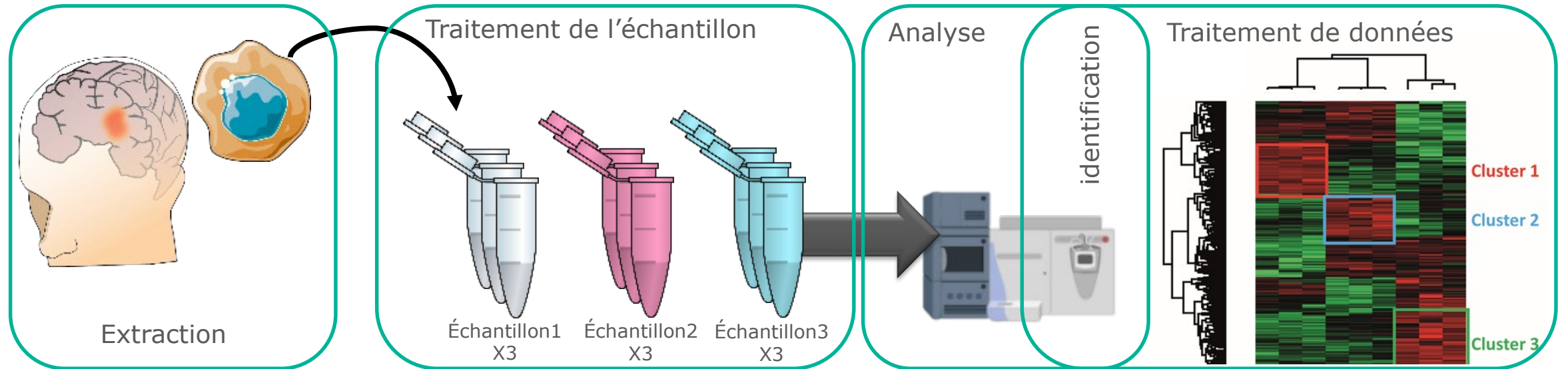


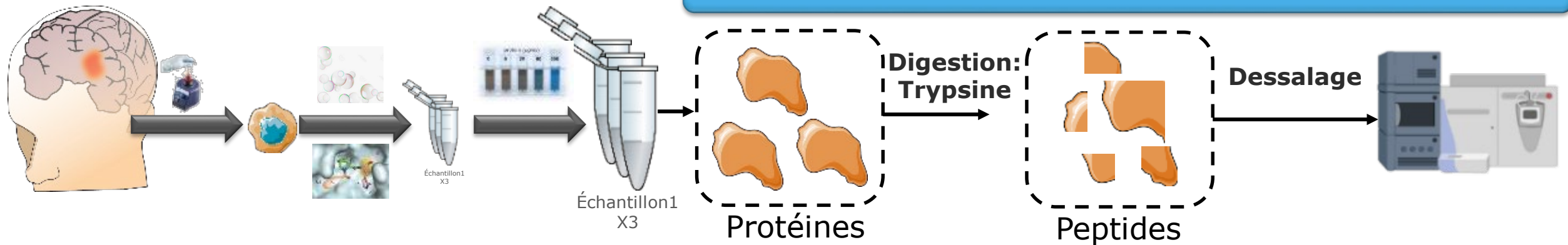
Bases de la protéomique

Principes et Applications

Les étapes pour la protéomique



Shotgun

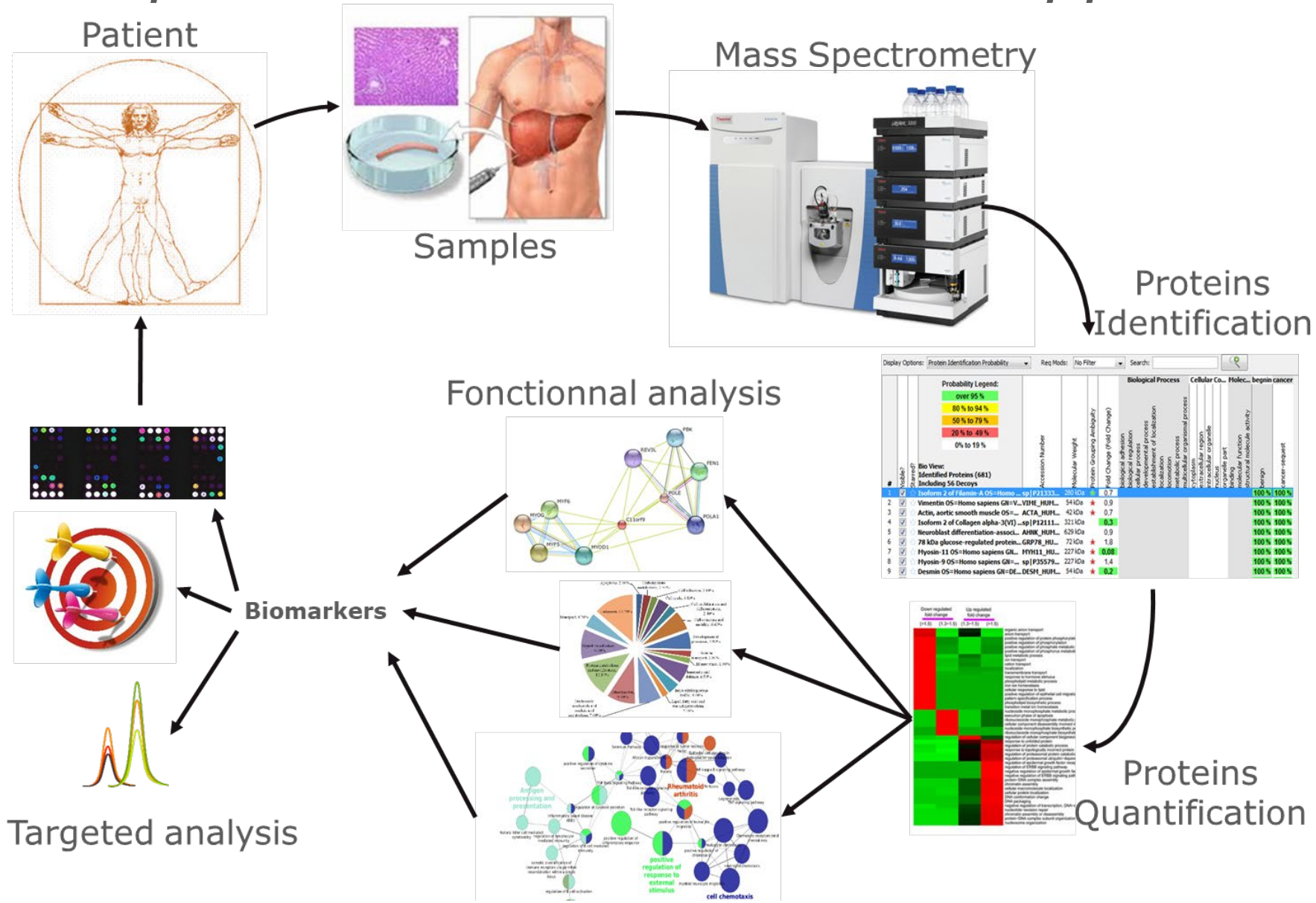


Quel sont les sites de clivage de la trypsine ?

