhasz‡, Stephen Martin‡,
Michael Bartlet-Jones§, Feng He¶, Allan Jacobson¶, and Darryl J. Pappin‡‖**

DOI: 10.1074/mcp.M400129-MCP200



DOI: 10.1038/nprot.2006.254

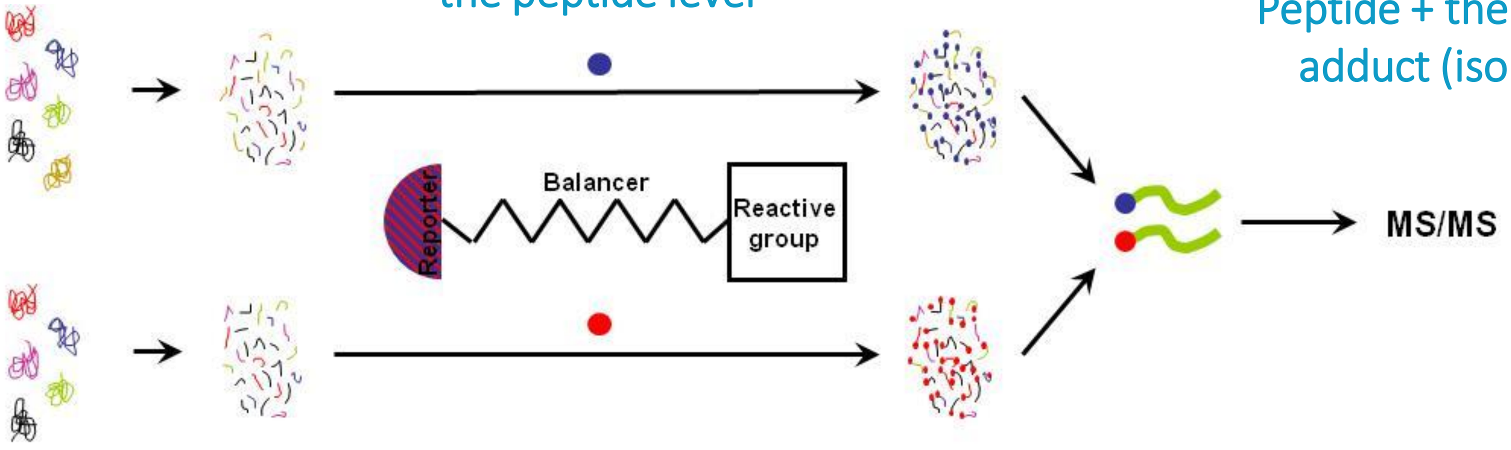
Chemical labelling: Isobaric tagging

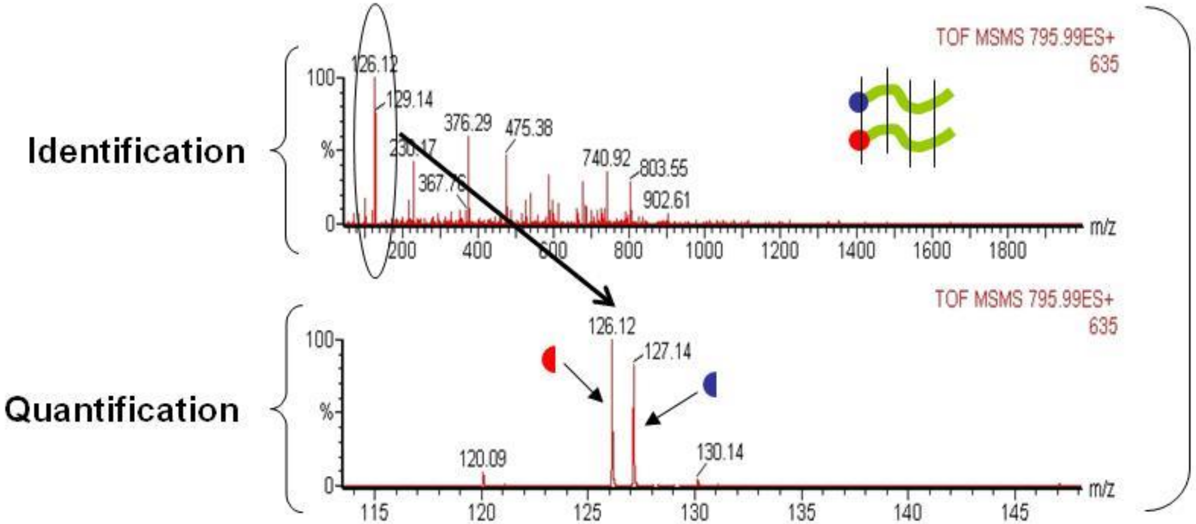


https://www.thermofisher.com

Differential labelling at the peptide level

Peptide + the same mass adduct (isobaric tag)





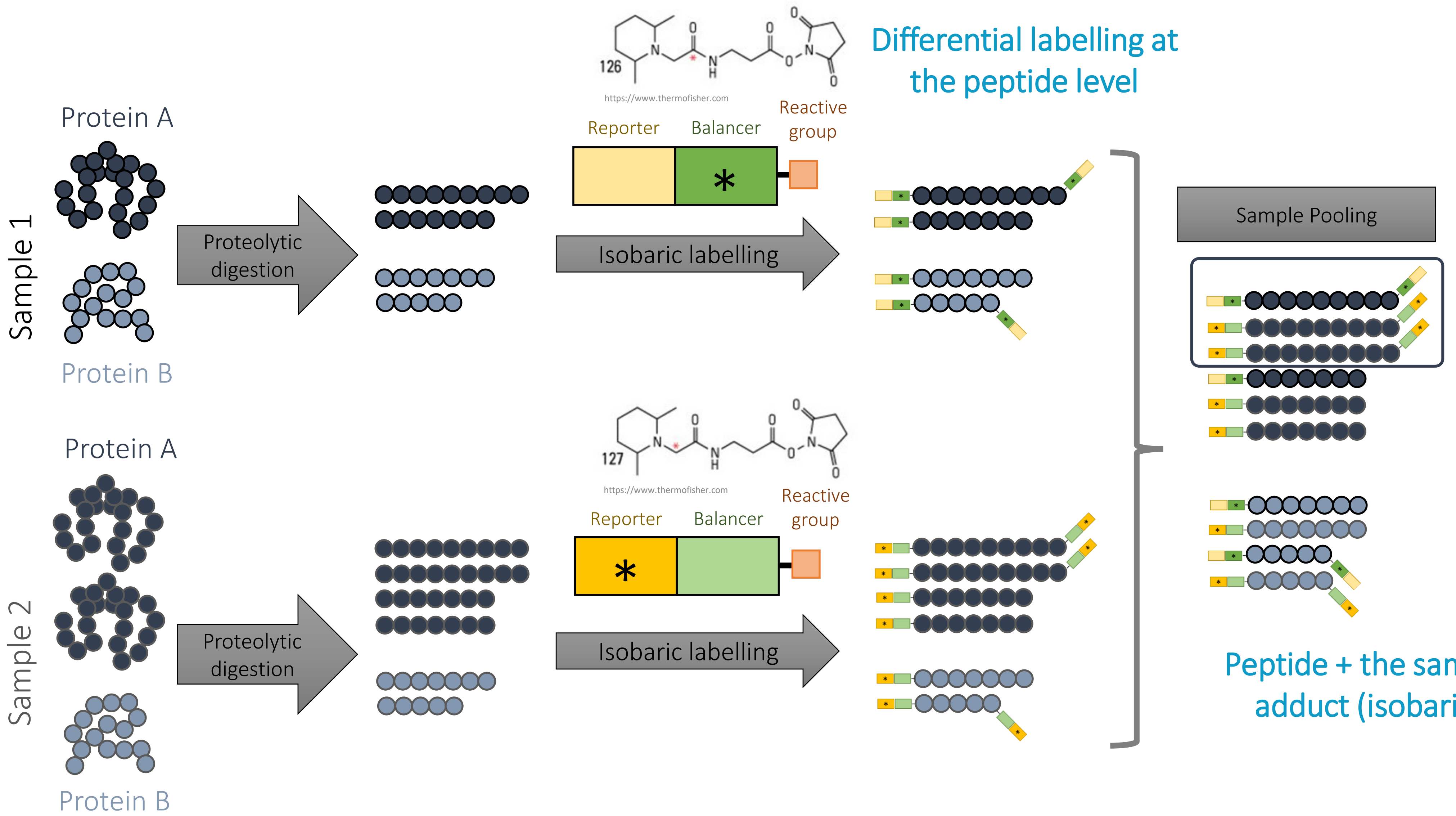
Peptide fragments EQUAL

Tandem mass spectrum

Reporter ions DIFFERENT

Dayon et al., Chimia, 2010, *64*, 132–135

Chemical labelling: Isobaric tagging - Sample preparation



Differential labelling at the peptide level

Sample 1

Protein A

Protein B

Proteolytic digestion

Reporter

Balancer

Reactive group

Isobaric labelling

Sample 2

Protein A

Protein B

Proteolytic digestion

Reporter

Balancer

Reactive group

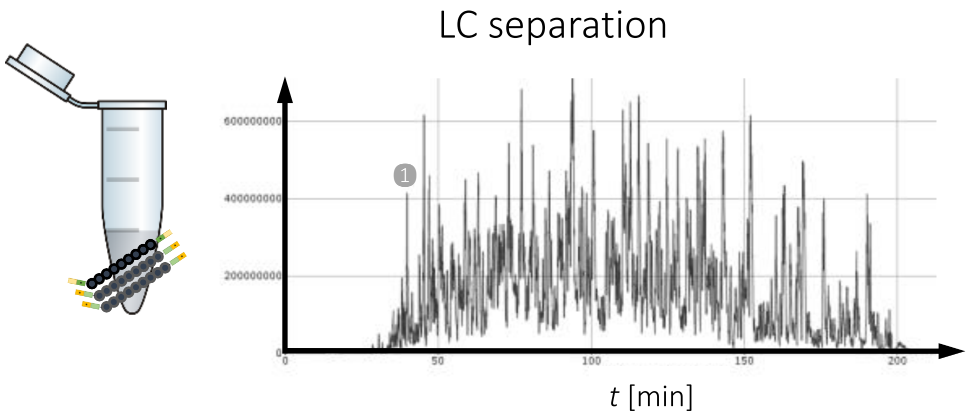
Isobaric labelling

Sample Pooling

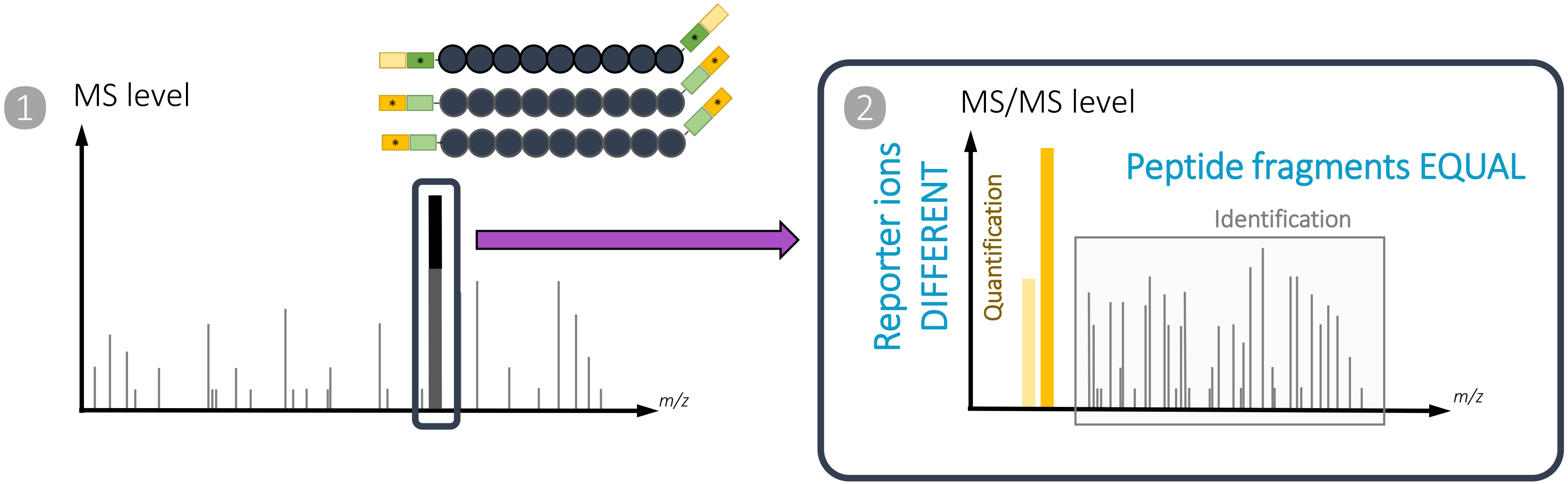
Peptide + the same mass adduct (isobaric tag)

