Different factors affect the efficiency of the protein expression, and on some of them we don't have any control.

We can define the species-specific codons based on the type of cell we are using. Once obtained our plasmid, it's better to perform a retro-translation in order to find the best codons.

For the back translation, you can do it:

- Starting from the primary sequence
- Starting from the DNA sequence in which we perform the translation

You can play with GC content to keep it in average.

Is the possibility to avoid the restriction site

Play with Codon usage, GC content because we know about the redundancy in codons. All codons share in e.coli and yeast and others but the codon proportion of a codon for one amino acid are not equal in them.

It is recommended to use the codon based on your recombinant.

We need to define the species-specific codon, according to the type of cell we are using.

If the restriction sites are inside the protein coding sequence we should remove them because they would truncate the sequence.

QUANTITATIVE PROTEOMICS

Why do we need quantitative proteomic analysis:

The expression of certain proteins changes upon cell stimulation or stress, or when cells differentiate or turn into a disease state. We need to achieve quantitative identification. Measuring the protein expression levels as well as characterizing their post-translational modifications gives us information on their cellular state.

The goal of quantitative proteomics is to evaluate the relative expression of proteins between two cellular samples being compared.

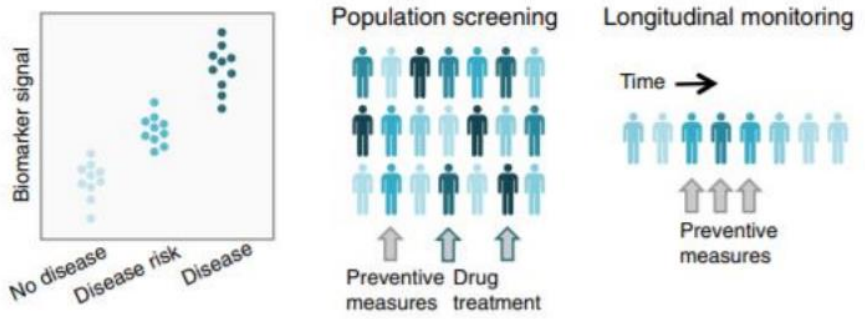
• comparison between tissue types.

• biomarker discovery – for comparing healthy and disease states.

• response to drug or pathogen treatment, in order to try to use personalized medicine.

• study stress responses.

QUANTITATIVE PROTEOMICS IN MOLECULAR MEDICINE:



Proteomics play a huge role in personalized medicine, there are two strategies:

a) **Population screening**: we try to identify those subjects who are at risk or are in the first stages.
b) **Longitudinal monitoring**: we take a subject, and we follow him/her for a long period. It tries to consider the high variability intra- and inter subjective.

Applications of quantitative proteomics in personalized medicine are typically based on biomarkers that reflect disease risk or disease status. Biomarkers are screened across individuals or patient cohorts.

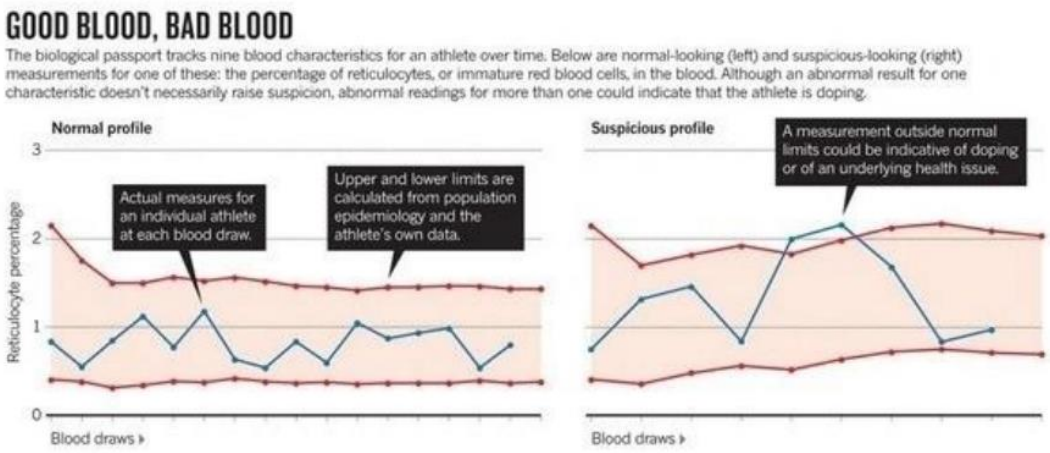
Longitudinal profiling of individuals allows the molecular profile of a person to be monitored over long-time frames, and comparison of each measurement with data for the same subject from previous time points yields more meaningful clinical information than comparison of a single measurement with the population average.

This personalized approach to molecular medicine is expected to enable early and highly sensitive detection of disease risk and is therefore the most effective in **disease prevention.**

LONGITUDINAL PROFILING: ATHLETE BIOLOGICAL PASSPORT

Biological passport: includes most relevant parameters for each athlete.

The fundamental principle of the Athlete Biological Passport (ABP) is to monitor selected biological variables over time that indirectly reveal the effects of doping rather than attempting to detect the doping substance or method itself.



Fluctuations are restricted in the range of physiological values.

QUANTITATIVE PROTEOMICS

Quantitative proteomics identifies proteins in order of their relative abundance--> untargeted analysis, more powerful but we are dealing with a lot of proteins, and we can normally detect those present in a higher amount.

Additive techniques, such as SILAC and isobaric labeling methods, are incorporated to obtain quantitative information. Quantitative proteomics can be used to compare the expression levels of 2000–6000 proteins in cells or tissues and 200–500 proteins in plasma.

TARGETED PROTEOMICS

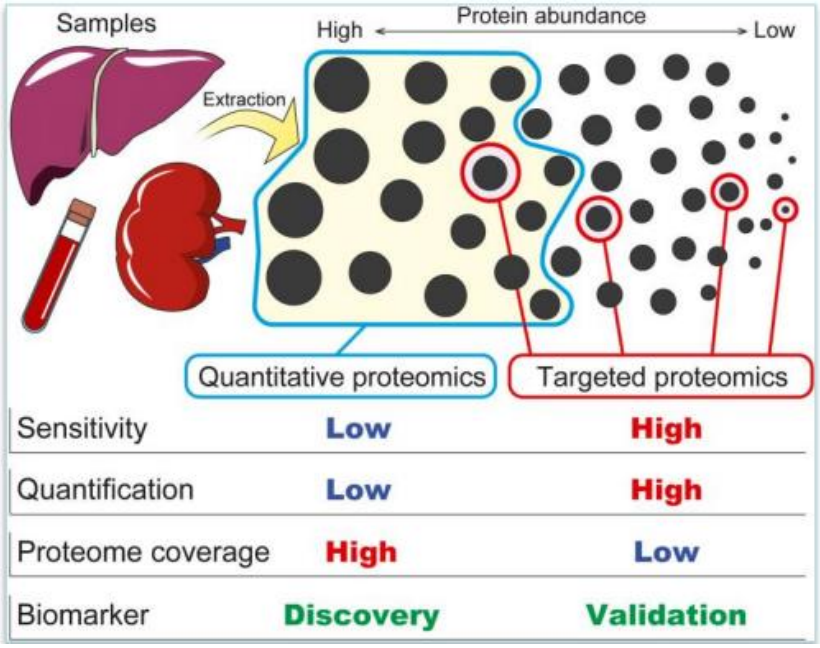
Targeted proteomics only measures proteins selected before the LC-MS measurement--> quantify only target proteins connected to biological events that are investigated.

Targeted proteomics is a method used to quantify only target proteins. Proteins are measured by the quantification of a specific digested peptide(s) using chromatographic information, and the same methodology was used for quantifying small molecules by LC-MS. Targeted proteomics can be used to measure 10–100 target proteins simultaneously.

Biomarker discovery pipeline

- **Quantitative** proteomics provides more comprehensive protein information, including protein expression profiles, and higher protein coverage than targeted proteomics. Less sensitive but we can identify more proteins.
- In contrast, **targeted** proteomics provides more sensitive and accurate quantitative protein information than quantitative proteomics.