Comment retire t-on les sels de l'échantillon ?

Mass spectrometry-based proteomics

La spectrométrie de masse en application clinique



Protein databases

Mix of proteins

Enzyme digestion

Parent selection

Experimental spectra

Proteins level

Peptides level

Peptide level

MS/MS

ADFALKLEVDSTR
PWRSDIGSETEKQ
DHVRELLVACMP
GEFWRSTYKDIST

Proteins
databank

ADFALK
LEVDSTR
PWR
SDIGSETEK

In silico
digestion

m/z 1403,21
SDIGSETEK
IDSSEGKET

In silico
peak list

Theoretical
spectra

Compare

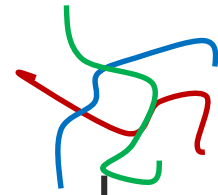
Search engine

Score

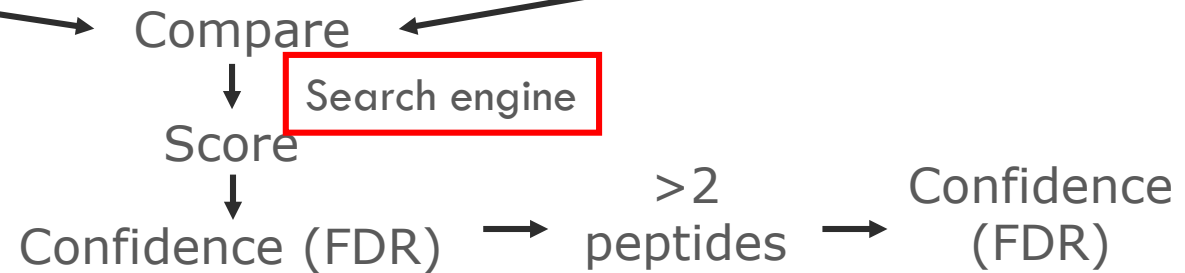
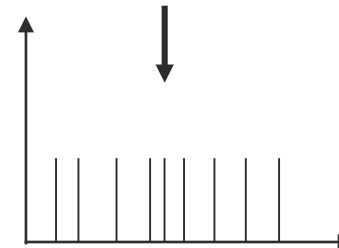
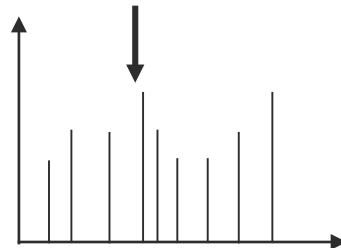
Confidence (FDR)

>2
peptides

Confidence
(FDR)



m/z 1403,21



Identification Protéique = Stratégie et instrument dédiées

Data analysis = étape critique

Logiciel dédié à la protéomique et à l'approche MS

Banque de données protéique (protein database) = **répertoire contenant les séquence en Acide Aminées des protéines**

Différente banque de données protéique:

- **Uniprot**: banque contenant les références SwissProt et Trembl (prédiction de séquence)

Nextprot: protéines humaines seulement + outil « *Peptide uniqueness checker* »

Protein databases

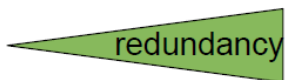
The image displays two screenshots of protein databases. The top screenshot is the UniProt homepage, featuring a search bar, navigation links (BLAST, Align, Retrieve/ID mapping, Peptide search, SPARQL), and a mission statement. It highlights UniProtKB (Swiss-Prot and TrEMBL), UniRef, UniParc, and Proteomes. A news section on the right mentions SARS-CoV-2 protein updates. The bottom screenshot is the NextProt homepage, showing a search interface with 'Simple search' and 'Advanced search (SPARQL)' options. It includes a 'Getting started' section with links to tools and a 'Data sources' section listing various databases. A 'Release 2020-07-17' section features a donut chart showing protein existence statistics: Predicted (46), Uncertain (609), Inferred from homology (68), Evidence at protein level (18263), and Evidence at transcript level (1187).



UniProtKB/ Swiss-Prot

high-quality manual annotation,
reviewed

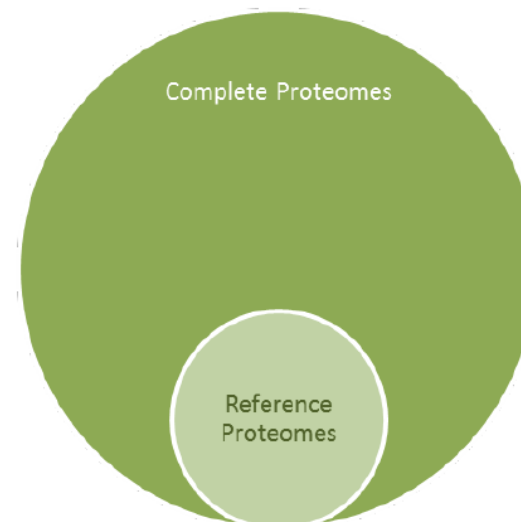
1 record per gene
in one species



UniProtKB/ TrEMBL

automatically annotated,
unreviewed

1 record for 100% identical full-
length sequences in one species



Complete proteomes

Complete sets of proteins thought to be expressed by organisms whose genomes have been completely sequenced.

Reference proteomes

Some proteomes have been (manually and algorithmically) selected as **reference proteomes**. They cover well-studied model organisms and other organisms of interest for biomedical research and phylogeny.

UniProtKB

Swiss-Prot
(546,238)

Manually
annotated and
reviewed.

UniRef

Sequence clusters

UniParc

Sequence archive

Proteomes

Supporting data

Literature citations

Taxonomy

Subcellular locations

TrEMBL
(82,126,897)

Automatically
annotated and not
reviewed.