Dayon and Affolter, Expert Rev. Proteomics, ***2020****, 17, 149-161*

Chemical labelling: Isobaric tagging - LC-MS and MS/MS analysis


LC separation

t [min]


① MS level

② MS/MS level

Reporter ions DIFFERENT

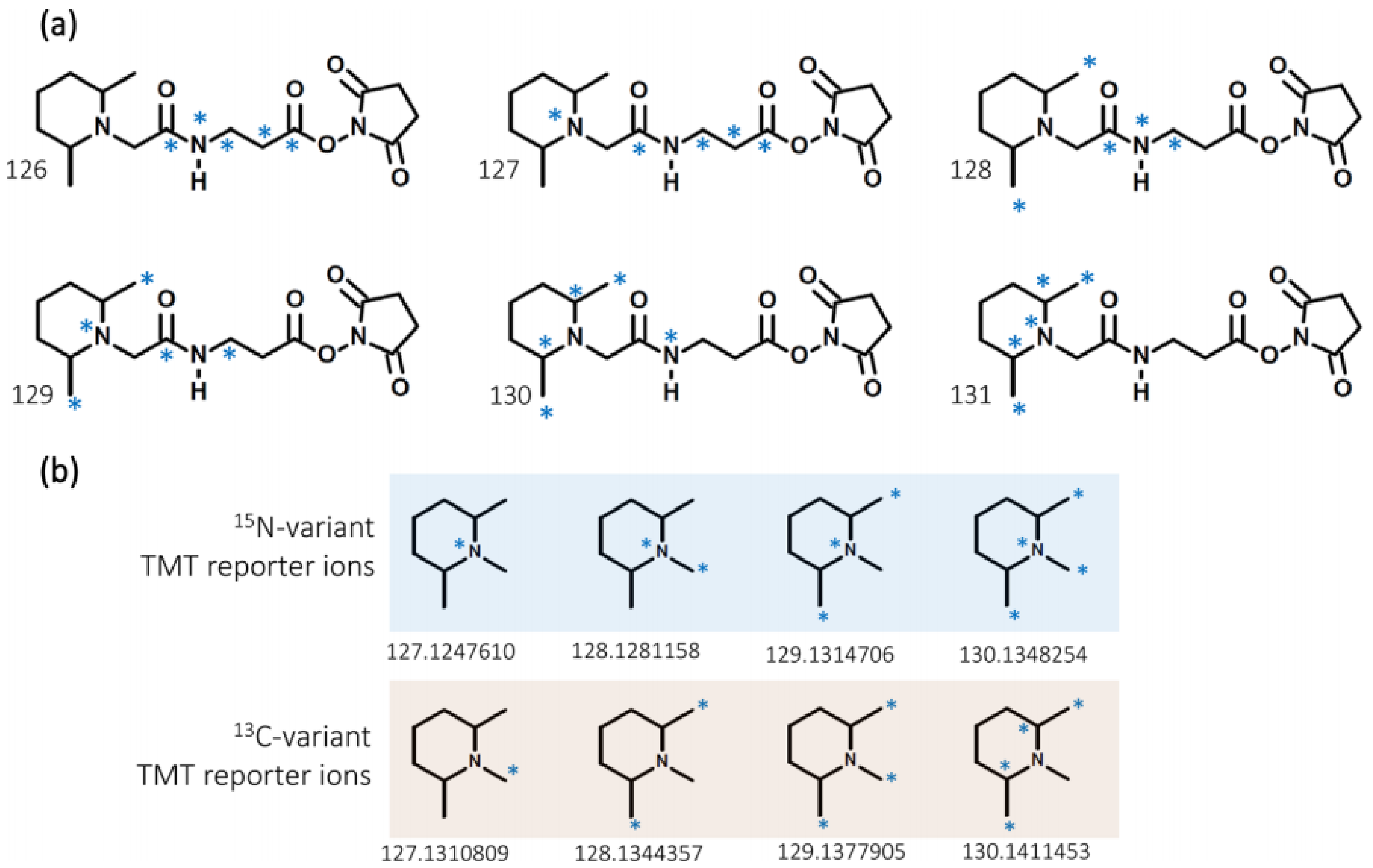
Quantification

Peptide fragments EQUAL

Identification

Dayon and Affolter, Expert Rev. Proteomics, ***2020****, 17, 149-161*

Tandem mass tags



DOI: 10.1021/pr500880b

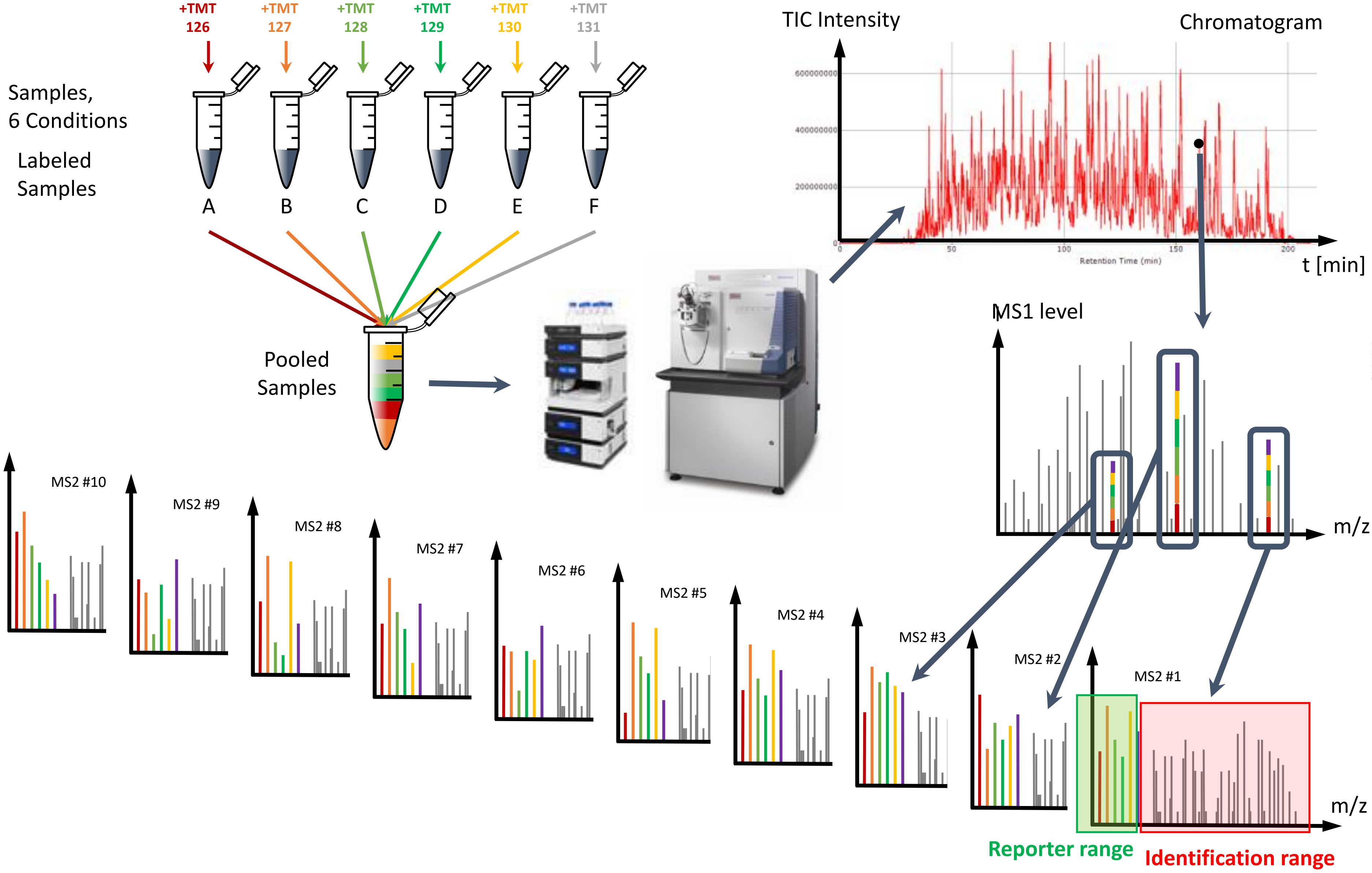
Targets:

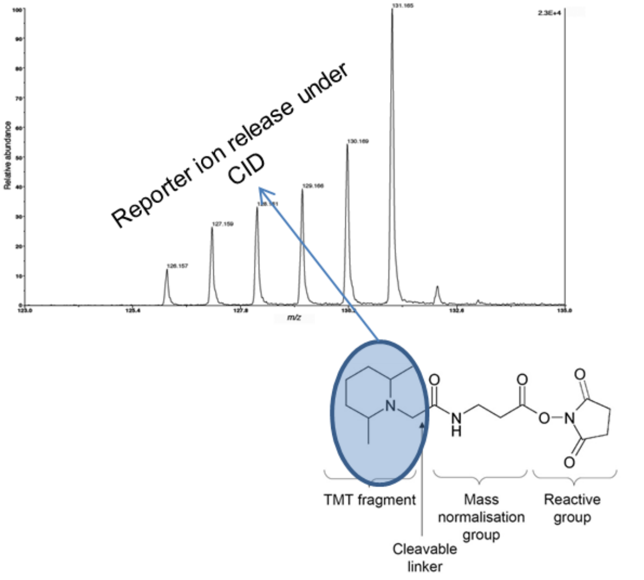
- N-terminus
- Lysine

Multiplexing:

- 2-plex
- 6-plex
- 10-plex
- 11-plex
- 16-plex…

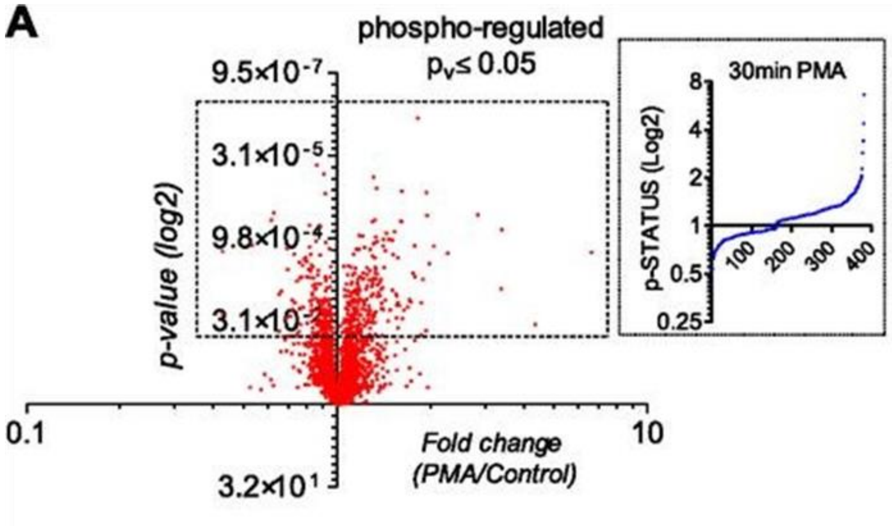
TMT technology



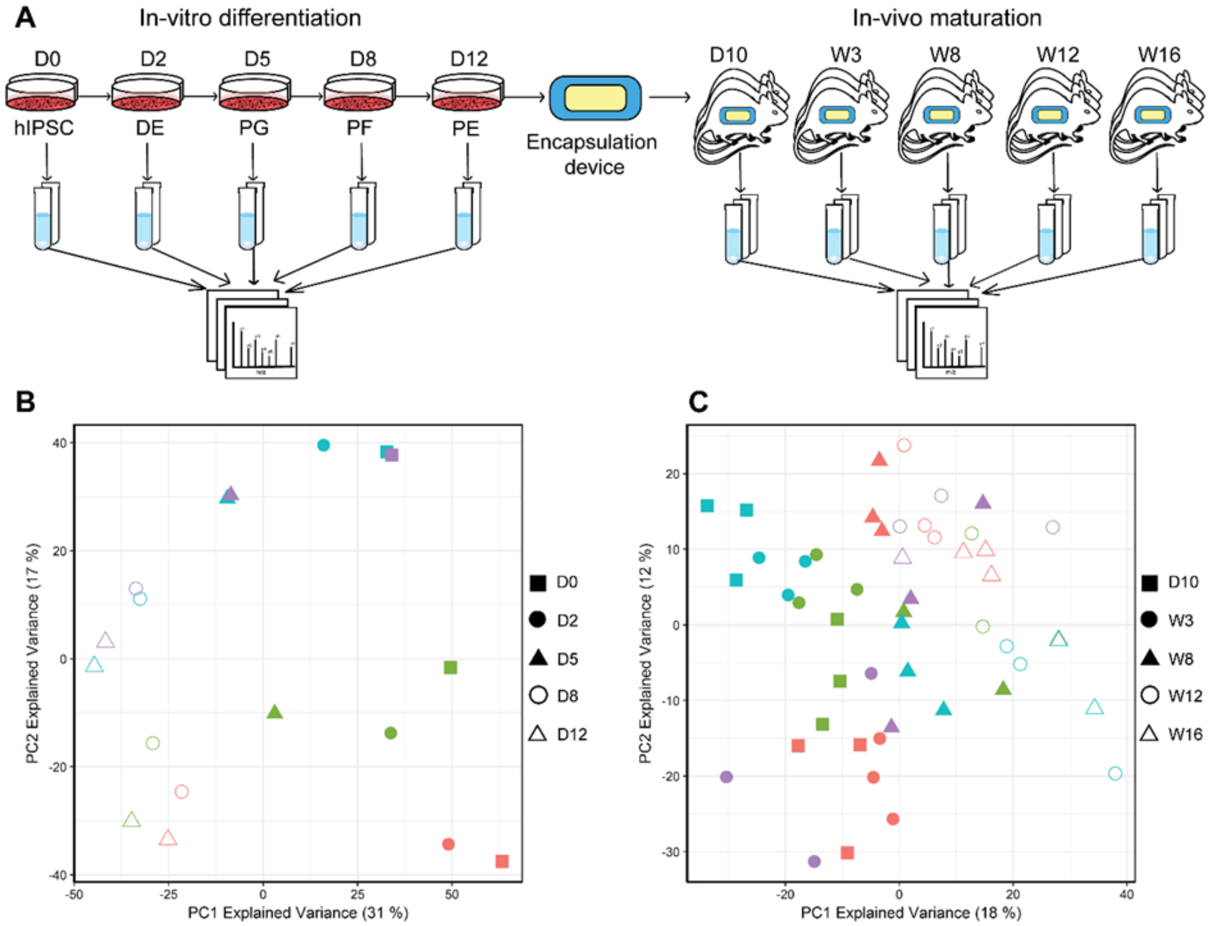


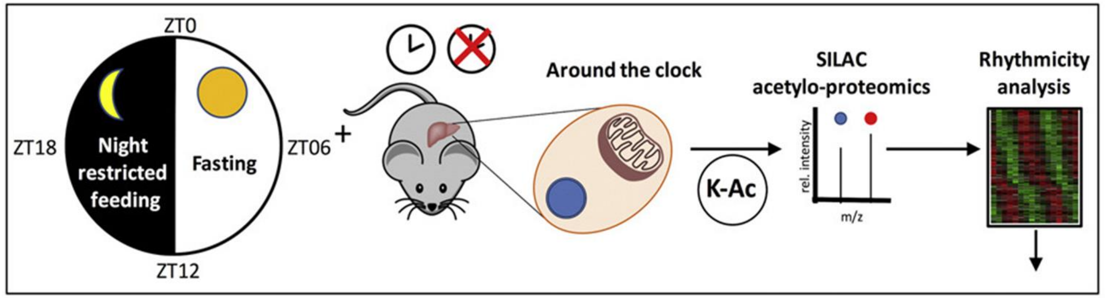
Dayon et al., Anal. Chem., **2008**, *80* (8), 2921–2931

Main applications: Response to *stimuli*



DOI: 10.1096/fj.201600837R

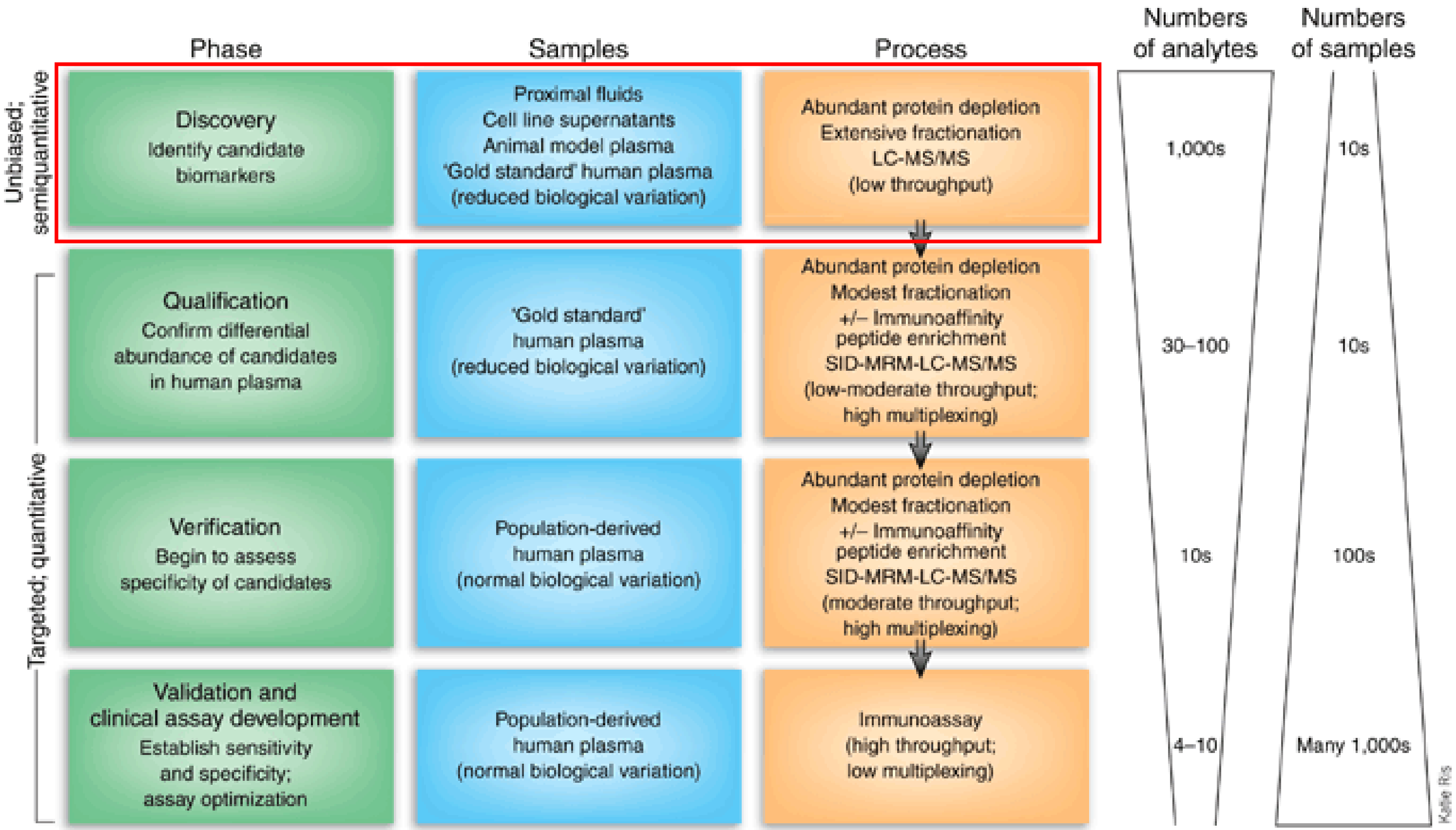




DOI: 10.1016/j.celrep.2017.07.065

DOI: 10.1002/prca.201600173

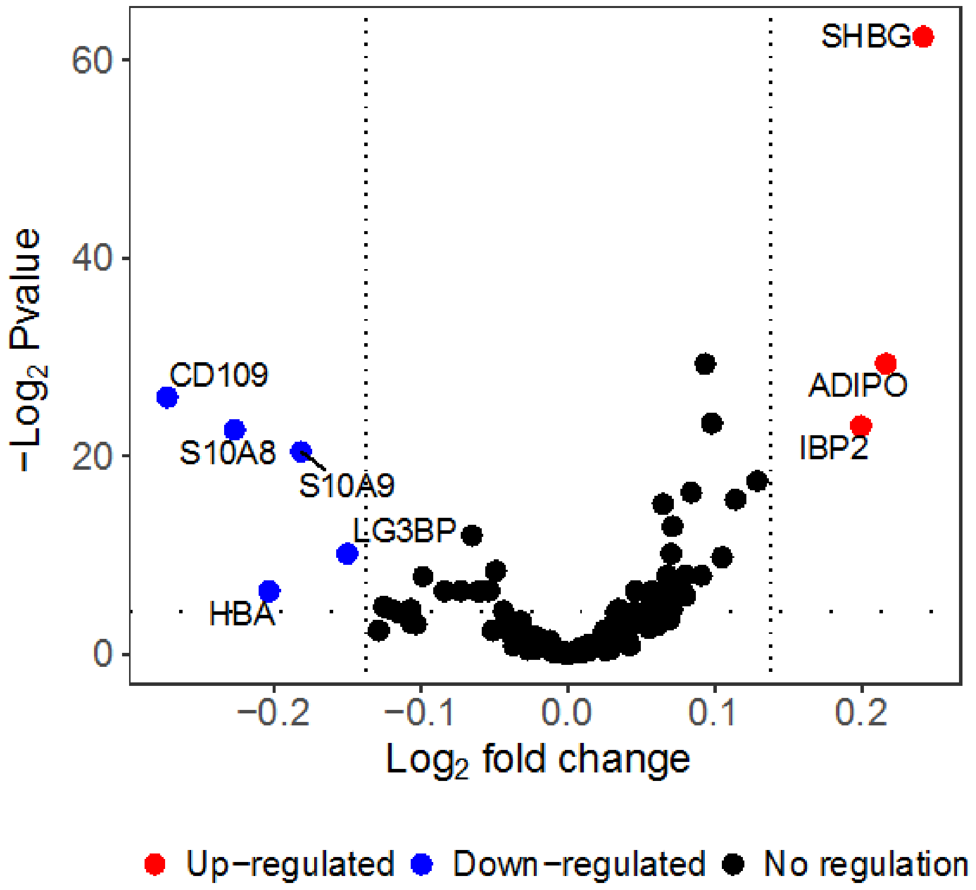
Main applications: Discovery of biomarkers