Considering these characteristics, biomarkers are discovered using quantitative proteomics to identify marker candidate proteins. Then, the identified marker candidate proteins in large sets of samples are accurately quantified using targeted proteomics to validate biomarker performance.



Workflow proteomics (bottom-up)

- Sample
- Cell lysis (unless plasma proteomics)
- Protein preparation
- Digestion (trypsin)
- Separate peptides
- MS and or...

We can have a first labelling when the cells are still alive (metabolic labelling: SILAC and SILAM), or later after obtained the peptides (in vitro labeling). Or we can skip labelling and go for a label free workflow.

Strategies available for mass spectrometer-based proteomic quantification

SILAC – **Stable Isotope Labelling with Amino Acids in Cell Culture** – cells grown in the presence of light (control) and heavy (treated) amino acids – measuring ratios at the peptide level via MS data. --> label cells when they are going in the culture, with heavy or light gas

ICAT - **Isotope Coded Affinity Tags** - two reagents – heavy and light – labelling done at the protein level – cysteine containing peptides – affinity selection – quantification via MS data. --> the tag is quite peculiar, we can have different parts of the molecule used for the quantification, usually binds to cysteine residues.

iTRAQ – **Isobaric Tagging Reagents for Relative and Absolute Quantification** - four (eight) reagents – multiplexing - labelling done at the peptide level – all peptides are labelled – quantification done via MS-MS data. – multi-dimensional LC-MS approach--> most powerful approach, we can also do a multiplexing (we can analyze up to 8 samples at one)

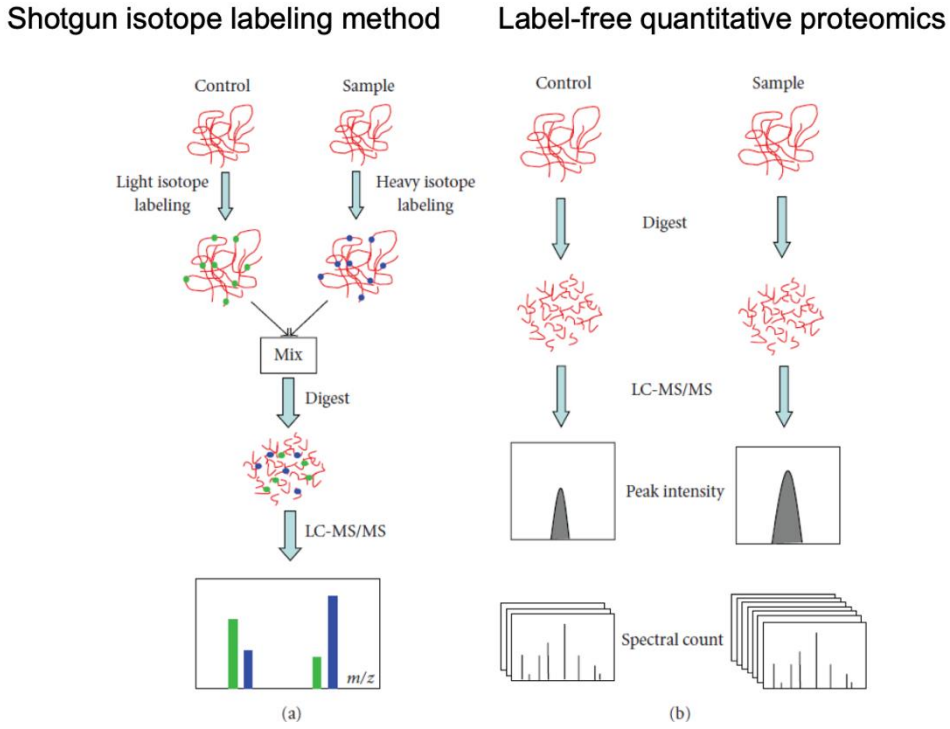
$^{18}O/ ^{16}O$ enzymatic labeling

Label-free

Questions are very general in the oral

Example:

- Which is the difference between label based and label free workflow?
- Difference between ICAT and iTRAQ



Difference between labeling and label free workflow:

Experiments with labeling enable us to achieve the quantitative information, we can mix samples and then understand which is which from the difference masses or fragmentation pathways.

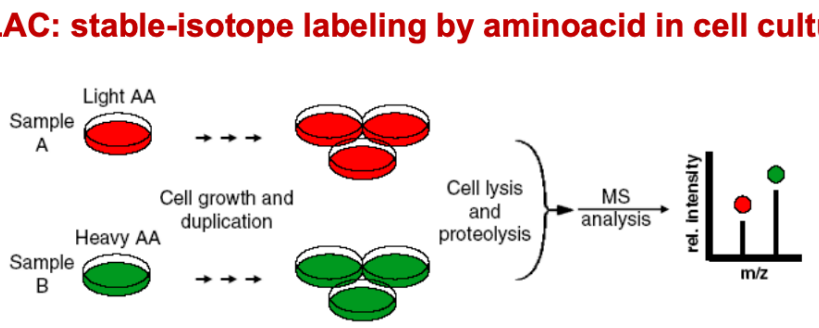
With label-free we keep everything separated and we'll try to achieve the quantification either with peaks intensity or spectral count.

SILAC: stable isotope labeling by amino acid in cell culture

• SILAC is a technique commonly used to perform quantitative proteomic studies. The approach is based on the use of stable isotopes, like 13C and 15N, which are characterized by not being radioactive, having the same chemical and physical properties of the normal 12C and 14N atoms, but with a different mass.

• **The stable isotopes are used to synthesize arginine and lysine amino acids**, which are then termed "heavy" because of their higher mass compared to the normal ones. The choice of arginine and lysine is preferred in cell culture since they are essential (not synthesized at all by the cell) or semi-essential (can be generated by the conversion of proline) amino acids, respectively. Moreover, they are well suited for the standard digestion of proteins into peptides with trypsin, which cuts C-terminally to Arg and Lys or endoprotease Lys-C, which cuts C-terminally to Lys.

In-vivo labeling (metabolic)

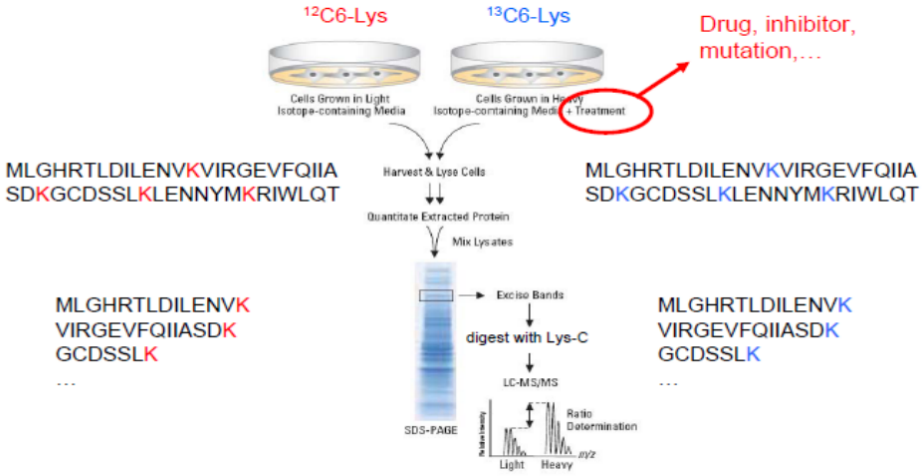


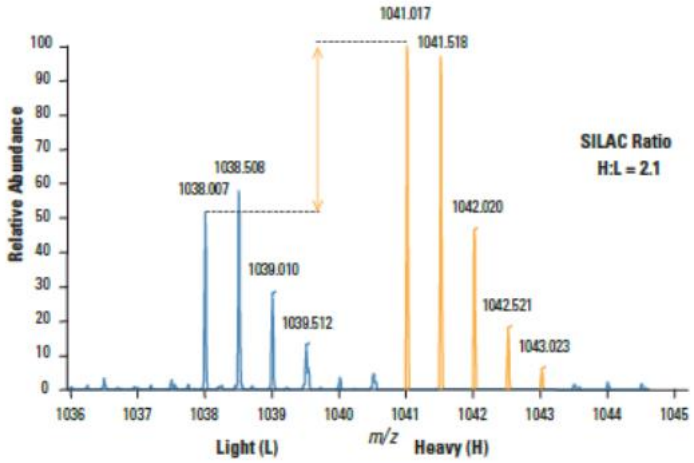
In this quantification procedure, labeled, essential amino acids (**deuterated leucine, Leu-d3**) are added to amino acid deficient cell culture media and are thus incorporated into all proteins as they are synthesized. No chemical labeling or affinity purification steps are performed. In a typical experiment, a cell population is treated in a specific way, such as cytokine stimulation. Protein populations from both this experimental sample and the control are then harvested, and because the label is encoded directly into the amino acid sequence of every protein, the extracts can be mixed directly. Purified proteins or peptides will preserve the exact ratio of the labeled to unlabeled protein, as no more synthesis is taking place.