Cross-ref. databases

Diseases

Keywords

UniProt

Proteomes

homo sapiens

Advanced

Search

BLAST

Align

Retrieve/ID mapping

Help

Contact

Proteomes results

About Proteomes

Filter by:

Download

Repeat search in UniProtKB (999,811)

	Proteome ID	Organism	Organism ID	Last modified	Protein count
<input type="checkbox"/>	UP000005640	Homo sapiens (Human)	9606	2015-08-09	70071

1 to 1 of 1

Show 25

Demo

Help video

UniProt

UniProtKB

organism:"Homo sapiens (Human) [9606]"

Advanced

Search

BLAST

Align

Retrieve/ID mapping

Help

Contact

UniProtKB results

About UniProtKB

Basket

Filter by:

Reviewed (20,196)

Unreviewed (128,790)

Popular organisms

Human (148,986)

Proteomes

UP000005640 (70,071)

View by

Taxonomy

BLAST

Align

Download

Add to basket

Columns

1 to 26 of 148,986

Show 25


Expand search "Homo sapiens (Human) [9606]" to include lower taxonomic ranks


	Entry	Entry name	Protein names	Gene names	Organism	Length
<input type="checkbox"/>	P31946	1433B_HUMAN	14-3-3 protein beta/alpha	YWHAB	Homo sapiens (Human)	246
<input type="checkbox"/>	P62258	1433E_HUMAN	14-3-3 protein epsilon	YWHAЕ	Homo sapiens (Human)	255
<input type="checkbox"/>	Q04917	1433F_HUMAN	14-3-3 protein eta	YWHAH, YWHA1	Homo sapiens (Human)	246
<input type="checkbox"/>	P61981	1433G_HUMAN	14-3-3 protein gamma	YWHAG	Homo sapiens (Human)	247
<input type="checkbox"/>	P31947	1433S_HUMAN	14-3-3 protein sigma	SFN, HME1	Homo sapiens (Human)	248
<input type="checkbox"/>	P27348	1433T_HUMAN	14-3-3 protein theta	YWHAQ	Homo sapiens (Human)	245

UniProtKB results

Basket

UniProtKB consists of two sections:

**Reviewed (Swiss-Prot) - Manually annotated**
Records with information extracted from literature and curator-evaluated computational analysis.

**Unreviewed (TrEMBL) - Computationally analyzed**
Records that await full manual annotation.

The UniProt Knowledgebase (UniProtKB) is the central hub for the collection of functional information on proteins, with accurate, consistent and rich annotation. In addition to capturing the core data mandatory for each UniProtKB entry (mainly, the amino acid sequence, protein name or description, taxonomic data and citation information), as much annotation information as possible is added.

[Help](#) [UniProtKB help video](#) [Other tutorials and videos](#) [Downloads](#)

Filter by

Reviewed (176)
Swiss-Prot

Unreviewed (1,202)
TrEMBL

Popular organisms

Human (20)

Mouse (7)








Rat (4)

Bovine (3)

Zebrafish (2)

BLASTAlignDownloadAdd to basketColumns

1 to 25 of 1,378 Show 25

	Entry	Entry name		Protein names	Gene names	Organism	Length	
<input type="checkbox"/>	P02144	MYG_HUMAN		Myoglobin	MB	Homo sapiens (Human)	154	
<input type="checkbox"/>	P04247	MYG_MOUSE		Myoglobin	Mb	Mus musculus (Mouse)	154	
<input type="checkbox"/>	P02192	MYG_BOVIN		Myoglobin	MB	Bos taurus (Bovine)	154	
<input type="checkbox"/>	Q9QZ76	MYG_RAT		Myoglobin	Mb	Rattus norvegicus (Rat)	154	
<input type="checkbox"/>	P84997	MYG_BUBBU		Myoglobin	MB	Bubalus bubalis (Domestic water buffalo)	154	
<input type="checkbox"/>	Q6VN46	MYG_DANRE		Myoglobin	mb zgc:65819	Danio rerio (Zebrafish) (Brachydanio rerio)	147	
<input type="checkbox"/>	P00016	MYG_XENLA		Myoglobin	mb	Xenopus laevis (Claw toad)	147	

Display

- Entry
- Publications
- Feature viewer
- Feature table

- None
- Function
- Names & Taxonomy
- Subcellular location
- Pathology & Biotech
- PTM / Processing
- Expression
- Interaction
- Structure
- Family & Domains
- Sequences (1+)
- Similar proteins
- Cross-references
- Entry information
- Miscellaneous

▲ Top

BLAST Align Format Add to basket History

Protein

Myoglobin

Gene

MB

Organism

Homo sapiens (Human)

Status

Reviewed

- Annotation score: ●●●●● - Experimental

Function

Serves as a reserve supply of oxygen and facilitates the movement of ox

Sites

Feature key	Position(s)	Description
Metal binding ⁱ	65	Iron (heme distal ligand) ⓘ 1 Pu
Metal binding ⁱ	94	Iron (heme proximal ligand) ⓘ

GO - Molecular functionⁱ

- heme binding ⓘ Source: InterPro
- metal ion binding ⓘ Source: UniProtKB-KW
- oxygen binding ⓘ Source: GO_Central ▾
- oxygen carrier activity ⓘ Source: UniProtKB-KW

Complete GO annotation on QuickGO ...

GO - Biological processⁱ

- brown fat cell differentiation ⓘ Source: Ensembl
- enucleate erythrocyte differentiation ⓘ Source: Ensembl
- heart development ⓘ Source: Ensembl
- oxygen transport ⓘ Source: GO_Central ▾
- response to hormone ⓘ Source: Ensembl

Names & Taxonomyⁱ

Protein names ⁱ	Recommended name: Myoglobin
Gene names ⁱ	Name:MB
Organism ⁱ	Homo sapiens (Human)
Taxonomic identifier ⁱ	9606 [NCBI]
Taxonomic lineage ⁱ	Eukaryota > Metazoa > Chordata > Craniata > Vertebrata > Euteleostomi > Mammalia > Eutheria > Euarchontoglires > Prim
Proteomes ⁱ	UP000005640 Component ⁱ : Chromosome 22

Organism-specific databases

EuPathDB ⁱ	HostDB:ENSG00000198125.12
HGNC ⁱ	HGNC:6915, MB
MIM ⁱ	160000, gene
neXtProt ⁱ	NX_P02144

Subcellular locationⁱ

Graphics by Christian Stolle & Seán O'Donoghue; Source:COMPARTMENTS

Manual annotation Automatic computational assertion

GO - Cellular component

Cytosol

cytosol ⓘ Source: Reactome

Extracellular region or secreted

extracellular exosome ⓘ Source: UniProtKB ▾

Complete GO annotation on QuickGO ...