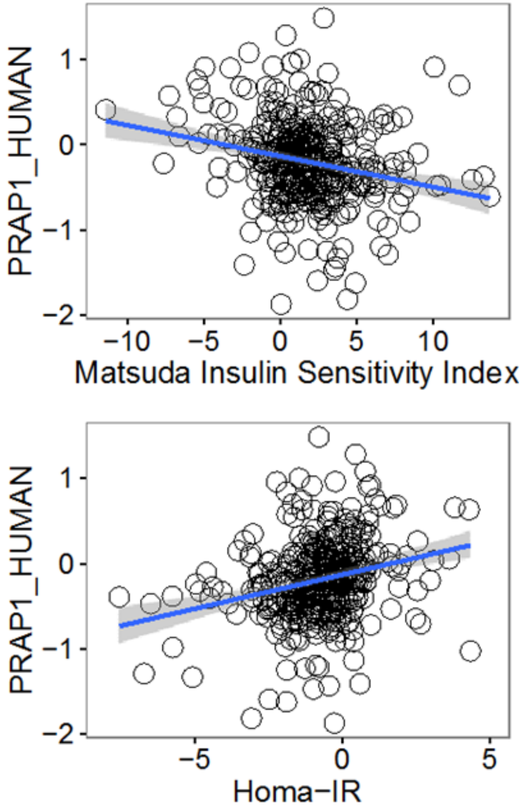


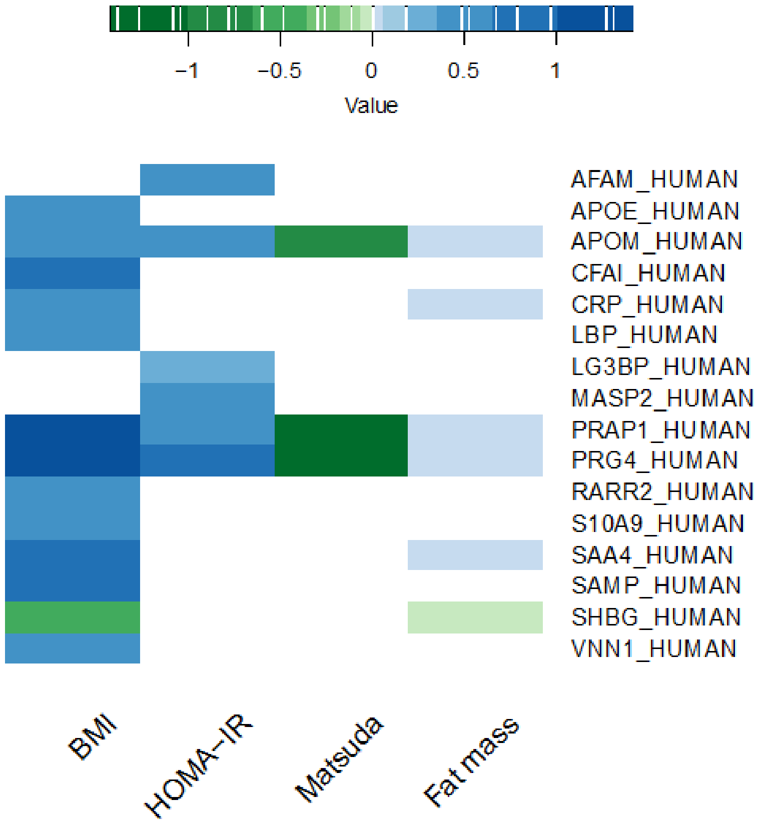
Phase	Samples	Process	Numbers of analytes	Numbers of samples
Discovery Identify candidate biomarkers	Proximal fluids Cell line supernatants Animal model plasma 'Gold standard' human plasma (reduced biological variation)	Abundant protein depletion Extensive fractionation LC-MS/MS (low throughput)	1,000s	10s
Qualification Confirm differential abundance of candidates in human plasma	'Gold standard' human plasma (reduced biological variation)	Abundant protein depletion Modest fractionation +/– Immunoaffinity peptide enrichment SID-MRM-LC-MS/MS (low-moderate throughput; high multiplexing)	30–100	10s
Verification Begin to assess specificity of candidates	Population-derived human plasma (normal biological variation)	Abundant protein depletion Modest fractionation +/– Immunoaffinity peptide enrichment SID-MRM-LC-MS/MS (moderate throughput; high multiplexing)	10s	100s
Validation and clinical assay development Establish sensitivity and specificity; assay optimization	Population-derived human plasma (normal biological variation)	Immunoassay (high throughput; low multiplexing)	4–10	Many 1,000s

DOI: 10.1038/nbt1235

Discovery of biomarkers





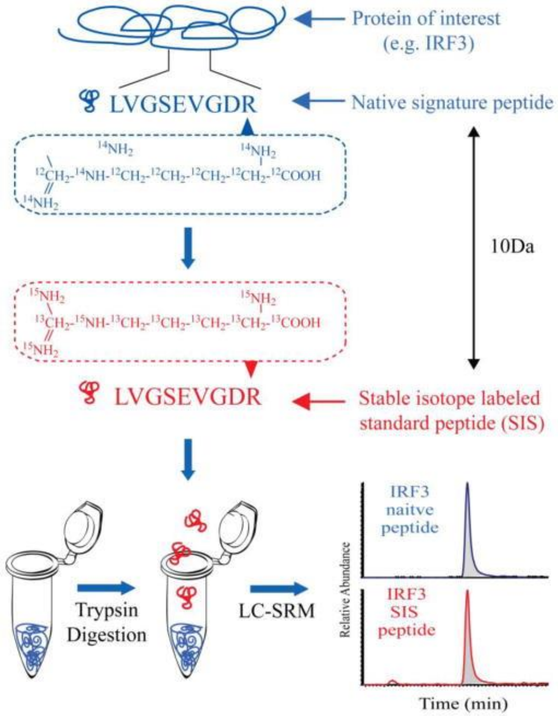


DOI: 10.1002/prca.201600150

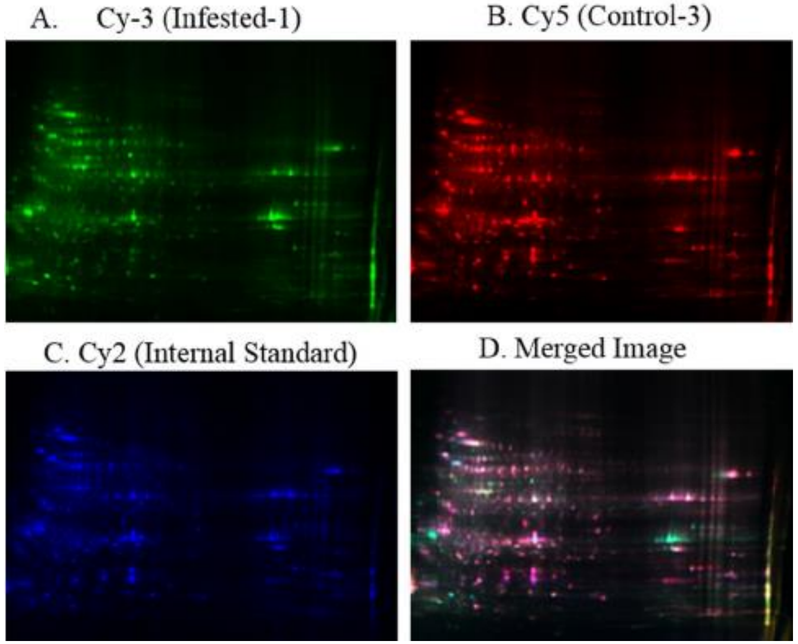
4.3. Other quantitative techniques

“Absolute” (or precise relative quantification) with targeted MS

Other techniques (*e.g.*, 2-dimensional differential gel electrophoresis (2D-DIGE))



DOI: 10.1016/j.ymeth.2013.02.001



DOI: 10.3390/ijms160819326

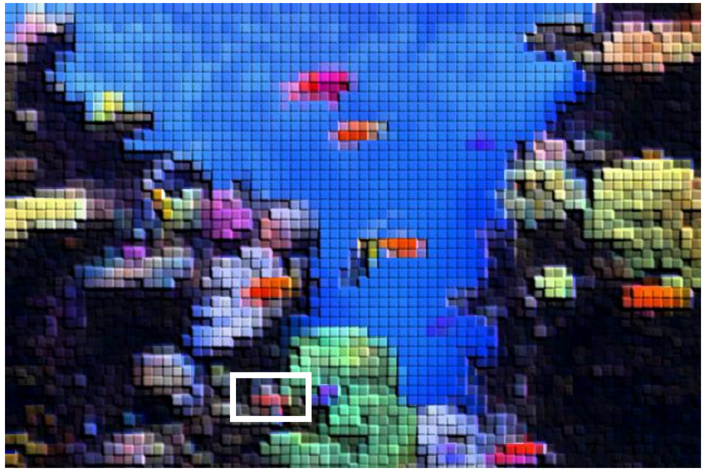
Quantitative "resolution"

Real "proteome" picture

Discovery/untargeted

Targeted

Semi/relative quantification

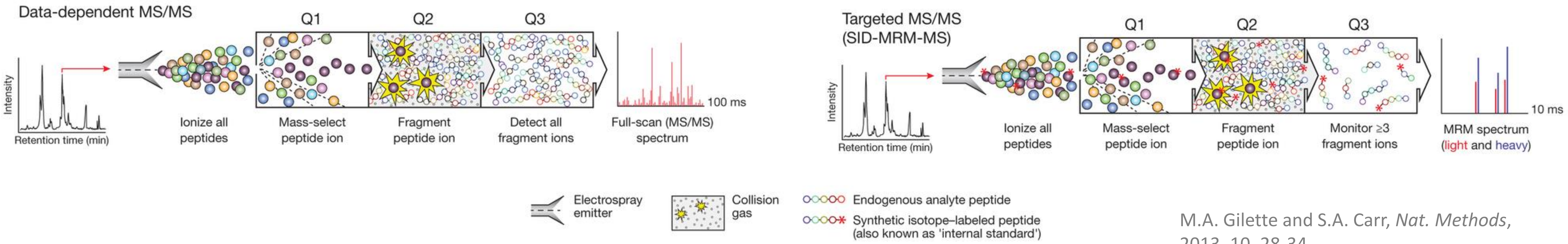


e.g., spectral counting

e.g., isobaric labelling

Precise "absolute" quantification

Discovery and Targeted MS



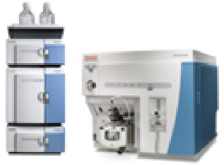
M.A. Gilette and S.A. Carr, *Nat. Methods*, 2013, 10, 28-34

ThermoFisher LTQ-Orbitrap

ThermoFisher Orbitrap LUMOS

Discovery Mode

- Relative quantification/differential display
- 1’000-10’000 protein analytes
- Specific, accurate & robust
- Moderate throughput
- Date-driven/hypothesis generation
- Plasma, CSF, tissue, cell, organelle
- Proteins and their PTMs

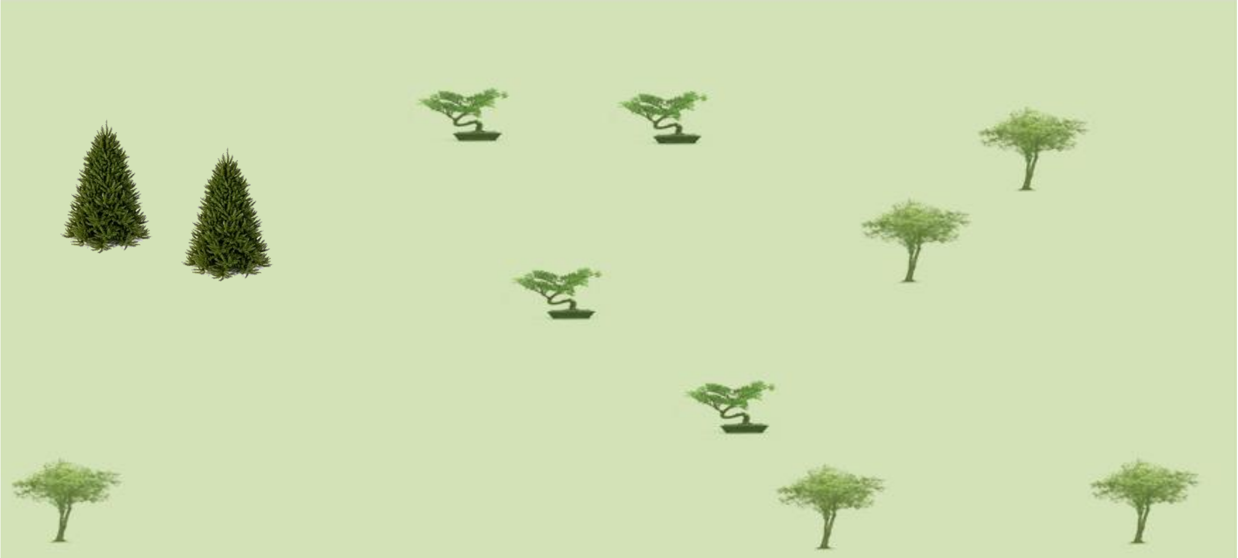


ThermoFisher TSQ Vantage

Targeted Mode

- “Absolute” quantification
- 10-100 protein analytes
- Specific, sensitive, accurate & robust
- “High” throughput
- Hypothesis-driven/hypothesis validation
- Plasma, CSF, tissue, cell, organelle
- Proteins and their PTMs