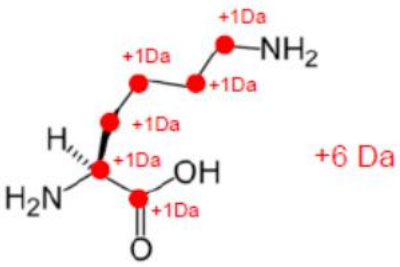
In most of the non-invasive samples of the personalized medicine, we are not able to grow up cells because the samples come from blood, urine, tears… where there aren't cells to cultivate. So, we are not able to perform SILAC.

SILAC approach

In SILAC, differentially labeled samples are mixed early in the experimental process, and analyzed together by LC-MS/MS. Since the labeling does not affect the chemical properties of the molecules, they co-elute from the LC column and analyzed together in the mass spectrometer. The peptide peaks of the differentially labeled samples can be very accurately quantified relative to each other to determine the peptide and protein ratios.







This is a relative quantification: we have a delta of 6Da. The same peak will be shifted of 3, by calculating the ratio of the height of the peak we can have the relative quantification.

Representative MS spectra generated using SILAC. Light and heavy ($^{13}C6$) L-lysine containing peptides (AEDNADTLALVFEAPNQEK) were analyzed by MS. Mass spectra of heavy peptides containing 13C6 L-lysine have an increased mass of 6Da and are shifted to the right of light peptide spectra by a mass to charge ratio (m/z) of 3 caused by a +2 ionization of peptides. Ratios are determined by comparing the heights of the peaks from light and heavy peptides.

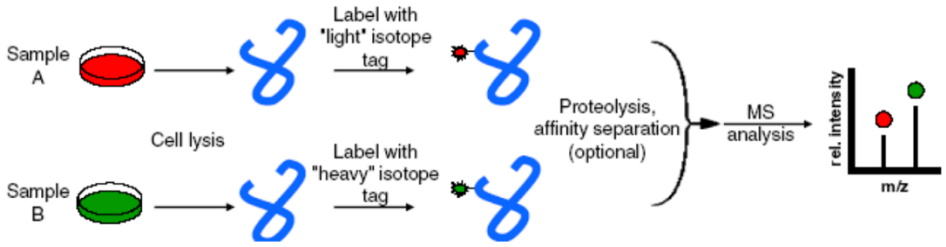
SILAC labeling of whole organisms: the SILAC mouse

Application of SILAC to whole organisms is valuable when in-vivo experiments are performed that analyze whole tissues rather than cultured cells. SILAC labeling of whole organisms requires preparation of **food that contains the SILAC amino acid** as the sole source and that is compatible with growth of the organism.

Full incorporation requires the turnover of the organism's entire proteome, which can be achieved by feeding them for more than one generation with food that contains heavy amino acids in place of normal ones. These then represent the real complexity of the organism and the specific organs and tissues of interest, including extracellular proteins and body fluids. Since arginine is not an essential amino acid in some tissues, most model organisms are labeled only with **lysine**.

We have most of the efforts directed to in vitro labeling in which we can start with any type of samples:

In-vitro labeling (chemical)

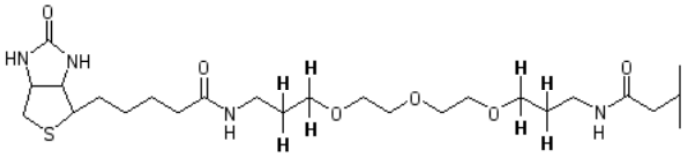


- Labeling with a heavy or light tag
- Digestion
- Separation
- Identification and quantification

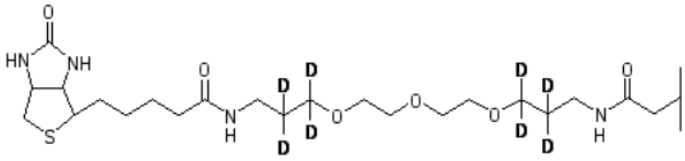
ICAT: Isotope-Coded Affinity Tag

A method for systematic identification and quantitation of proteins by using **isotope coded affinity tag** reagents. This approach uses two reagents, a light version and a heavy one, which have a mass difference of 8 (same with the SILAC), the only difference is the linker (in the light version there is H, in the heavy D).

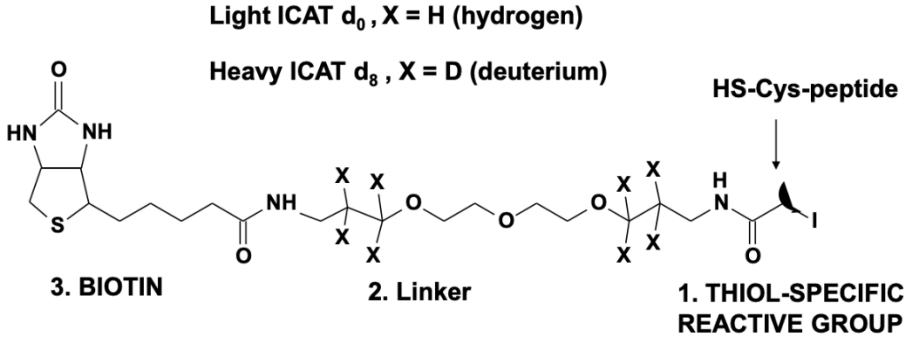
light version with hydrogens



Heavy version with deuteriums



STRUCTURE OF THE ICAT REAGENT



ICAT reagents consist of <u>3 functional components</u>:

1. a **reactive thiol-specific group** capable of **labeling** a defined amino acid side chain (e.g., iodacetamide to modify cysteine residues)

2. An **ethylene glycol linker** (the light and heavy part) that occurs in deuterated and isotopically normal form and provides the basis for quantification --> the only difference between light and heavy ICAT reagent

3. **Biotin** (Every time you'll find biotin in an experiment, it means that you'll use it for retrieving your molecule, your target peptide. Biotin provides an affinity tag for the selective isolation of tagged peptides)

ICAT reagent will bind to cysteine residues.

For the quantitative comparison of two proteomes, one sample is labeled with the isotopically light (d0) probe and the other with the isotopically heavy (d8) version.

To minimize error rate, both samples are then combined, digested with a protease (trypsin), and subjected to avidin affinity chromatography to isolate peptides labeled with isotope-coded tagging reagents. These peptides are then analyzed by liquid chromatography-mass spectrometry (LC-MS).

The ratios of signal intensities of differentially mass-tagged peptide pairs are quantified to determine the relative levels of proteins in the two samples.

The principle is the same for the SILAC, but it's more general, we can label every sample.