Pathology & Biotechⁱ

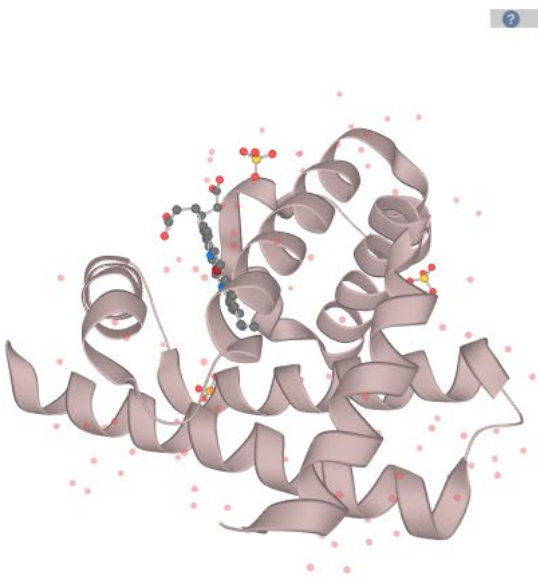
Molecule processing

Feature key	Position(s)	Description	Actions	Graphical view	Length
Initiator methionine ⁱ		Removed 2 Publications ▾			
Chain ⁱ (PRO_0000053303)	2 – 154	Myoglobin	Add BLAST		153

Amino acid modifications

Feature key	Position(s)	Description	Actions	Graphical view	Length
Modified residue ⁱ	4	Phosphoserine By similarity ▾			1
Modified residue ⁱ	68	Phosphothreonine By similarity ▾			1

Structureⁱ



PDB Entry	Method	Resolution	Chain	Positions	Links
3RGK	X-ray	1.65 Å	A	2-154	PDBe RCSB PDB PDBj PDBsum

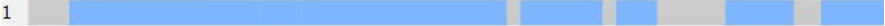
Sequence (1+)ⁱ

Sequence statusⁱ: Complete.
Sequence processingⁱ: The displayed sequence is further processed into a mature form.
This entry has 1 described isoform and 5 potential isoforms that are computationally mapped. [Show all](#) [Align All](#)

P02144-1 [UniParc] [FASTA](#) [Add to basket](#)
[Hide](#)

10	20	30	40	50
MGLSDGEWQL	VLNVNGKVEA	DIPGHGQEV	IRLFKGPET	LEKFDKFKHL
60	70	80	90	100
KSEDEMKASE	DLKKHGATVL	TALGGILKKK	GHHEAEIKPL	AQSHATKHKI
110	120	130	140	150
PKYLEFISE	CIIQVLQSKH	PGDFGADAQG	AMNKALELFR	KDMASNYKEL
GFQG				