Discovery and Targeted MS

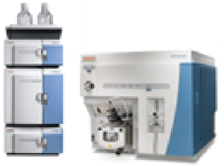


ThermoFisher LTQ-Orbitrap

ThermoFisher Orbitrap LUMOS

Discovery Mode

- Relative quantification/differential display
- 1’000-10’000 protein analytes
- Specific, accurate & robust
- Moderate throughput
- Date-driven/hypothesis generation
- Plasma, CSF, tissue, cell, organelle
- Proteins and their PTMs

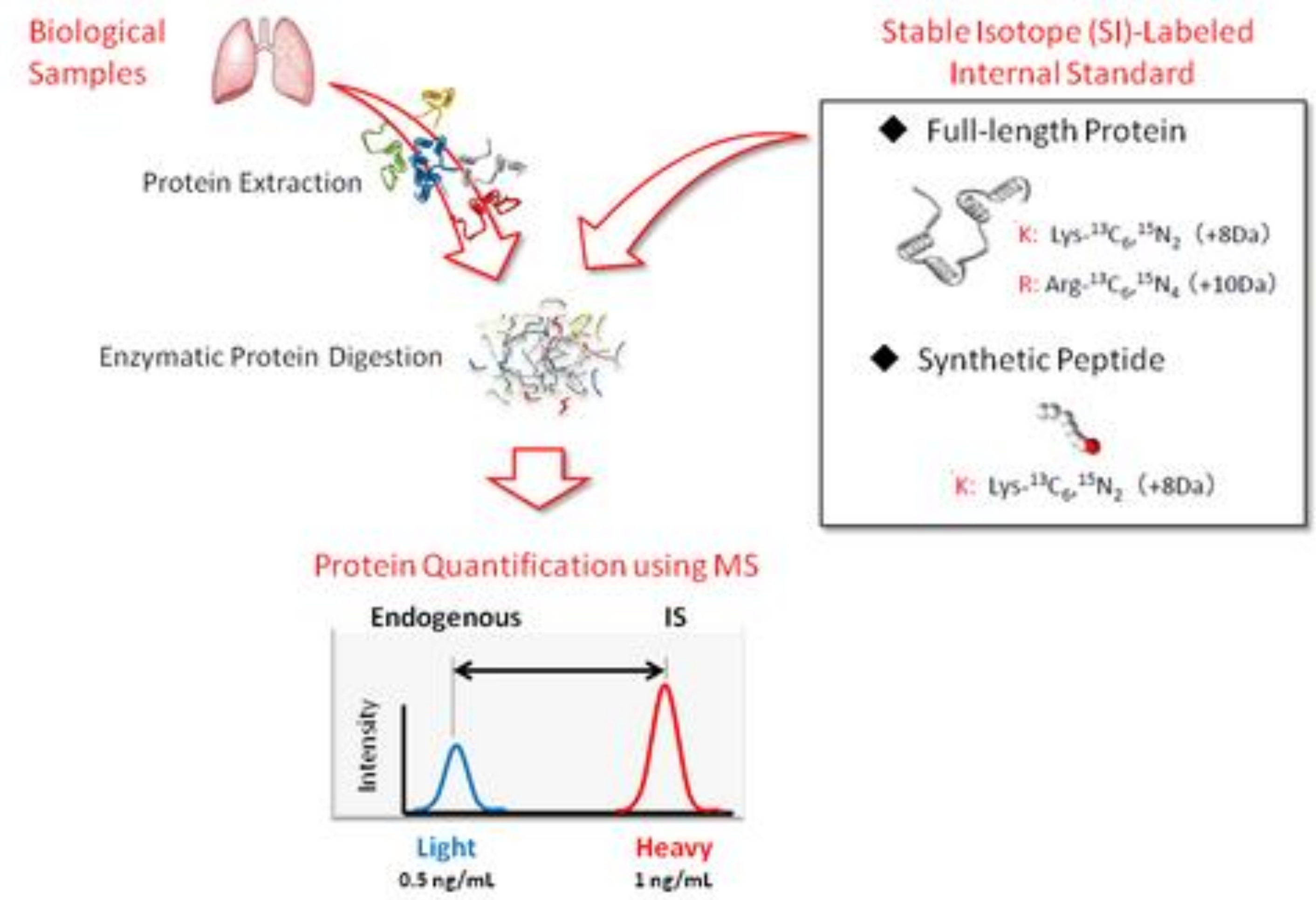


ThermoFisher TSQ Vantage

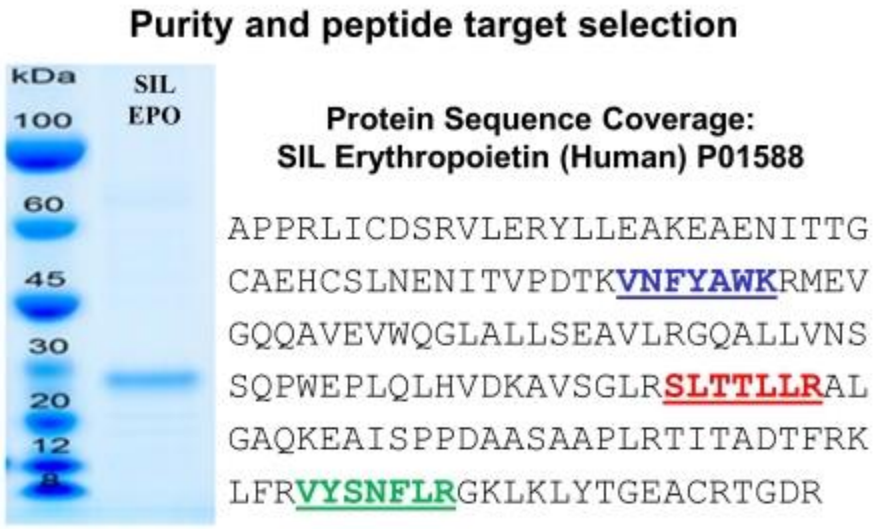
Targeted Mode

- “Absolute” quantification
- 10-100 protein analytes
- Specific, sensitive, accurate & robust
- “High” throughput
- Hypothesis-driven/hypothesis validation
- Plasma, CSF, tissue, cell, organelle
- Proteins and their PTMs

Stable isotope dilution(SID)



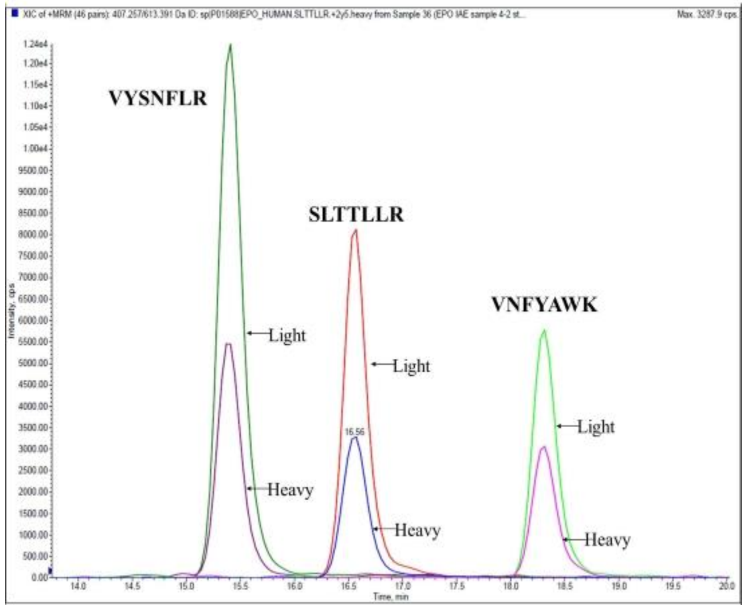
https://www.proteomicslaboratory.com/



Protein Sequence Coverage:
SIL Erythropoietin (Human) P01588

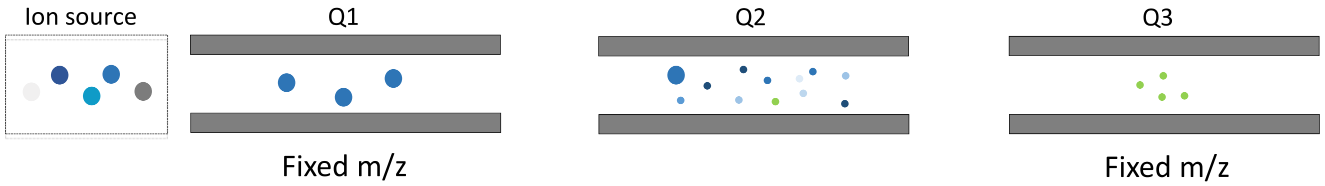
APPRLICDSRVLERYLLEAKEAENITTG
CAEHCSLNENITVPDTK**VNFYAWK**RMEV
GQQAVEVWQGLALLSEAVLRGQALLVNS
SQPWEPLQLHVDKAVSGLR**SLTTLLR**AL
GAQKEAISPPDAASAAPLRTITADTFRK
LFR**VYSNFLR**GKLKLYTGEACRTGDR

https://www.sigmaaldrich.com



https://www.sigmaaldrich.com

Selected reaction monitoring (SRM) and mSRM (MRM)



Ion source

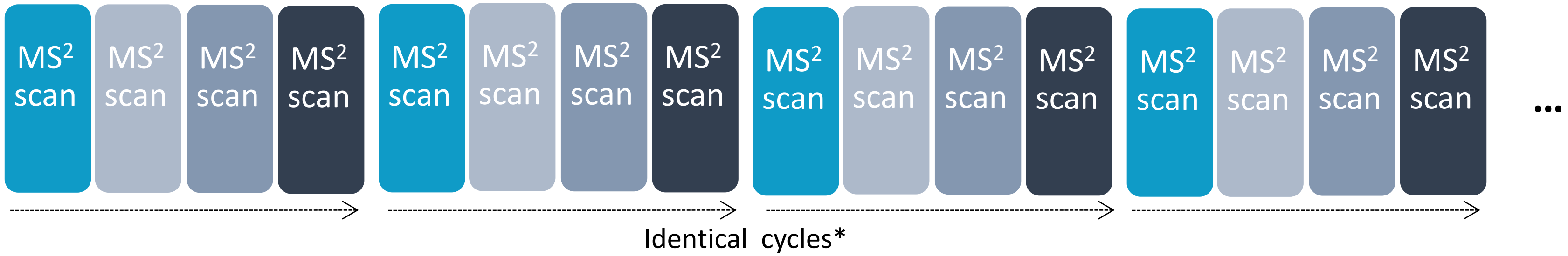
Q1

Q2

Q3

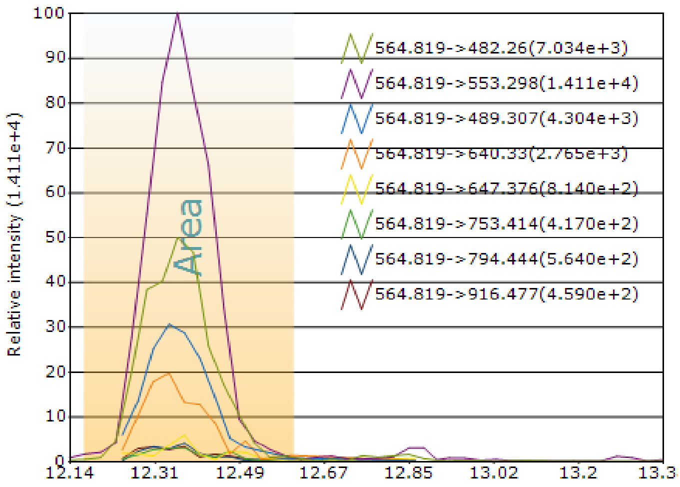
Fixed m/z

Fixed m/z



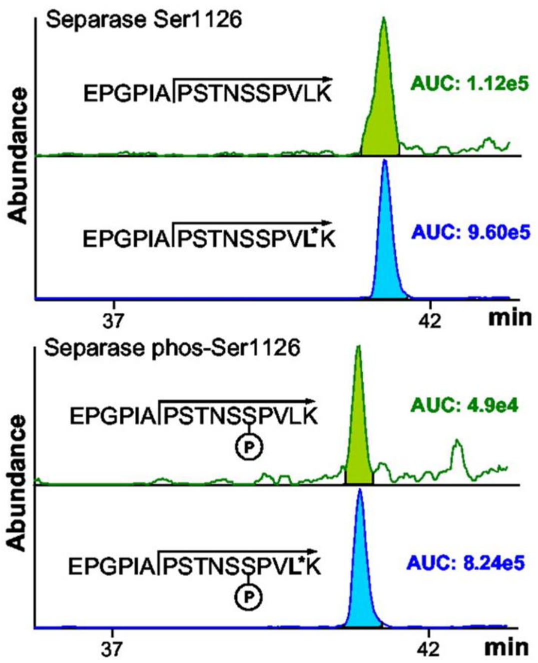
MS^2 scan MS^2 scan MS^2 scan MS^2 scan
MS^2 scan MS^2 scan MS^2 scan MS^2 scan
MS^2 scan MS^2 scan MS^2 scan MS^2 scan
MS^2 scan MS^2 scan MS^2 scan MS^2 scan …