Limitation of this approach:

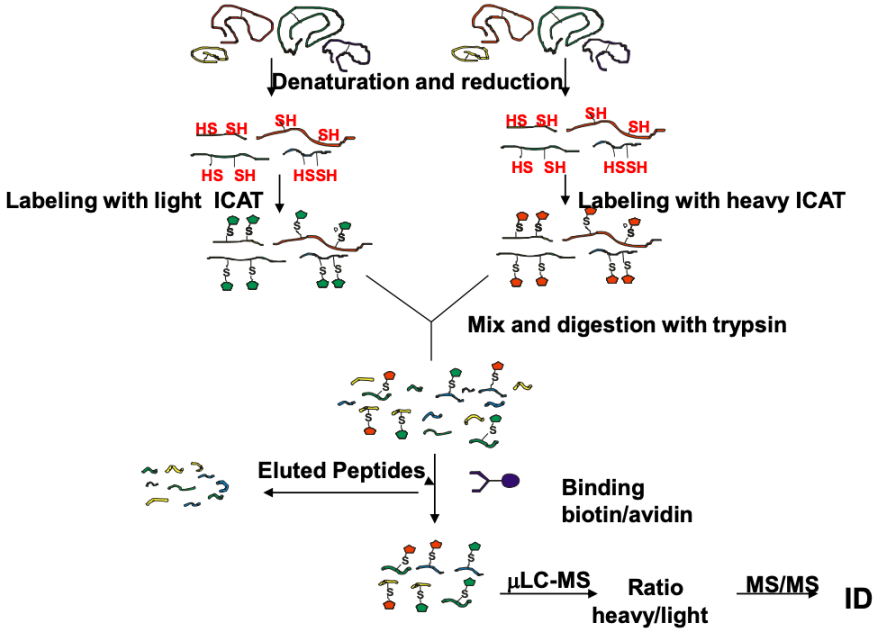
Cysteine is not that much frequent in peptides

METHODOLOGY

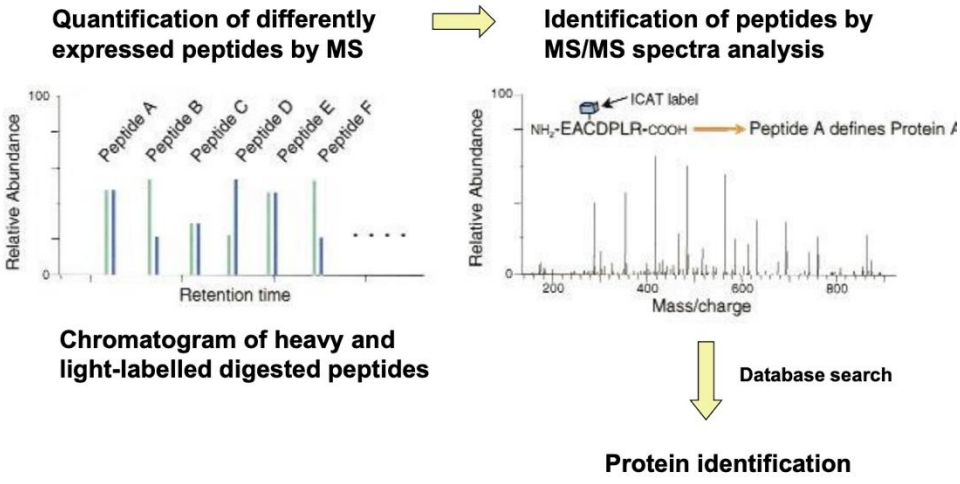
- We denature and reduce our proteins in order to have free cysteine
- We do the labeling
- Mix and digest with trypsin

Not all the peptides are labeled (not all contain cysteine)

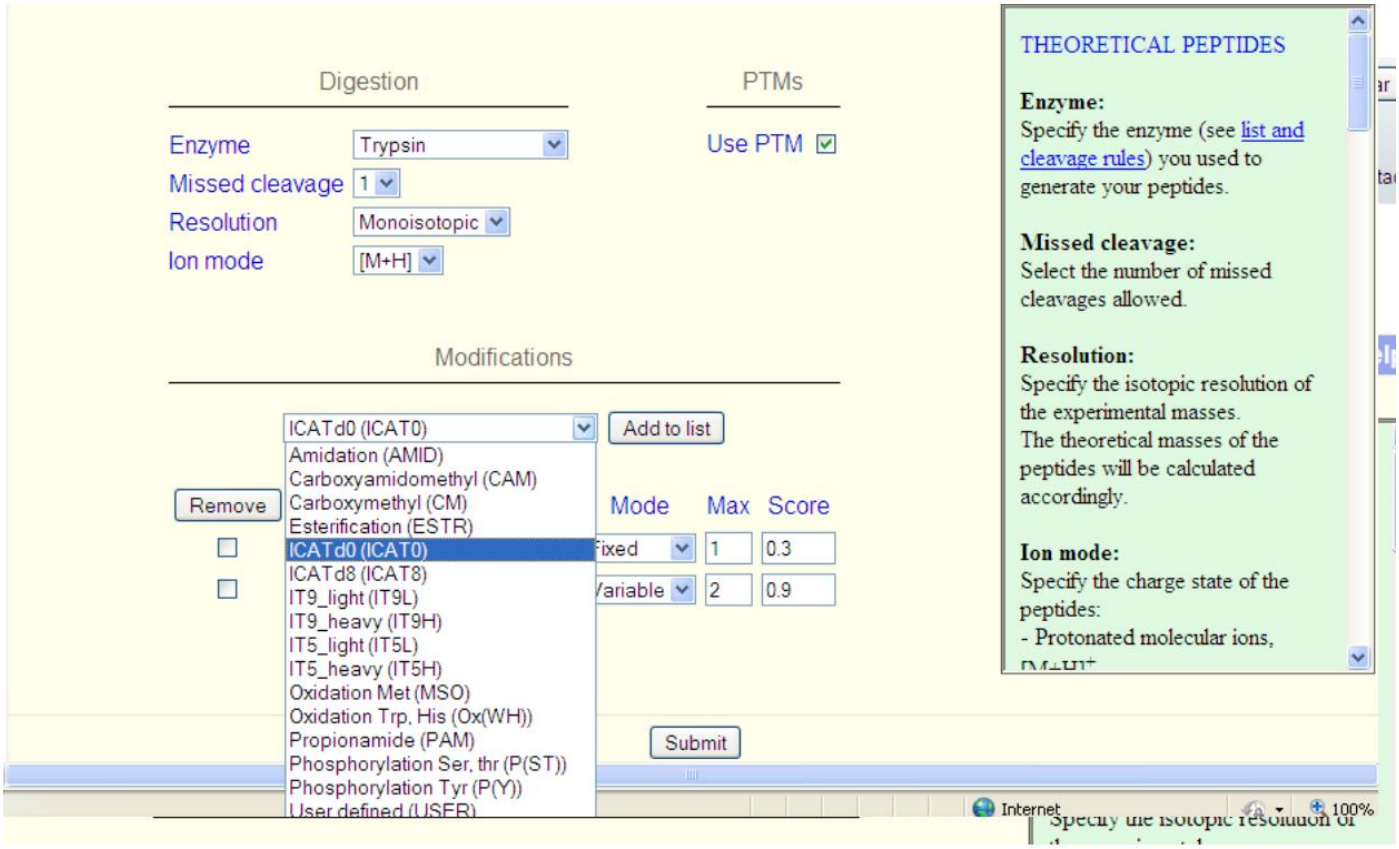
- We are going to rescue the cysteine-containing peptides with some general avidin beads, which bind biotin.
- We analyze only the labeled peptides and calculate the ratio heavy/light



MS spectra interpretation: With MALDI-TOF we are not able to achieve chromatogram of heavy and light-labelled digested peptides.



How to use MASCOT for ICAT experiments?



Modifications can be naturally present in the protein (in several diseases we have a different path of modification) or introduced by the workflow.

It's less performing than the iTRAQ but still one of the most used.