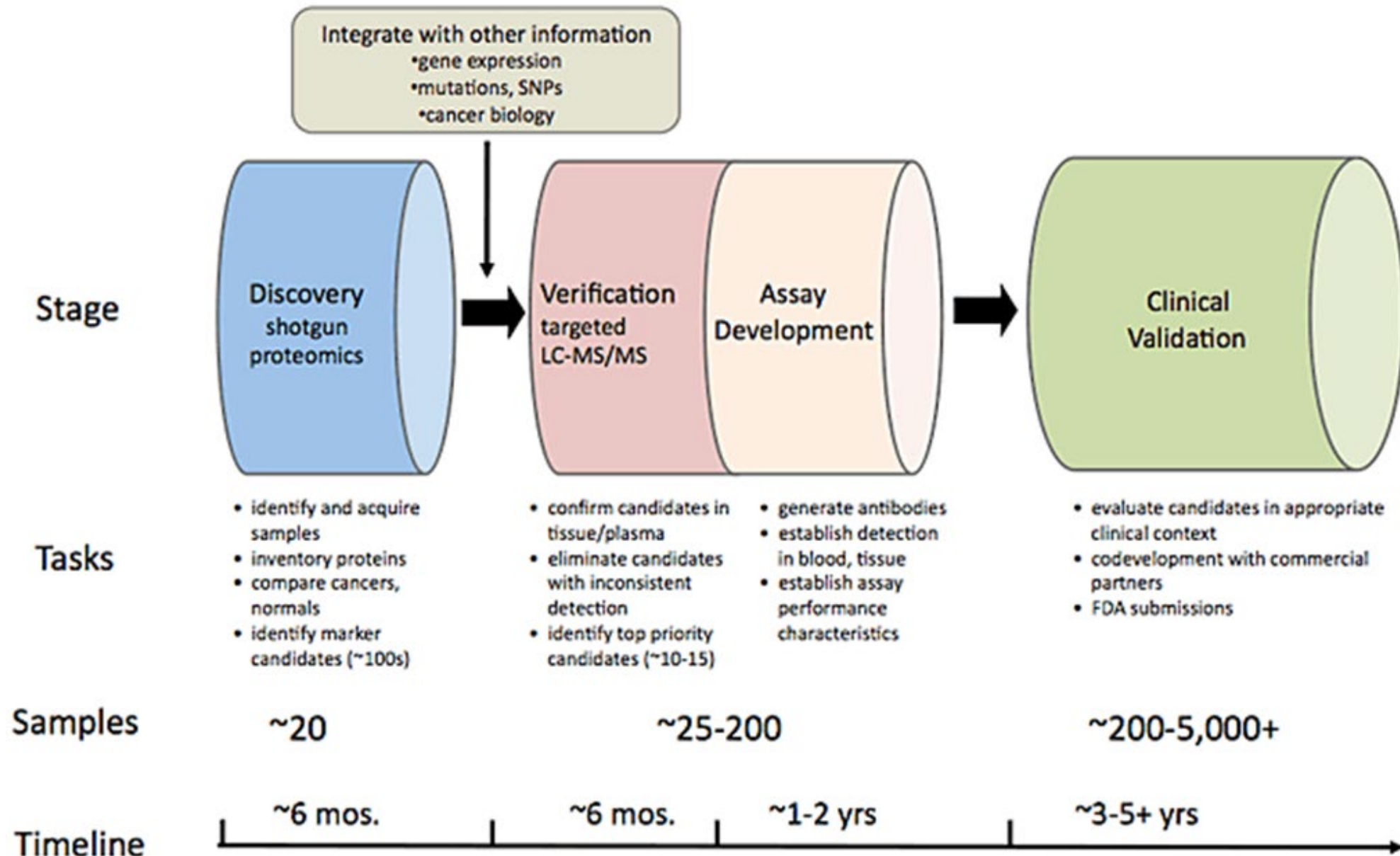
Secondary structure



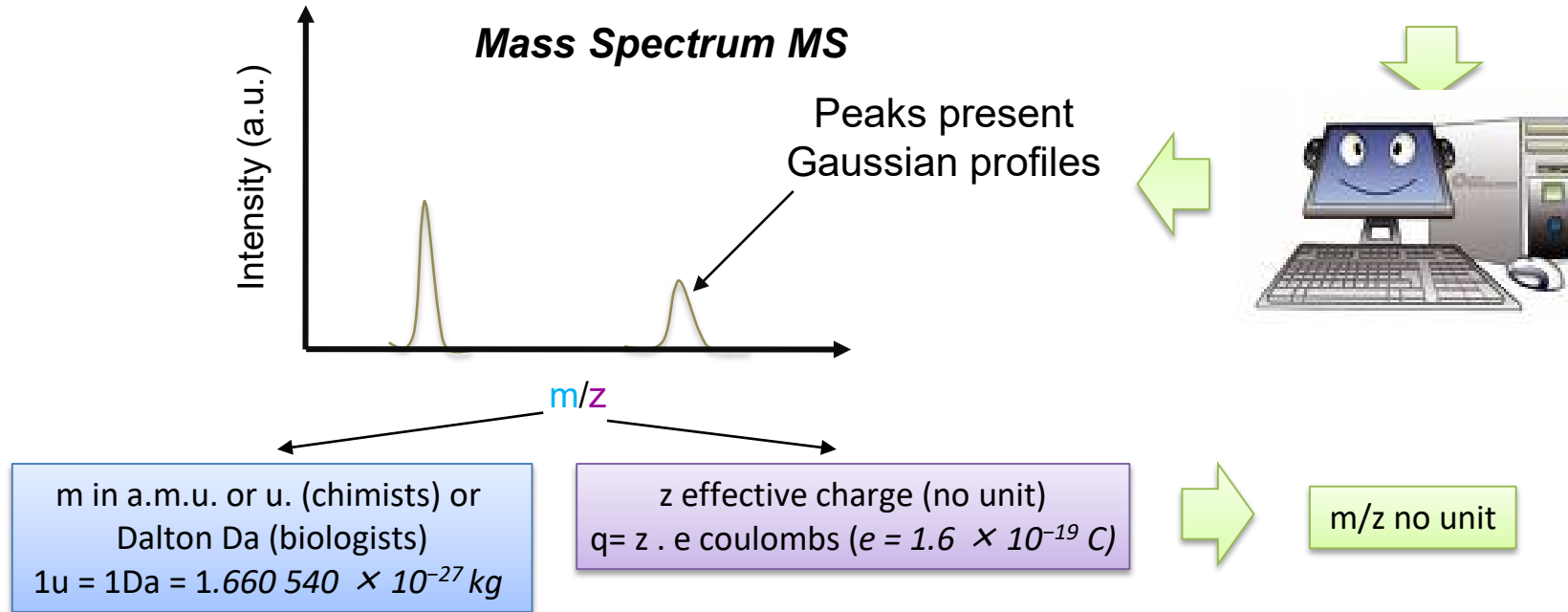
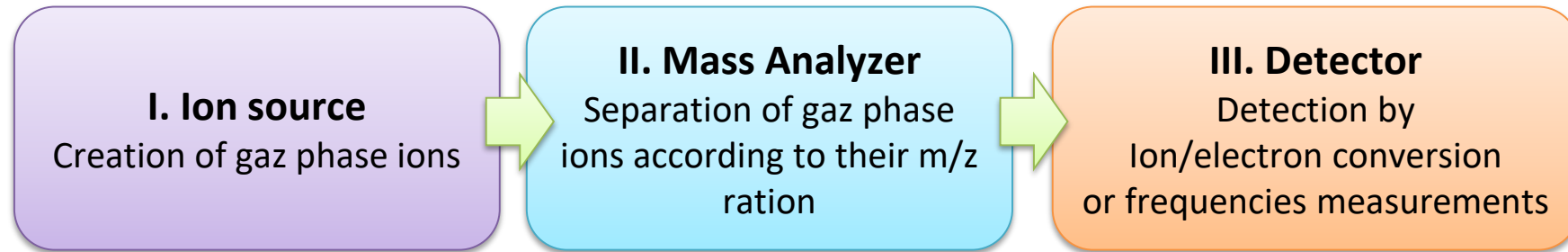
Legend: ■ Helix ■ Turn ■ Beta strand ■ PDB Structure known for this area

Length: 154
Mass (Da): 17,184
Last modified: January 23, 2007 - v2
Checksumⁱ: F6A41F19A525F09C
[BLAST](#) [GO](#)

Mass spectrometry



Mass spectrometry

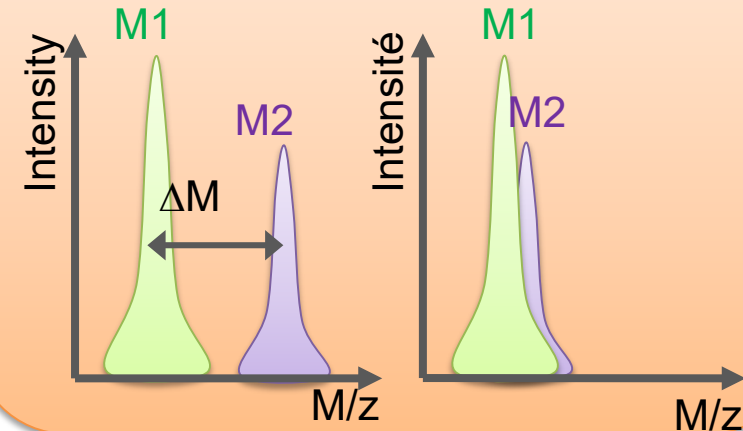


Since 1961 the *unified atomic mass [u.]* is defined as 1/12 of the mass of one atom of nuclide ^{12}C which has been assigned to 12 u. exactly by convention