Identical cycles*



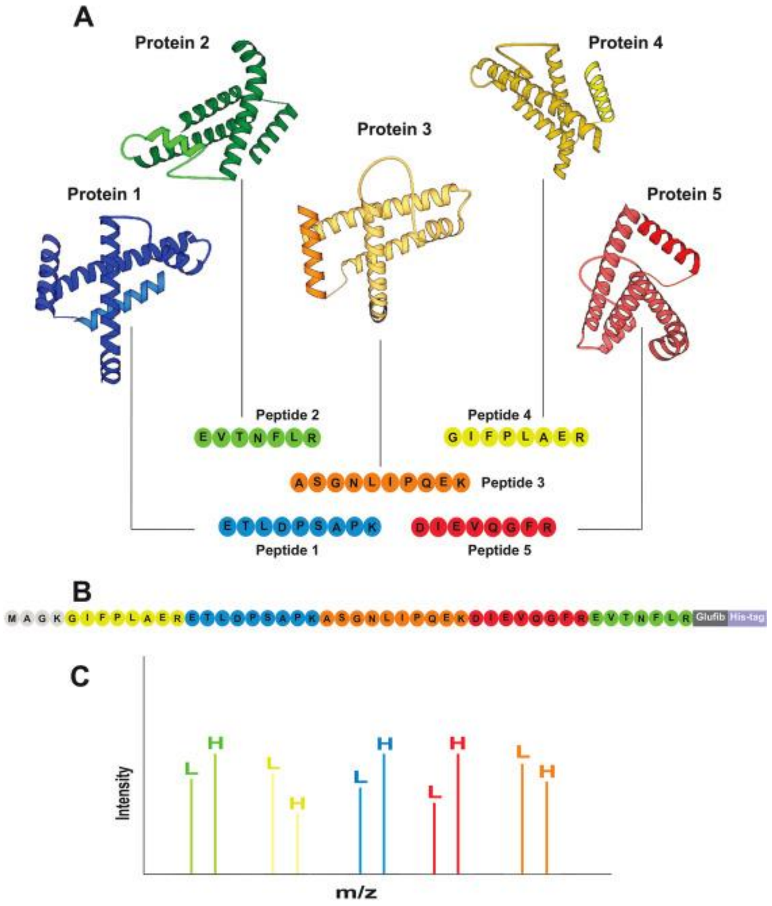
Choice of standards

AQUA internal standard peptides



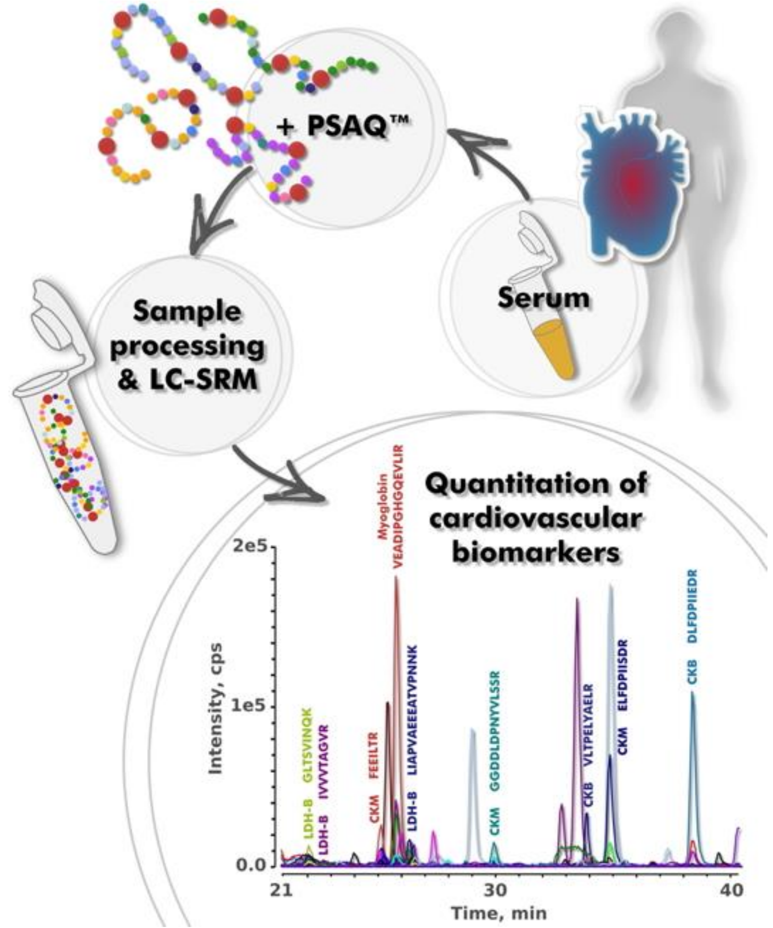
DOI: 10.1073/pnas.0832254100

Concatenations of tryptic Q-peptides (QconCATs)



DOI: 10.1016/j.ijms.2015.08.003

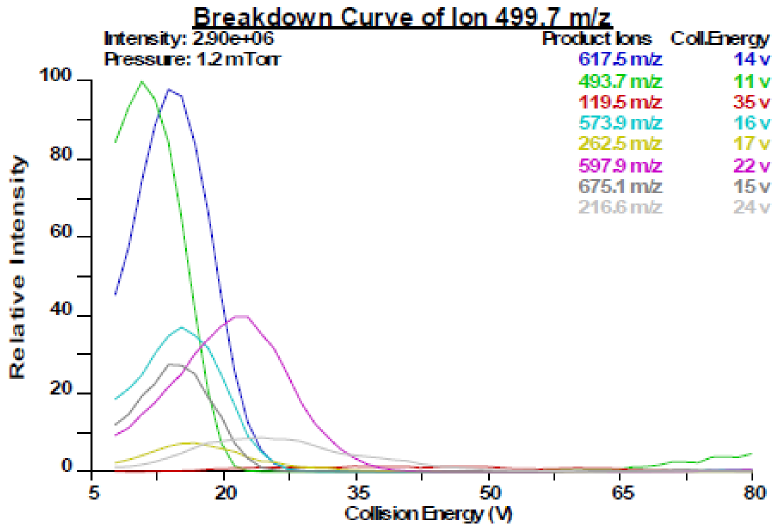
Protein standard absolute quantification (PSAQ)

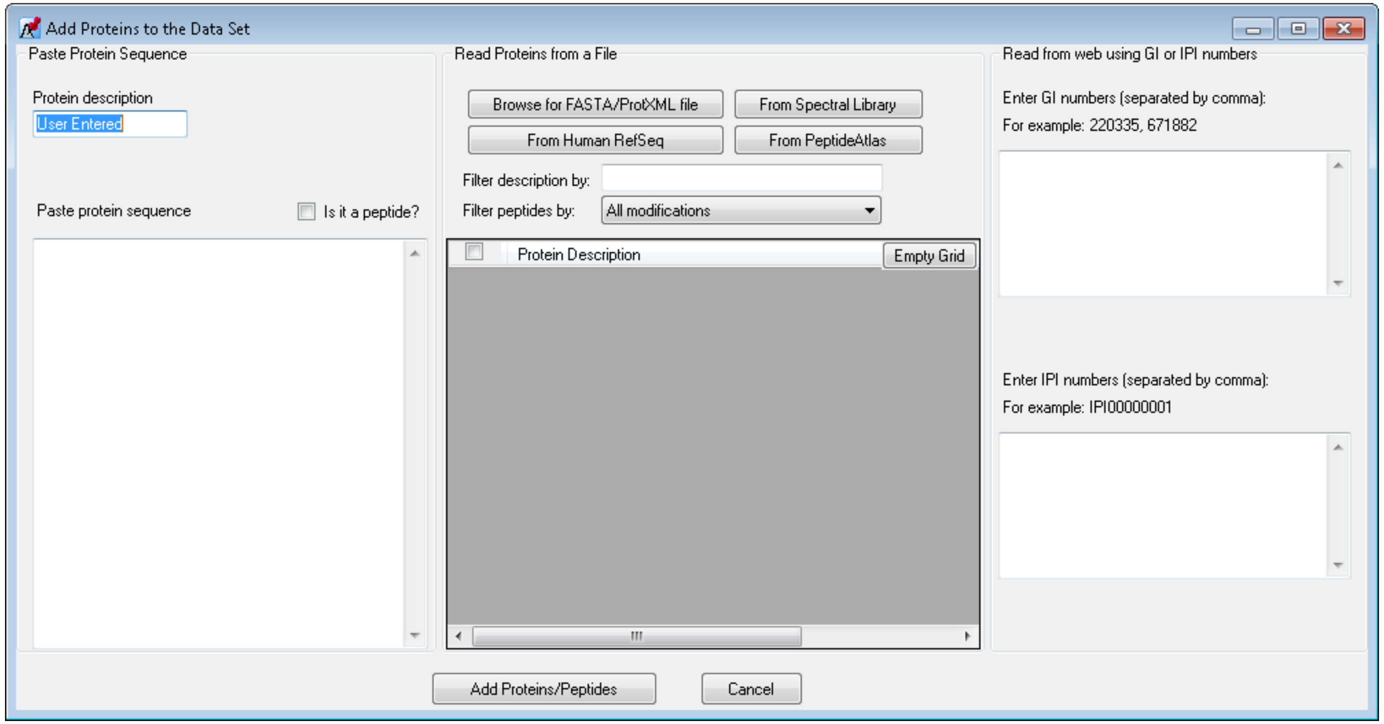


http://big.cea.fr/drf/big/english/pages/news/scientific-results/l27_vbrun.aspx

Targeted method development (1)

- Choice of peptides (http://www.uniprot.org/blast/; https://www.nextprot.org/viewers/unicity-checker/app/index.html)
- Synthesis of light and heavy peptides
- Study with MS/MS (infusion)