Advantages

• Estimates relative protein levels between samples with a reasonable level of accuracy (within 10%)

• Can be used on complex mixtures of proteins

• Cys-specific label reduces sample complexity

• Peptides can be sequenced directly if tandem MS-MS is used

Disadvantages

• Yield and non-specificity

• Expensive

• Meaning of relative quantification information

• No presence of cysteine residues or not accessible by ICAT reagent

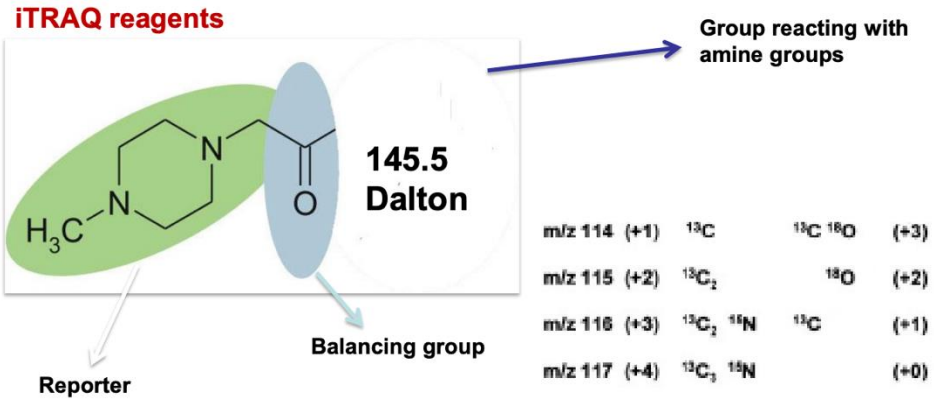
SILAC vs ICAT

Advantages of SILAC over ICAT include the fact that almost 70% of unique tryptic peptides in the human genome contain at least one leucine, while only ~25% contain cysteine, the common target for chemical tagging.

A disadvantage of SILAC is that it is limited to cells that can be grown in culture--> but with SILAC you have a very high efficiency of the labeling because it is performed with essential aa.

iTRAQ (isobaric Tagging for Relative and Absolute Quantitation)

Most powerful approach.

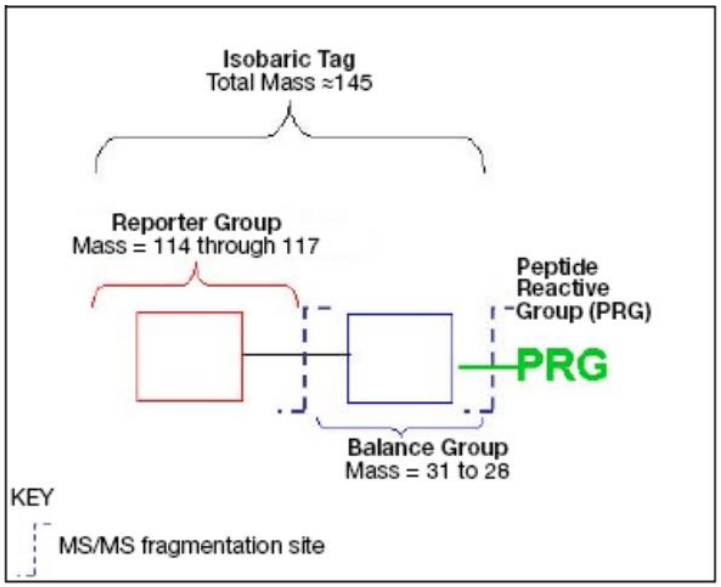


4 different iTRAQ reagents, 4 different reporters, same total mass (=> “isobaric”), same behavior before fragmentation, different behaviors after fragmentation.

The components are:

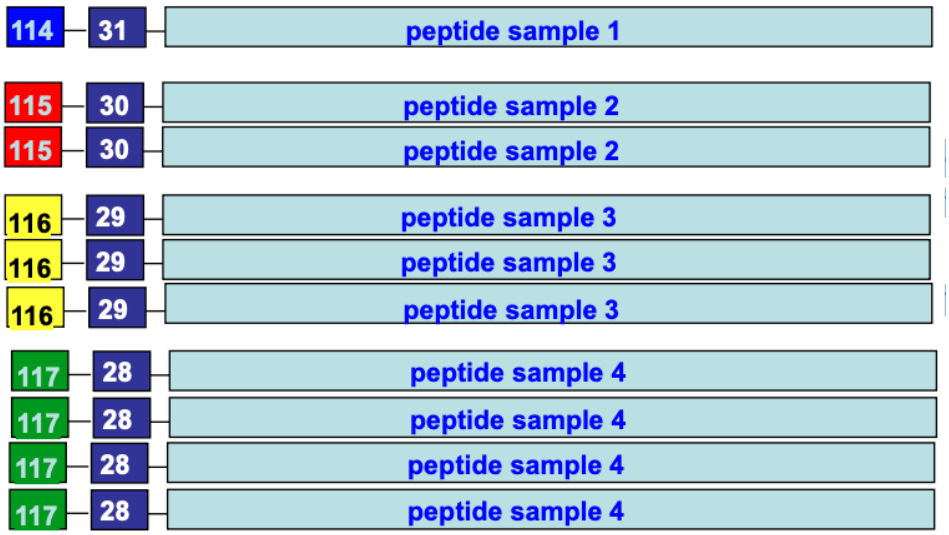
- Group reacting with amine groups
- Reporter group
- Balancing group

The sum of reports and balancing is always the same = isobaric tag (145 is the total mass)

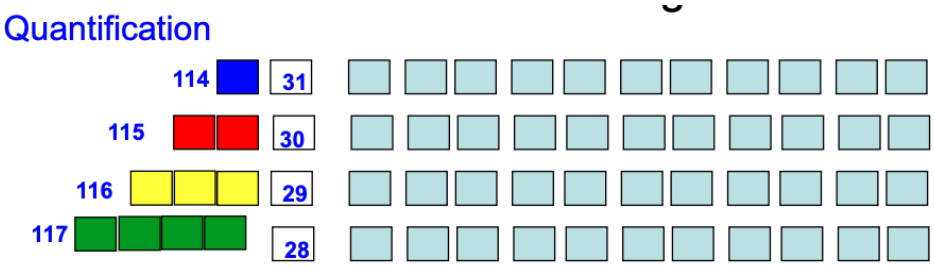


ITRAQ

4 different samples in which the same peptide is labelled with different iTRAQ labels. We can see that we have different amounts of the same protein in the different samples.



Collision Induced Fragmentation --> Peptide Identification (with normal techniques as PMF). In a tandem MS, you separate all the ions and then recover the info.



Sequence ions combined from all four peptides Eg. Peptide in sample 4 is four times more abundant than in sample 1.

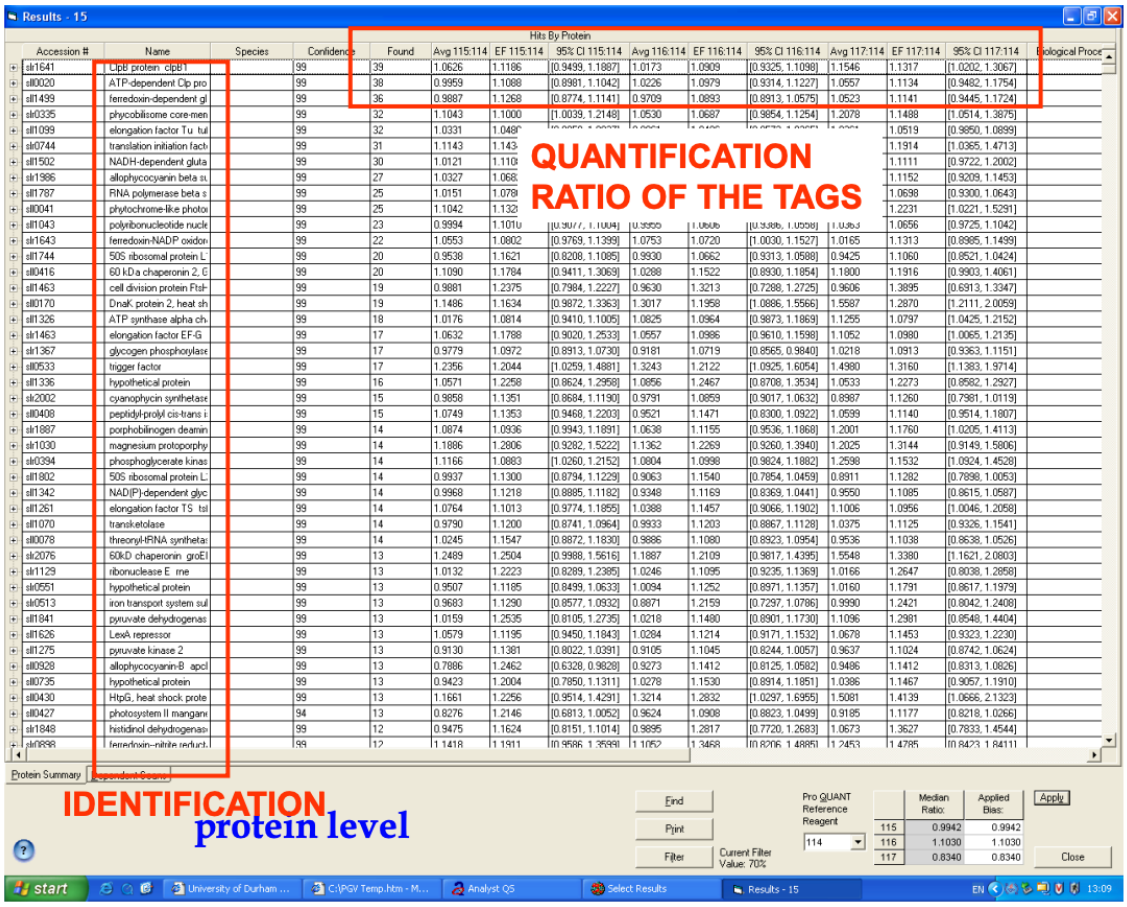
Data identification and quantification using ProQuant software.