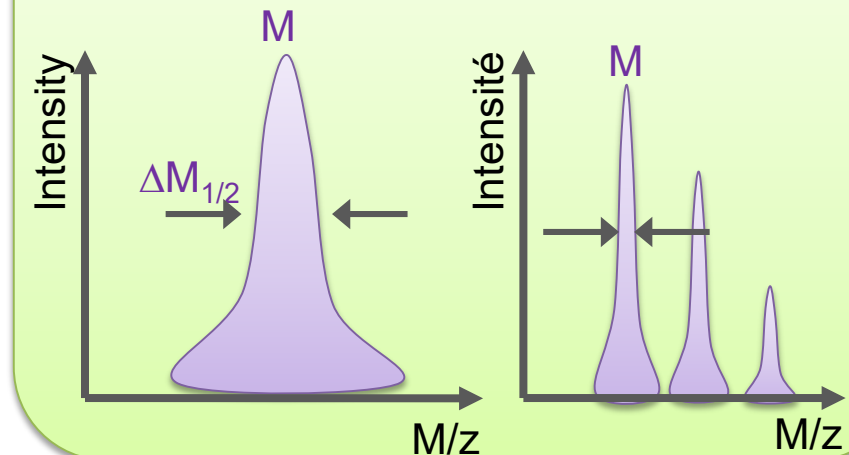
Characteristics of the analyzer

- ✓ Sensitivity
- ✓ LOD (Limit of Detection)
- ✓ Mass range
- ✓ Resolving power (RP)
 - Definition $RP = M/\Delta M$
 - Capacity to separate two peaks
 - Ex: $RP = 2000$ & $M = 2000$ peaks distant by 1 mass unit are separated (2000/2001)
 - $RP = 2000$ & $M = 200$ peaks distant by 0.1 mass unit are separated (200/200,1)
- ✓ Résolution
 - Definition $Rs = M/\Delta M_{1/2}$
 - $M = m/z$ value of a peak
 - ΔM Peak width at half maximum (FWHM Full Width at Half Maximum)
 - Describe the peak width
 - Ex: Quadrupoles $R =$ up to a few 1000's
 - Time-of-Flight $R =$ up to 30 000
 - FT-ICR $R =$ up to 1 000 000

Resolving Power



Mass resolution



Accuracy & precision

Mass accuracy = **exactitude**

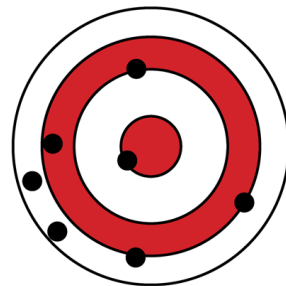
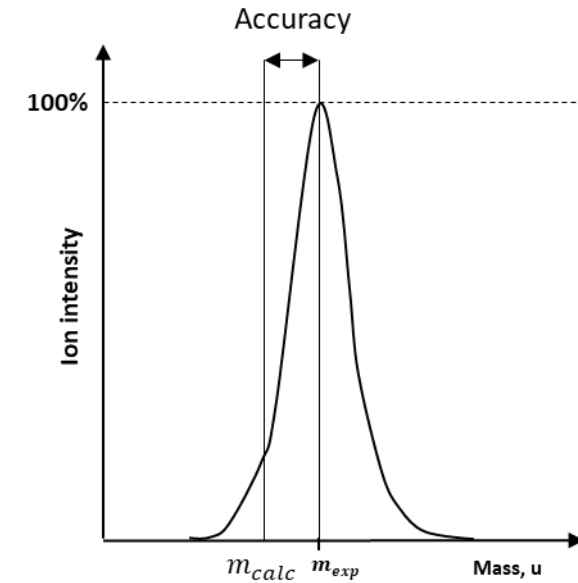
$$\text{Accuracy} = \left(\frac{m_{\text{exp}} - m_{\text{calc}}}{m_{\text{calc}}} \right) \times 10^6 \text{ ppm}$$

Theoretical mass: 400.000

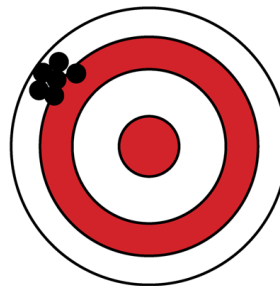
Measured Mass mesurée: 400.003

Difference: 0.003

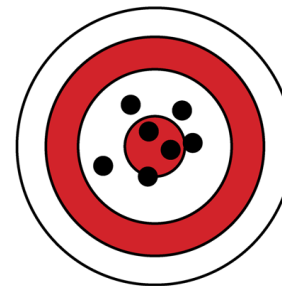
Error: $\frac{0.003}{400.000} \times 10^6 = 7.5 \text{ ppm}$



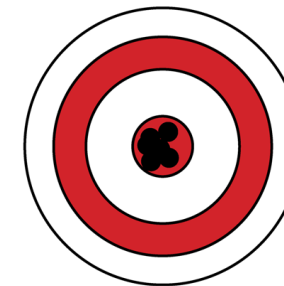
Low accuracy
Low precision



Low accuracy
High precision



High accuracy
Low precision



High accuracy
High precision

La gamme dynamique

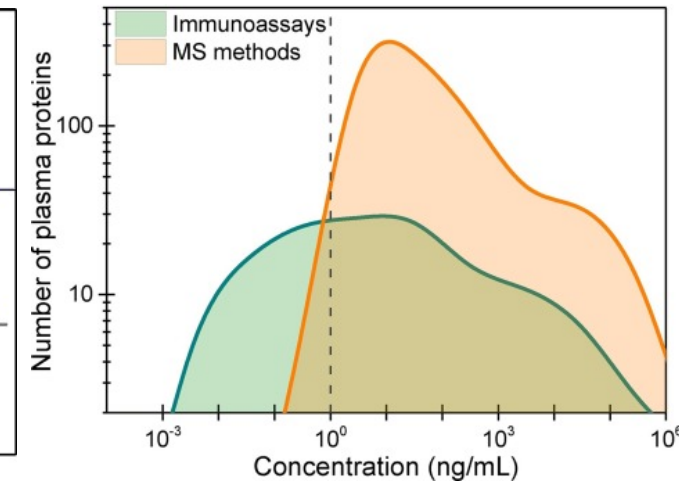
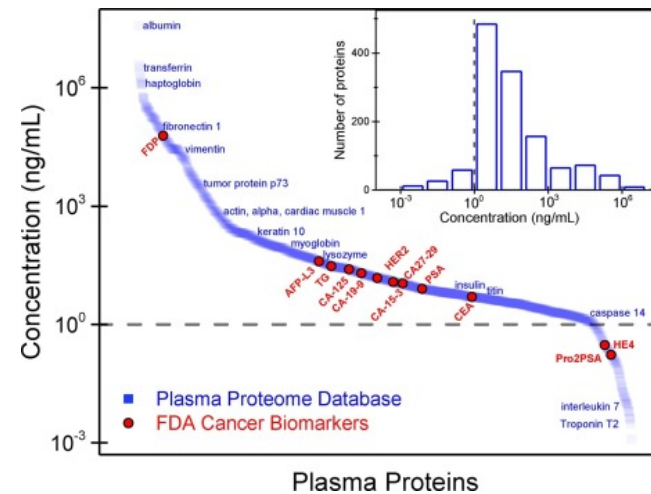
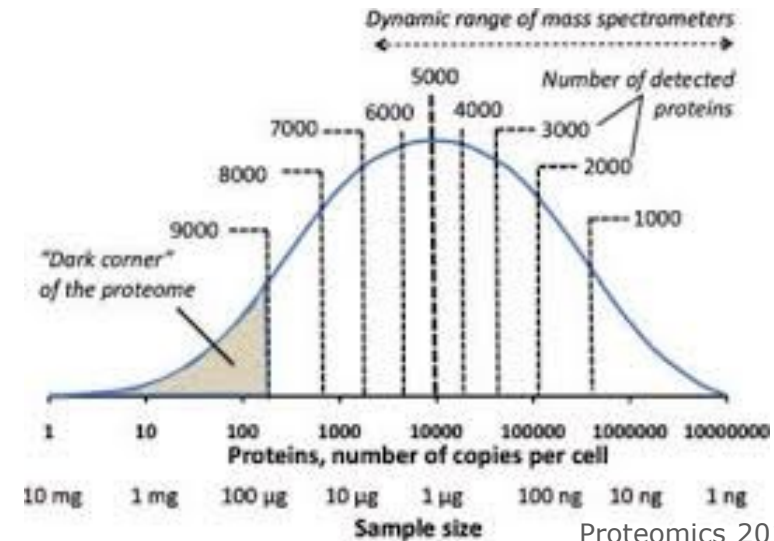
Capacité de la MS a détecter des molécules avec une grande différence de masse