Protein description: User Entered

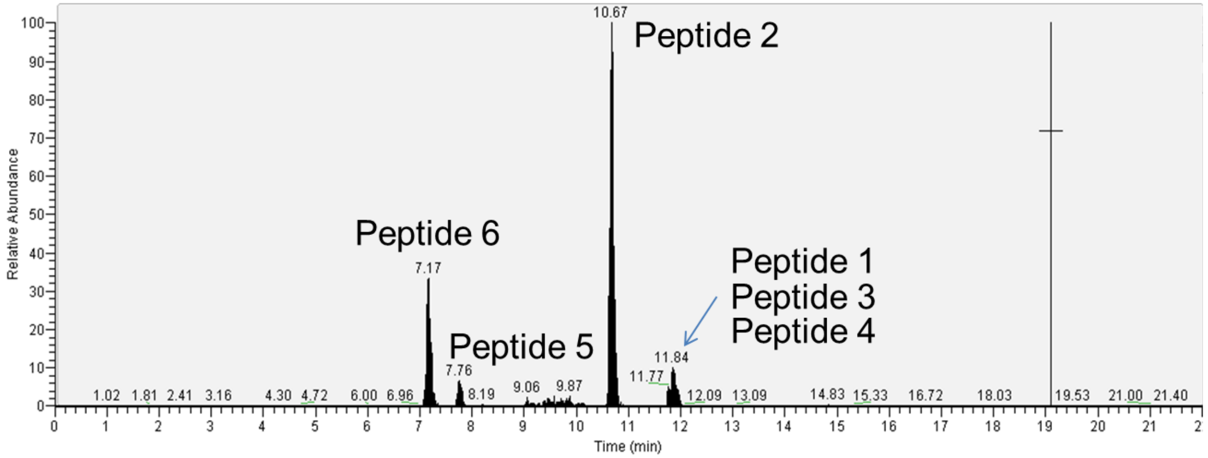
Filter peptides by: All modifications

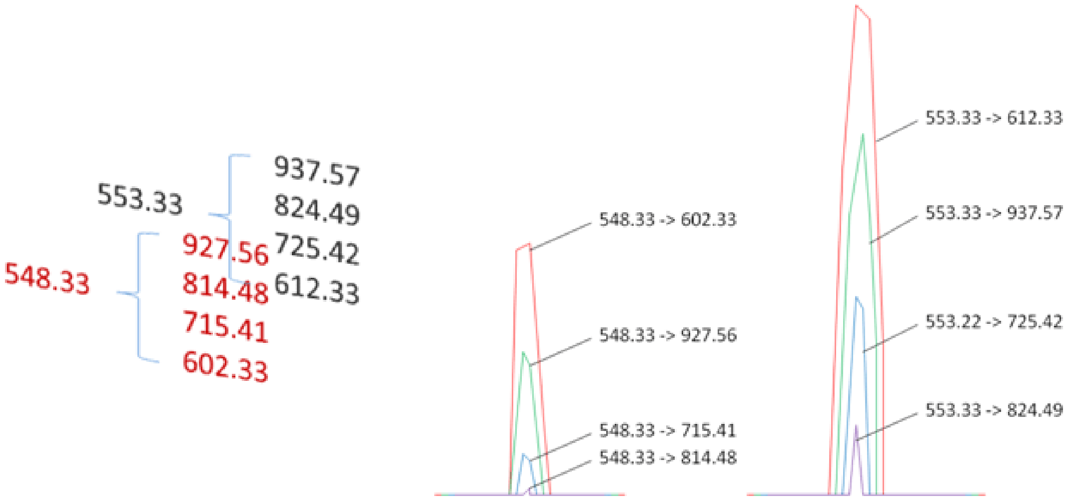
Enter GI numbers (separated by comma):
For example: 220335, 671882

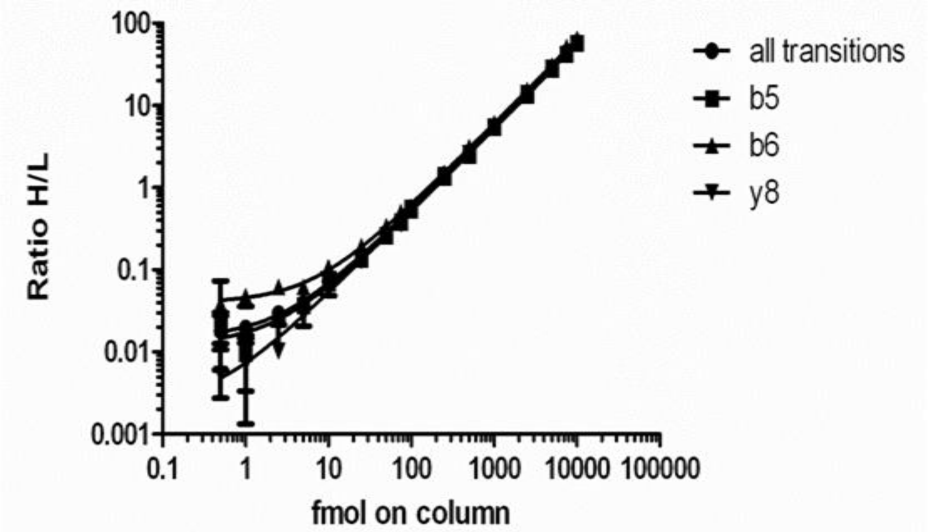
Enter IPI numbers (separated by comma):
For example: IPI00000001

Targeted method development (2)

- Study with LC MS/MS (in buffer)
- Study with LC MS/MS (in sample matrix)
- Assay characterization (in sample matrix)



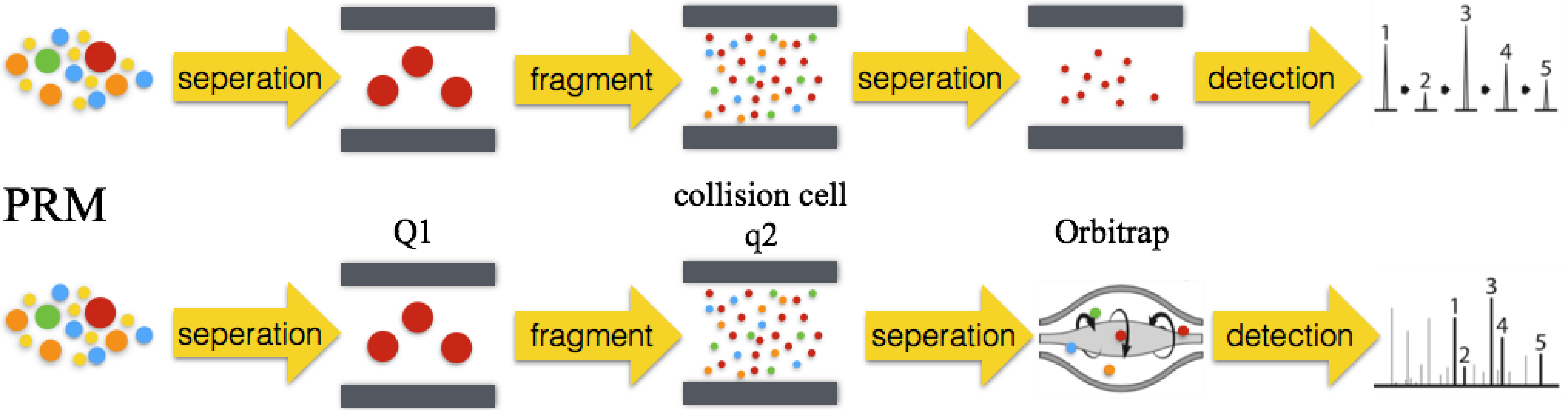




https://moritz.systemsbiology.org/resources/proteomics-program/targeted-proteomics/

SRM Assay development

Parallel Reaction Monitoring (PRM)



SRM/MRM

Q1

collision cell
q2

Q3

seperation

fragment

seperation

detection

PRM

Q1

collision cell
q2

Orbitrap

seperation

fragment

seperation

detection

Precursor
ion

Fragment
ions

https://www.creative-proteomics.com

Parallel Reaction Monitoring (PRM)

Peptide selection – Fragmentation – Fragment analysis

Proteins and peptides selection

Transitions selection

Collision Energy optimisation

Sample preparation optimisation

LC-MS method development

Method development

Selectivity

Sensitivity

Higher Specificity

Accuracy

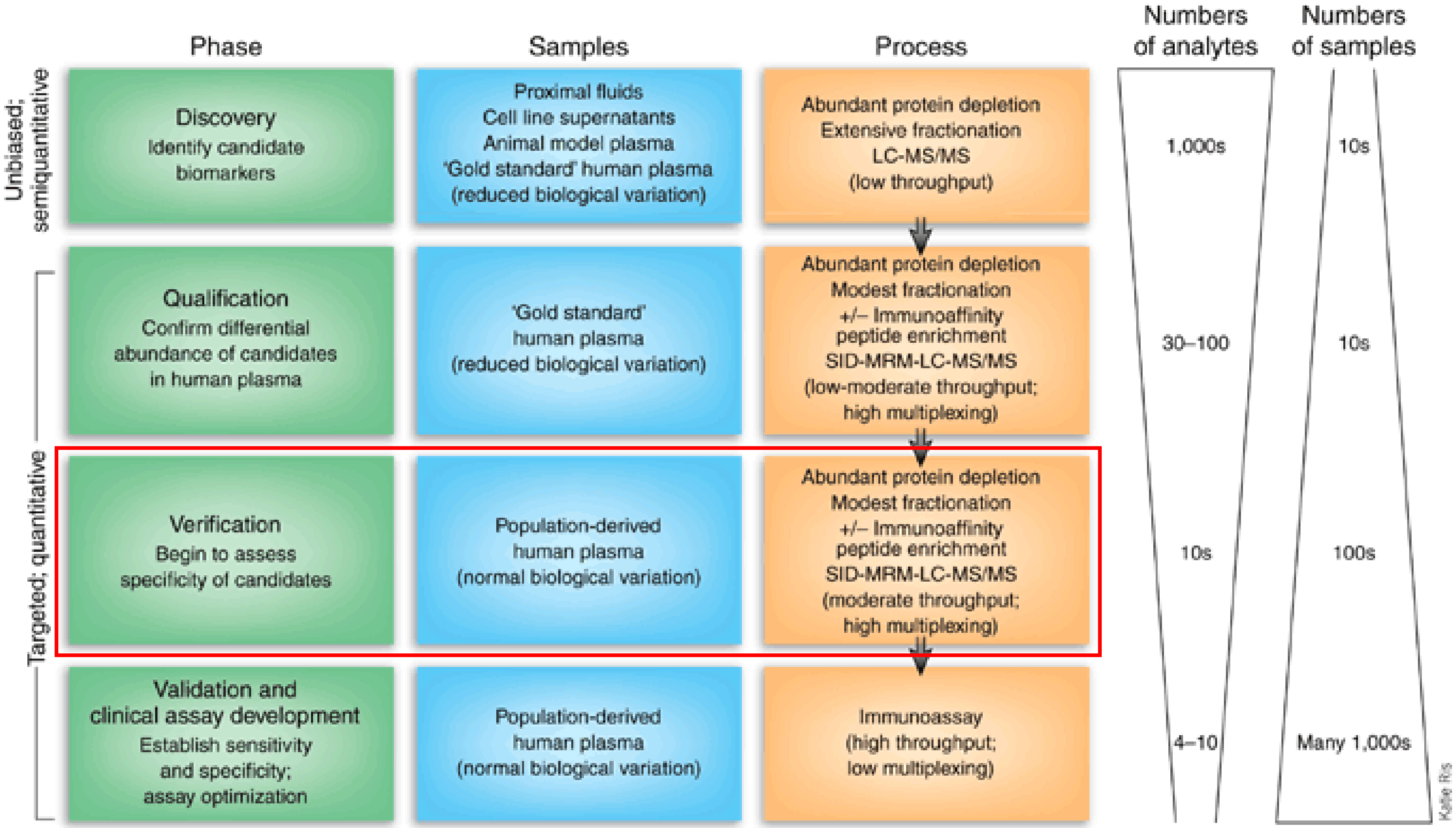
No missing values

Method characteristic

Method development time divided by 3
Simplified workflow
More confident in quantification results

Courtesy of Charlotte Macron

Main applications: Verification of biomarker discoveries

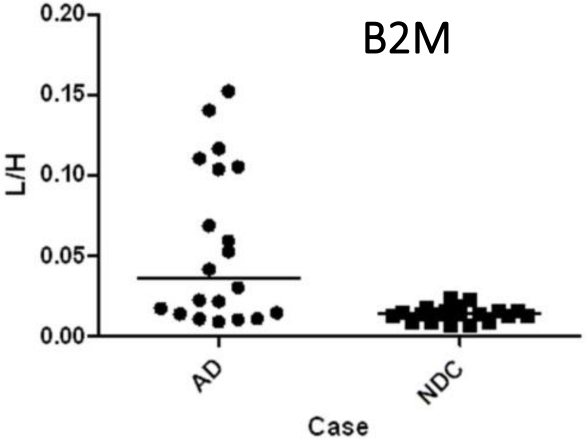


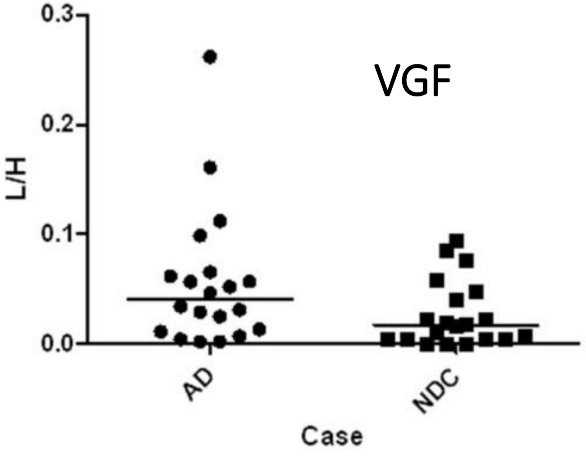
	Phase	Samples	Process	Numbers of analytes	Numbers of samples
Unbiased; semiquantitative	**Discovery** Identify candidate biomarkers	Proximal fluids Cell line supernatants Animal model plasma 'Gold standard' human plasma (reduced biological variation)	Abundant protein depletion Extensive fractionation LC-MS/MS (low throughput)	1,000s	10s
Targeted; quantitative	**Qualification** Confirm differential abundance of candidates in human plasma	'Gold standard' human plasma (reduced biological variation)	Abundant protein depletion Modest fractionation +/– Immunoaffinity peptide enrichment SID-MRM-LC-MS/MS (low-moderate throughput; high multiplexing)	30–100	10s
	Verification Begin to assess specificity of candidates	Population-derived human plasma (normal biological variation)	Abundant protein depletion Modest fractionation +/– Immunoaffinity peptide enrichment SID-MRM-LC-MS/MS (moderate throughput; high multiplexing)	10s	100s
	Validation and clinical assay development Establish sensitivity and specificity; assay optimization	Population-derived human plasma (normal biological variation)	Immunoassay (high throughput; low multiplexing)	4–10	Many 1,000s