For each peptide separated by LC, fragmentation is performed and the 114 – 117 region of the MS-MS fragmentation spectra is analyzed in order to quantify the 4 tags.

It uses a novel sequence matching algorithm in a database search to match the peptide amino acid sequence and produce a protein identification.

Tabulates all the peptide identification data with the quantification data and produces an output with a statistical measure of both the identification and quantification data.

ProQuant data output from a single fraction from 1 experiment (image below). The data consists of an identification for different proteins and the quantification ratio of the tags. From the data we can extrapolate both absolute and relative quantification.



Labeling-based quantification approaches: limitations

- Increased time and complexity of sample preparation
- requirement for higher sample concentration
- high cost of the reagents
- incomplete labeling
- requirement for specific quantification software

(iTRAQ allows the comparison of multiple samples at the same time, no more than 8)

Label-free approaches

Label-free approaches can be divided into two distinct groups according to the method used for data extraction

1. SPECTRAL COUNTING METHODS: counting the number of peptides or spectra assigned to a given protein (the more the peaks, the more abundant will be the protein)
2. PRECURSOR SIGNAL INTENSITY

The extraction of the area of the precursor ions' chromatographic peaks – area under the curve (AUC)- or MS signal intensity methods.

they all include the following fundamental steps:

(i) sample preparation including protein extraction, reduction, alkylation, and digestion;

(ii) sample separation by liquid chromatography (LC or LC/LC) and analysis by MS/MS

(iii) data analysis including peptide/protein identification, quantification, and statistical analysis

Each sample is separately prepared, then subjected to individual LC-MS/MS or LC/LC-MS/MS runs

Protein quantification is generally based on two categories of measurements.

1. measurements of ion intensity changes such as peptide peak areas or peak heights in chromatography.

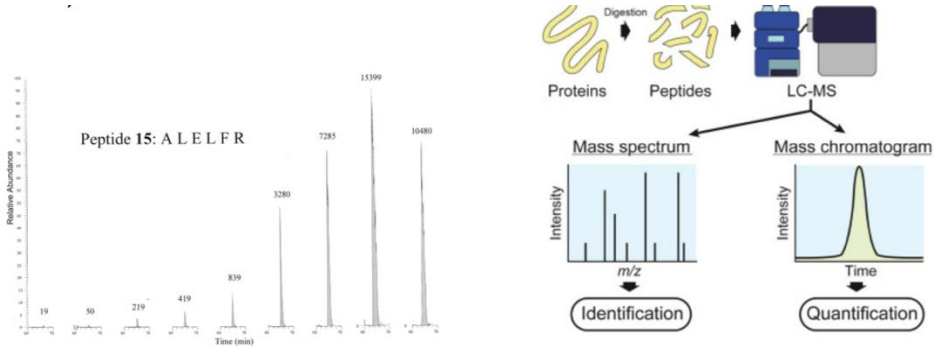
2. spectral counting of identified proteins after MS/MS analysis.

Peptide peak intensity or spectral count is measured for individual LC-MS/MS or LC/LC-MS/MS runs and changes in protein abundance are calculated via a direct comparison between different analyses.

Relative Quantification by Peak Intensity of LC-MS

Signal intensity from electrospray ionization (ESI) correlates with ion concentration.

The label free quantification of peptide/protein via peak intensity in LC-MS was first studied by loading 10 fmol–100 pmol of myoglobin digests to nano-LC and analyzing by LC/MS/MS. When the chromatographic peak areas of the identified peptides were extracted and calculated, the peak areas were found to increase with increased concentration of injected peptides. After the peak areas of all identified myoglobin peptides were combined and plotted against the protein amount, the **peak area was found to correlate linearly to the concentration of protein** (r2 = 0.991).



Changes of chromatographic peak area of peptide **15** (ALELFR) with different amounts of myoglobin injected. Horse myoglobin (1 mg; 60 nmol) was digested with trypsin, and 10, 100, 250, 500, 1000, 5000, 25000, 50000, and 100000 fmol samples were analyzed

Works good with standard samples, not with complex samples (biological samples) because of the interference.

PROBLEMS

• even the same sample can result in differences in the peak intensities of the peptides from run to run

• any experimental drift in retention time and m/z will significantly complicate the direct, accurate comparison of multiple LC-MS datasets

• the large volume of data collected during LC-MS/MS analysis of complex protein mixtures requires the data analysis of these spectra to be automated

Relative Quantification by Spectral Count

• relative protein quantification is achieved by comparing the number of identified MS/MS spectra from the same protein in each of the multiple LC-MS/MS or LC/LC-MS/MS datasets.

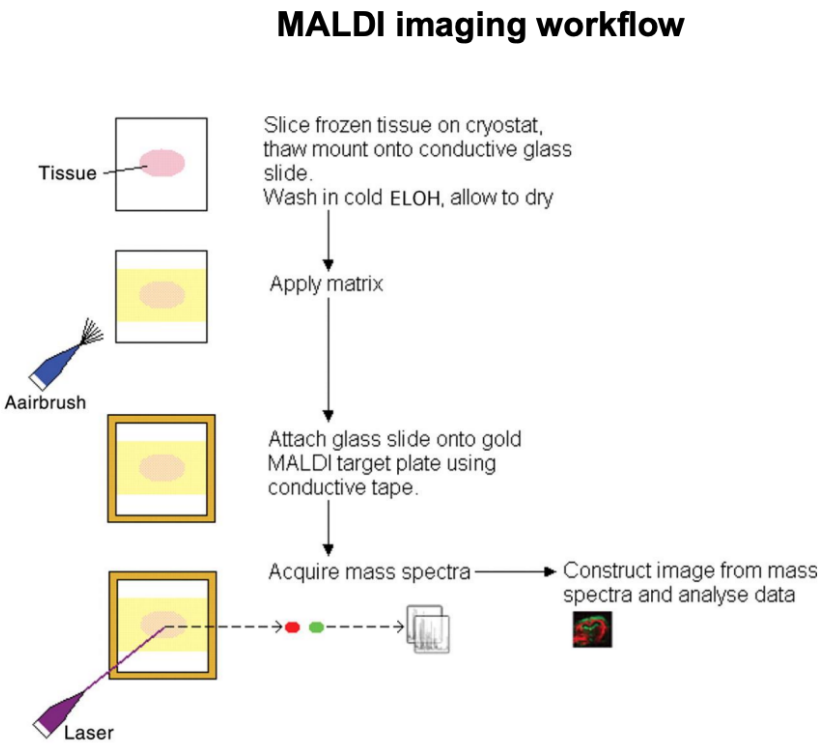
• This is possible because an increase in protein abundance typically results in an increase in the number of its proteolytic peptides, and vice versa. This increased number of (tryptic) digests then usually results in an increase in protein sequence coverage, the number of identified unique peptides, and the number of identified total MS/MS spectra (spectral count) for each protein.