Certaine protéine on une différence d'expression très grande (**ordre de magnitude 7**) :

De 1 copie par cellule contre 10 million

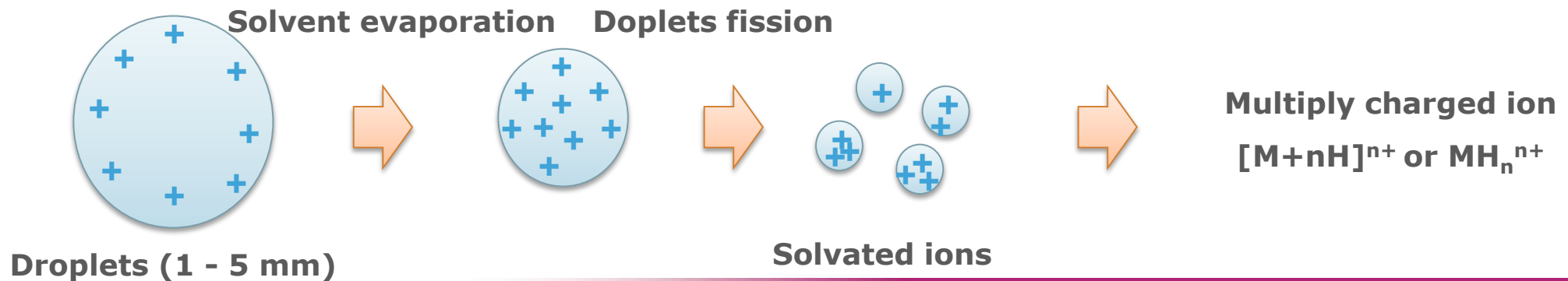
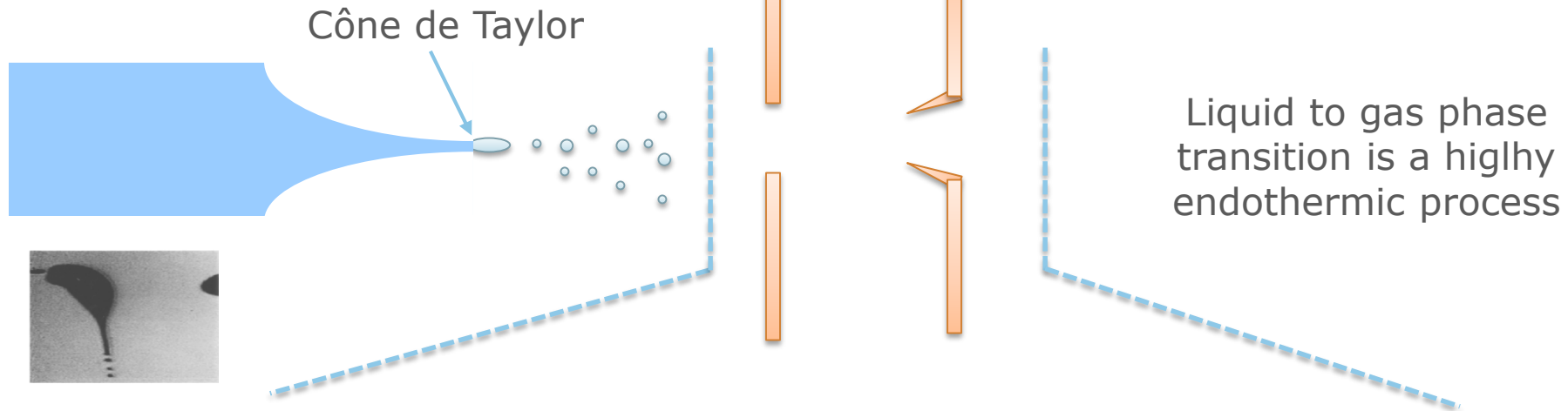
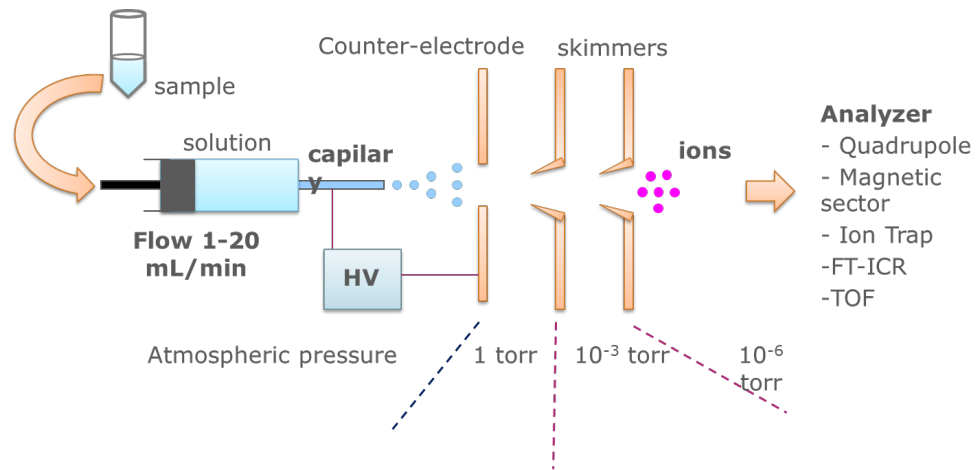
La gamme dynamique en MS vas jusqu'à un ordre de magnitude 4 à 5