DOI: 10.1038/nbt1235

Verification of biomarker discoveries

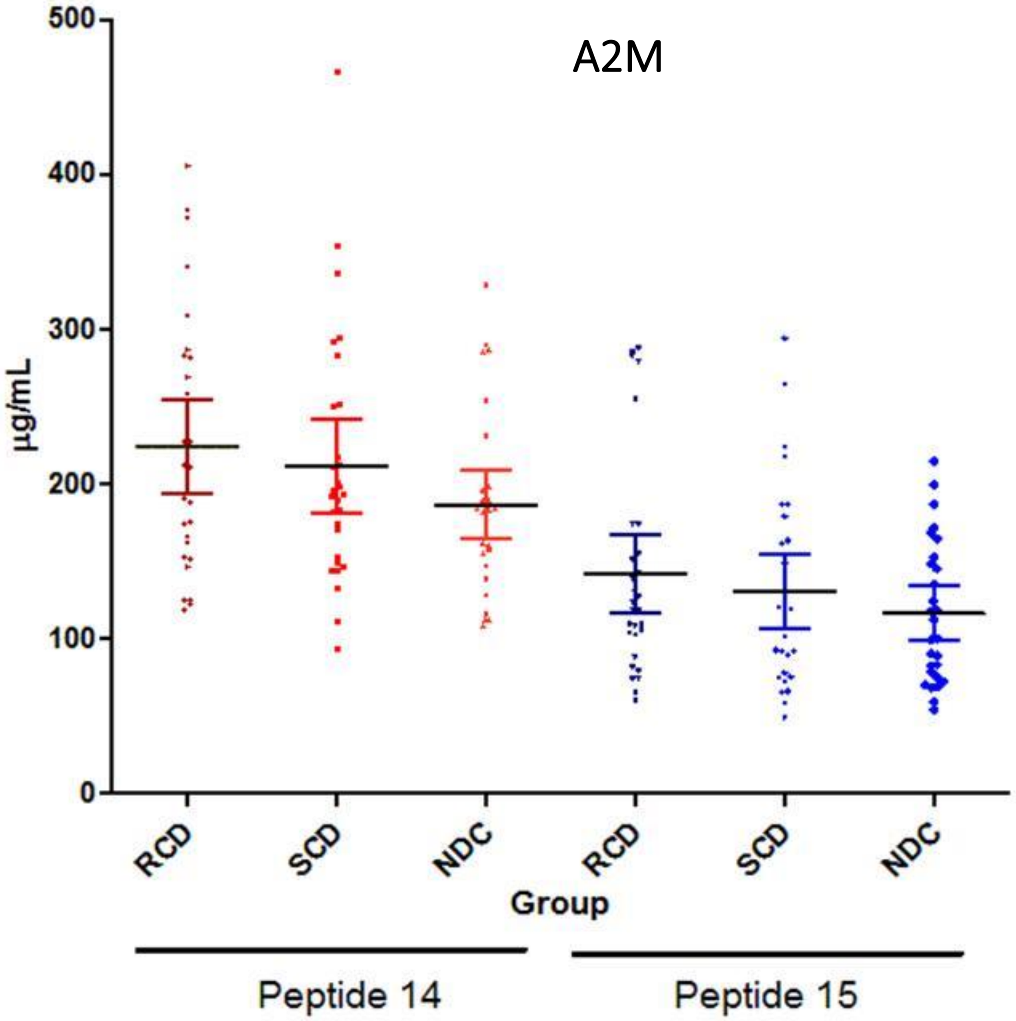
CSF





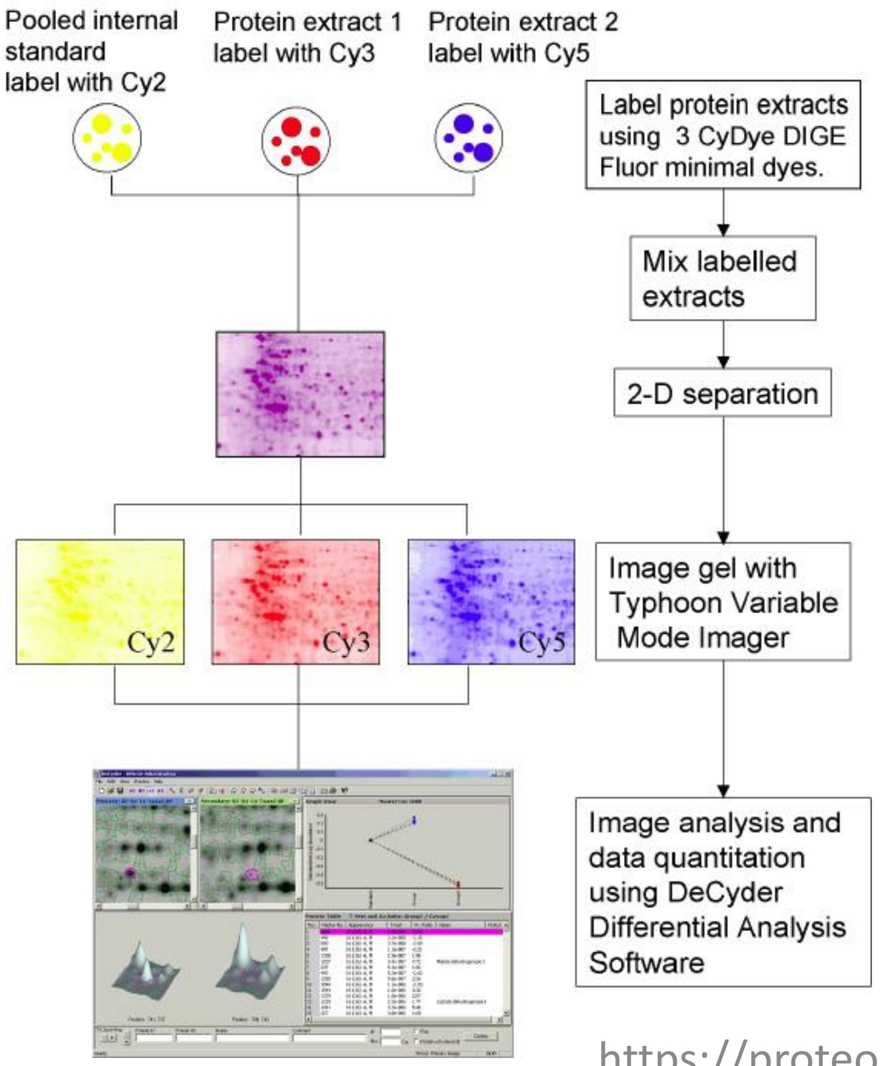
Alzheimer's and Dementia, Volume 7, Issue 4, Supplement, July 2011, Pages S150-S151

Plasma



Alzheimer's and Dementia, Volume 7, Issue 4, Supplement, July 2011, Page S333

2-dimensional differential gel electrophoresis (2D-DIGE))



Label protein extracts using 3 CyDye DIGE Fluor minimal dyes.

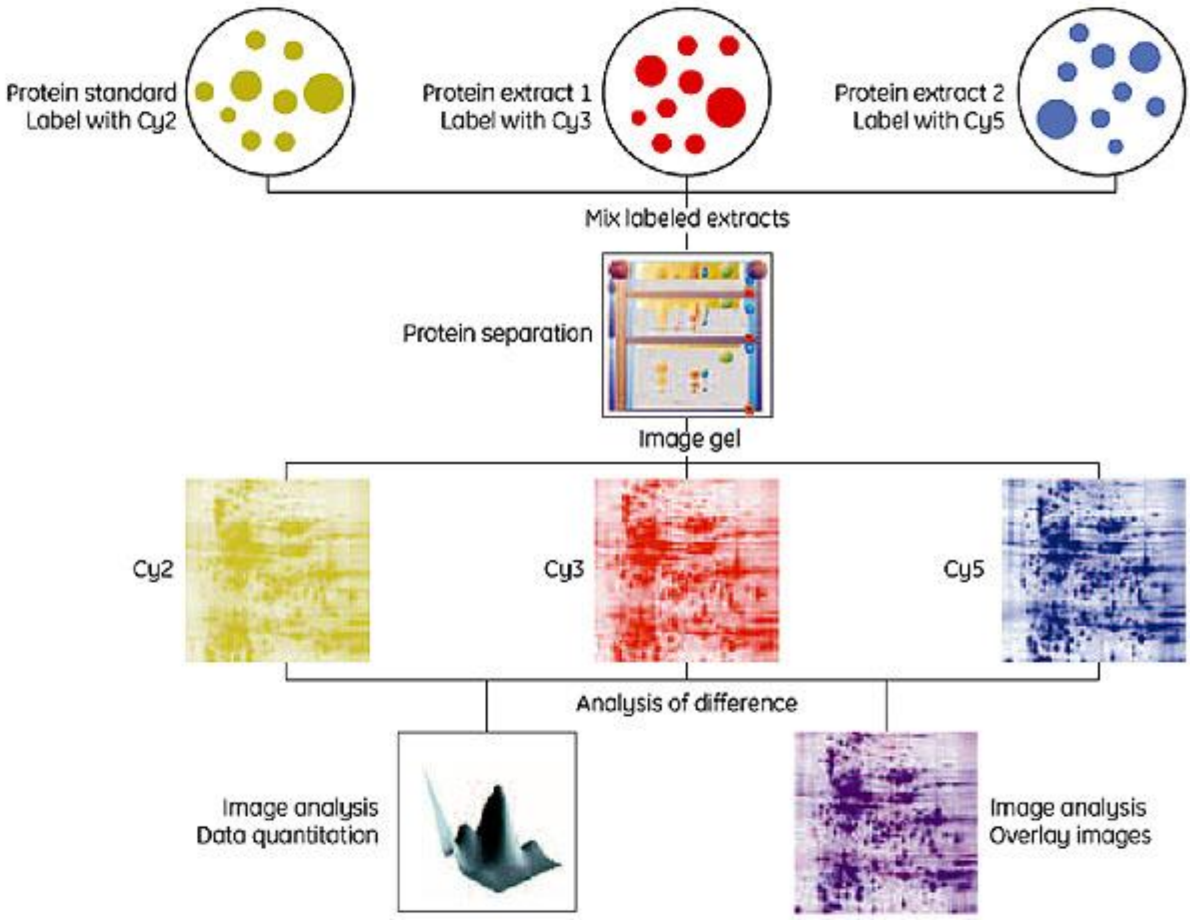
Mix labelled extracts

2-D separation

Image gel with Typhoon Variable Mode Imager

Image analysis and data quantitation using DeCyder Differential Analysis Software

https://proteomics.cf.gu.se



http://pawatbioinformatics.exteen.com/

Summary

- In proteomics, quantification is key to compare different biological conditions

- Different quantitative techniques are used to obtain relative quantification or precise “absolute” quantification of peptides and proteins

- MS is an alternative to standard biochemical methods (*e.g.*, immuno-assays) that allows easily multiplexing of samples and/or analytes

- In the next chapters, we will see the bioinformatic tools used to process the MS data and some concrete applications of those MS-based workflows to study proteomes in biology and clinical research