• No specific tools or algorithms have been developed specially for spectral counting due to its ease of implementation.

MALDI-TOF MS: tissue imaging

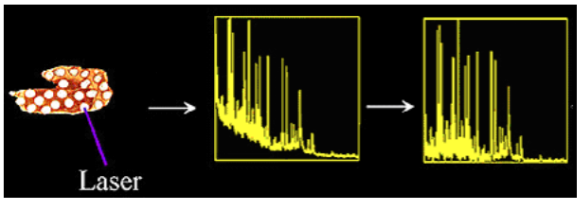
We can increase the relevance of information and achieve also spatial information, where are the proteins located (fundamental also in cancer research).

Here instead of having a sample hit with the laser we have a slice of tissue attached to the matrix. The rest of the workflow is the same.

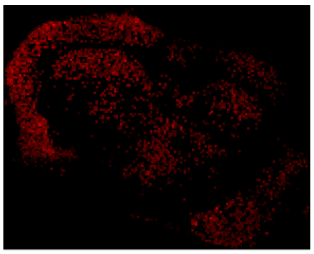


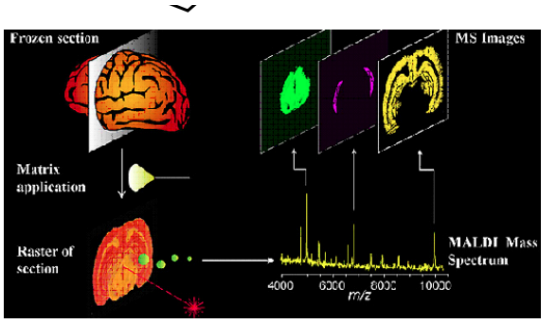
Mass Spectrom Rev. 2019 Oct 11. doi: 10.1002/mas.21602. Unsupervised machine learning for exploratory data analysis in imaging mass spectrometry. Verbeeck N, Caprioli RM, Van de Plas R

Profiling vs Imaging



Analysis of many samples and of many spots for every section of tissue, with the aim to compare classes of samples





Experiments performed on a single representative sample, with the aim to obtain 2D gel images of the distribution of proteins in the tissue

Complementary configuration with respect to profiling, in this case we use just one sample.

3D MALDI IMAGING

It is an expansion of MALDI IMAGING 2D which allows the visualization of the distribution of the proteins present in a tissue through a tridimensional representation. It can be applied both to known and unknown proteins.

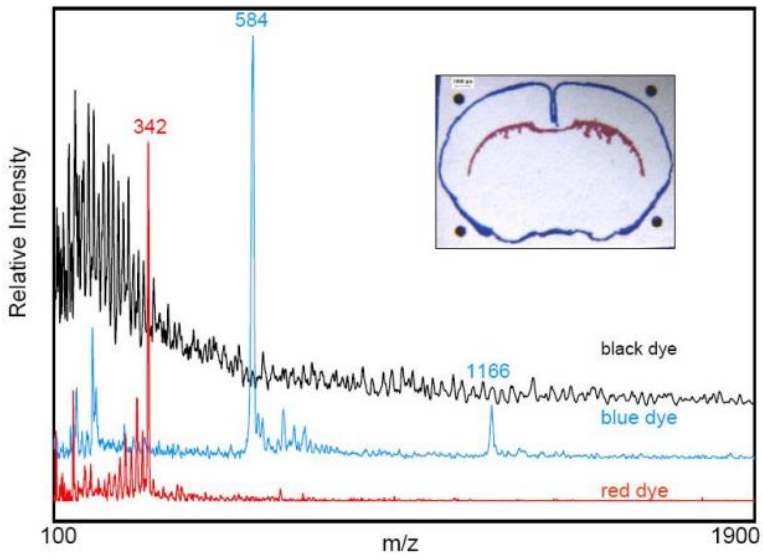
We can obtain a reconstruction of a tridimensional tissue by using the information coming from different tissue slices.

1. The material of interest (i.e.: brain) is cut into consecutive sections;

2. For each section the 2D-MALDI analysis is performed, and the m/z peaks acquired.

3. **All the consecutive 2D-images are combined in order to reconstruct the 3D image of the sample**

DEVELOPMENT OF THE TECHNIQUE

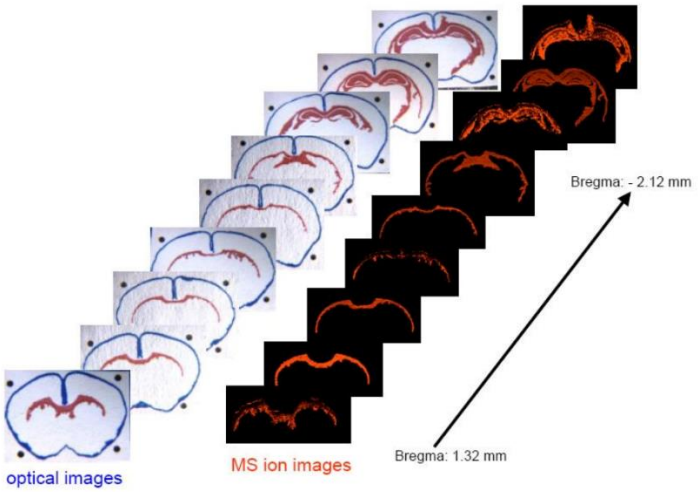


Blue ink→ signals at m/z 584 and 1166

Red ink → signal at m/z 342

Black ink → signals with m/z below 300

The expression of the proteins in the whole sample is mimicked



What information can we learn from IMS?

Similar info to immunohistochemistry (IHC) --> one of the more efficient but you need antibodies or fluorescence tagging

A limitation is that we cannot cover all the proteins in a sample: not all proteins will ionize.

How IMS differs:

• More Versatile

- No prior knowledge or antibodies needed

- Not limited to proteins- usable to examine lipids, metabolites, peptides, small molecules

• Examine distribution many molecules in same image

• Lower spatial resolution

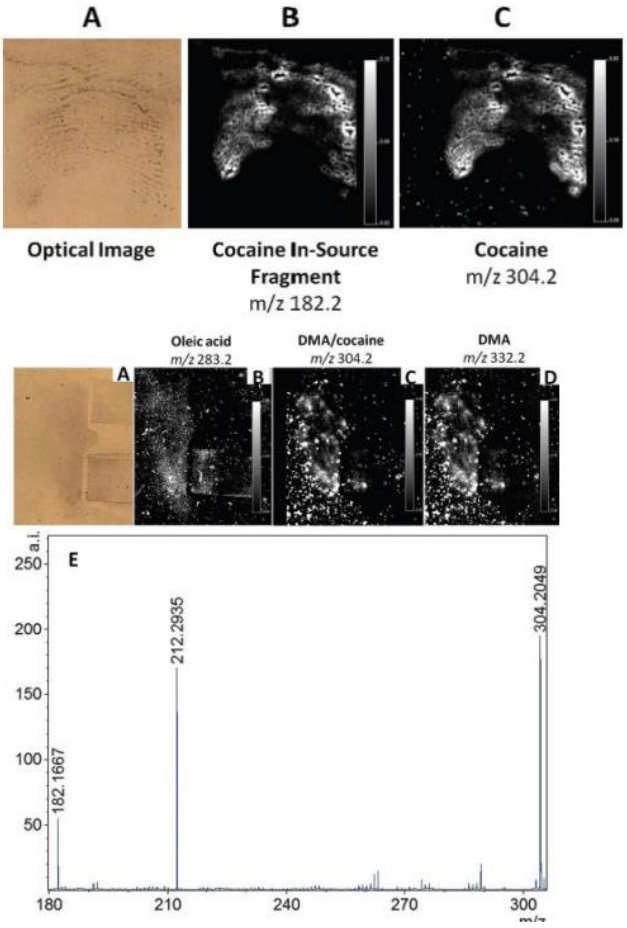
• Not comprehensive- not all molecules ionize well

SOME APPLICATIONS: MALDI IMAGINE IN THE CRIME SCENE

MALDI MSI analysis of a primary fingermark lift recovered on a textured light frame at a seized cannabis farm. MALDI MSI analysis of a primary fingermark lift recovered on the interior of a window frame following enhancement with carbon black powder. (A) Shows the optical image of the enhanced mark; (B) and (C) show the MALDI MS images of two ions at m/z 304.1 and 182.1, normalized against the matrix peak at m/z 190.0, which were assigned tentatively to

cocaine and its in source ion fragment respectively; these ions were later confirmed using by transfer fragmentation analysis.