Besoin de séparer les molécules entre elles



Seminars in Cancer Biology Vol 52, Part 1, 2018, 26-38

Electrospray ionization



Chromatographie Liquide

mixture of components



column
(stationary phase)

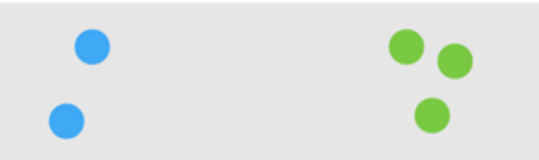
flow of mobile phase



most hydrophobic components interact with the column best



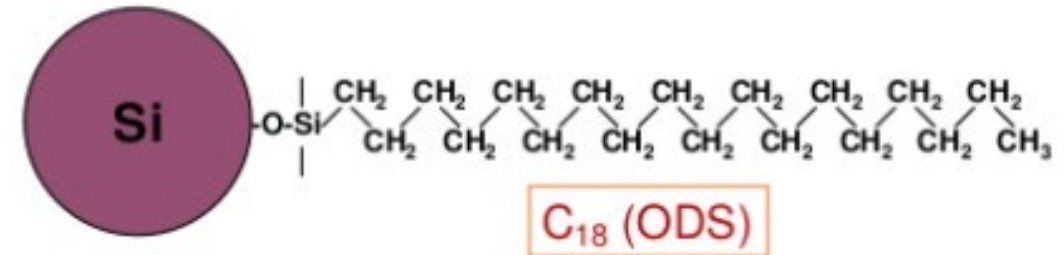
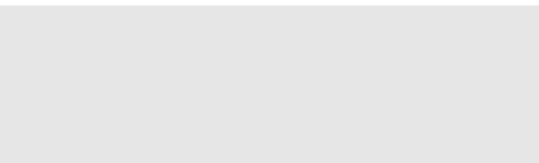
least hydrophobic components elute first



product elutes



most hydrophobic components elute last



MS-based proteomics

MS-based analysis

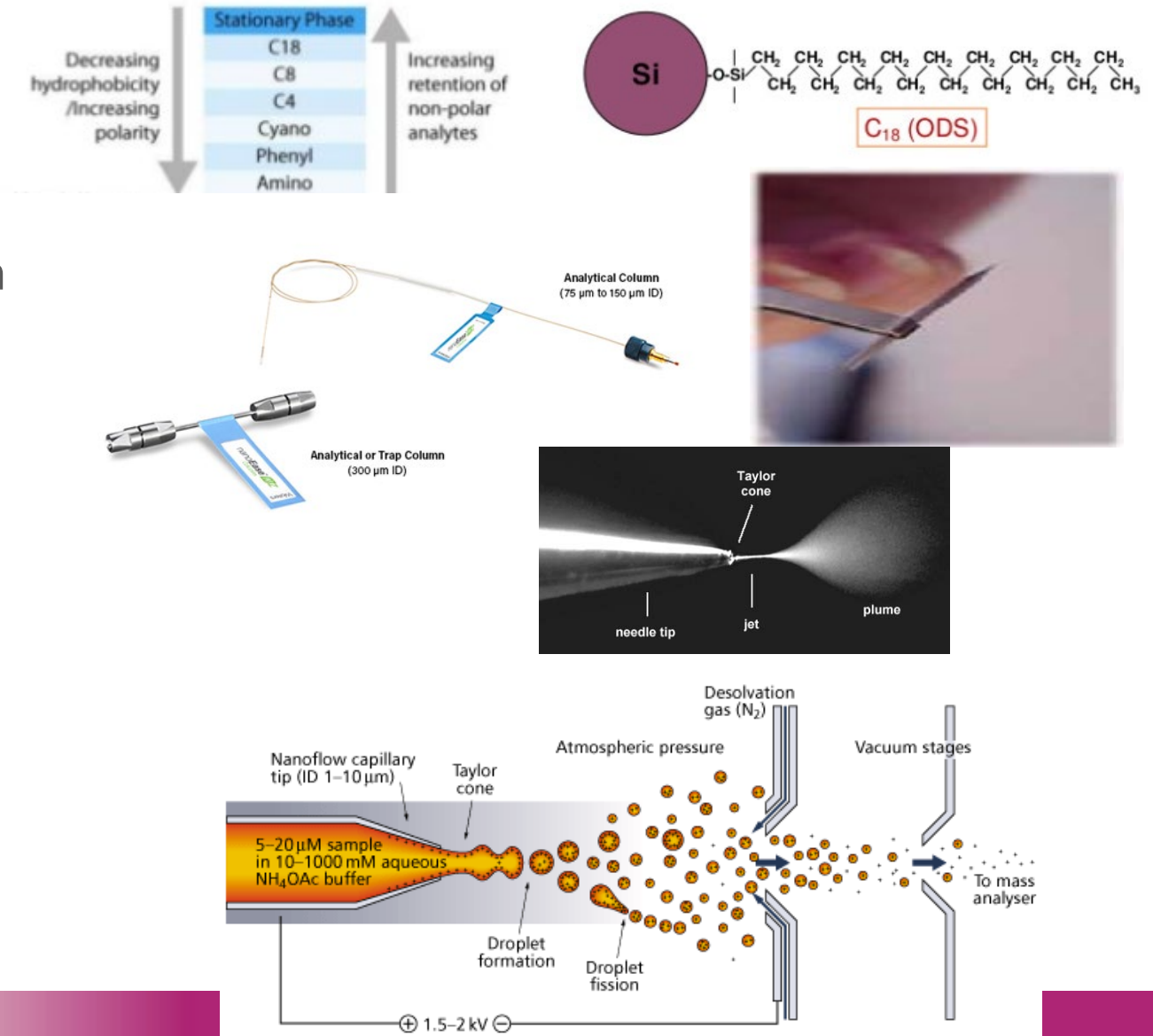
Séparation des peptides par chromatographie liquide couplée à une source nanoESI

Colonne en phase inverse C18 de 5cm à >1m

Utilisation de **nano débit** « nanoflow » (300 nanolitre/minute)

Chromatographie spécialisée (High pressure) (HPLC or uHPLC)

élution par gradient d'Acétonitrile



Identification protéique nécessite des instruments performants

Dont dépend « l'accuracy » et la précision de mesure

L'efficacité de fragmentation

La stratégie MS/MS utilisable :

- DDA
- DIA
- Ciblé...