MALDI MSI analysis of a primary fingermark lift recovered on a textured light plug socket at a seized cannabis farm. (A) Shows the optical image of the primary lift after enhancement with aluminium powder; (B) shows the molecular distribution of oleic acid at m/z 283.2; (C)–(D) show the molecular images of two ions at m/z 304.2 and 332.2, normalised against the matrix peak at m/z 190.0, which were tentatively assigned to n-alkyl dimethylbenzylammonium ion (DMA), with the possibility of the former ion actually being cocaine, and with the latter ion carrying two extra CH2 units for DMA. The MS/MS spectrum in panel E confirms the presence of both DMA and cocaine which share the same molecular mass through their characteristic ion fragments at m/z 212.2934 and 182.1667 respectively.



MALDI IMS and diagnostics

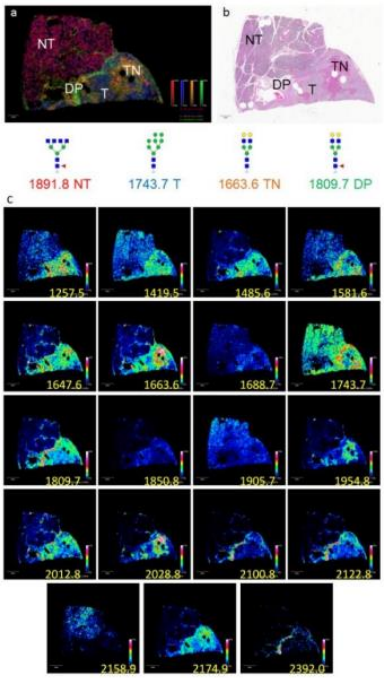
MALDI imaging mass spectrometry (MALDI IMS) technology creates a link between the molecular assessment of numerous molecules and the morphological information about their spatial distribution.

e.g.The application of MALDI IMS on formalin-fixed:

MALDI-IMS of a Human Pancreas FFPE Tissue Block. An FFPE block of pancreatic tissue from a human patient was cut at 5 um prior to and selected for MALDI-IMS. Histopathology found four unique regions in this tissue block.

The tissue block contained tumor tissue, non-tumor tissue, fibroconnective tissue representing desmoplasia surrounding the tumor tissue, and necrotic tissue (b). MALDI-IMS was able to distinguish these four regions based on specific ions after MALDI-IMS. M/z = 1891.80 (red) is found in the non-tumor (NT) region of the pancreas and corresponds to ex3dHex1HexNAc6, while m/z = 1743.64 (blue) represents Hex8HexNAc2 and is predominant in the tumor region (T) of

the tissue. Desmoplasia (DP) is represented by m/z = 1809.69 (green) corresponding to Hex5dHex1HexNAc4. In the region where necrosis was identified (TN), m/z = 1663.64 (orange) was elevated corresponding to Hex5HexNAc4.

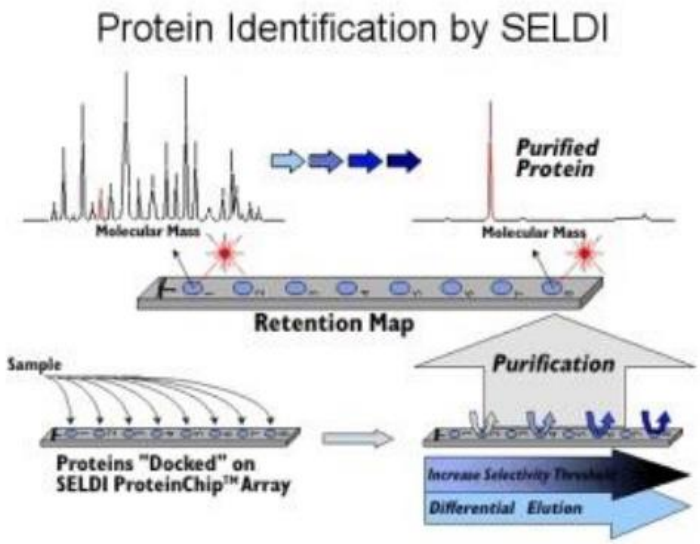


Surface Enhanced Laser Desorption/Ionization (SELDI) technology:

The information achieved with proteomics can go also further: the MALDI plate is a plate with inert material that doesn't react with my molecules. But with a functionalized material (the surface is not inert, it can bind my molecule, with some washing steps I can collect the molecule and get rid of the others) I could enrich my target protein or analyte.

Surface Enhanced Laser Desorption/Ionization (SELDI) technology

SELDI-TOF-MS is a variation of matrix-assisted laser desorption/ionization (MALDI) that uses a target modified to achieve biochemical affinity with the analyte compound.



APPLICATIONS--> in all cases that we apply the MALDI

• Fingerprinting/Biomarker discovery: Detect proteins from two different states e.g. disease versus normal tissue samples.

• Toxicology Detection of toxicity biomarkers

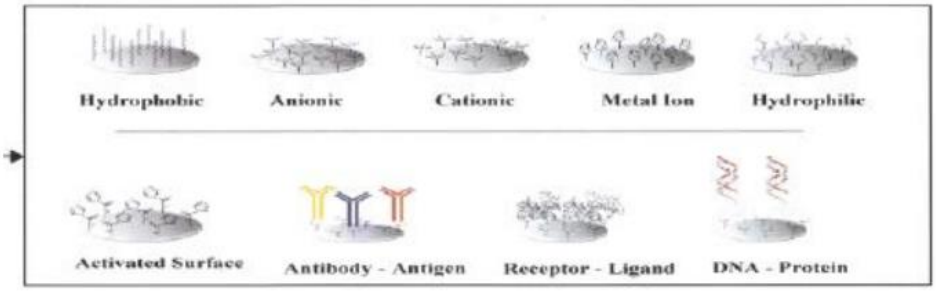
• Protein characterization Including the identification of novel ligands and protein characteristics.

In SELDI, the protein mixture is spotted on a surface modified with a chemical functionality. Some proteins in the sample bind to the surface, while the others are removed by washing.

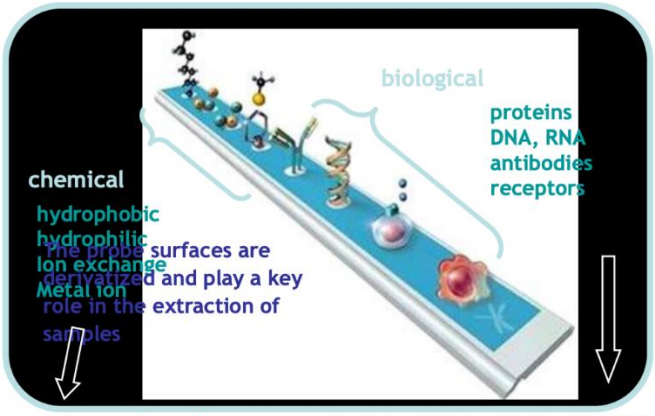
After washing the spotted sample, the matrix is applied to the surface and allowed to crystallize with the sample peptides.

Binding to the SELDI surface acts as a separation step and the subset of proteins that bind to the surface are easier to analyze.

Common surfaces include CM10 (weak-positive ion exchange), H50 (hydrophobic surface, similar to C6-C12 reverse phase chromatography), IMAC30 (metal-binding surface), and Q10 (strong anion exchanger). Surfaces can also be functionalized with antibodies, other proteins, or DNA.



Surfaces



Designed for *categories of proteins*

Designed for the interaction with a *single target protein*