

PROTEINOMIQUE

Plan

- Rappels sur la traduction
 - Définitions
 - Les acteurs de la synthèse protéique
 - Mécanisme de la traduction de l'ARNm
- La protéomique
 - Définition-objectifs
 - Human Proteome Project (HPP)
 - Profilage ou Caractérisation des Protéines
- Technique d'analyse du proteome: Spectrophotométrie de Masse (SM)
 - Définition
 - Principes
 - Identification des protéines

Rappels sur la traduction

- **Les protéines** sont des molécules composées d'un enchaînement d'acides aminés. Elles remplissent de nombreuses fonctions et contribuent ainsi à la réalisation du phénotype d'un organisme.
 - Protéine de structure
 - Hormone
 - Enzyme
 - Protéine de transport
 - Immunité
 - Régulation des processus biologique
- **La synthèse des protéines** résulte de l'expression de l'information génétique. La séquence en nucléotides d'un gène portée par une molécule d'acide désoxyribonucléique (ADN), code la séquence en acides aminés d'une protéine donnée en passant par la transcription en ARNm. Elle se déroule au niveau du cytoplasme et au niveau des ribosomes

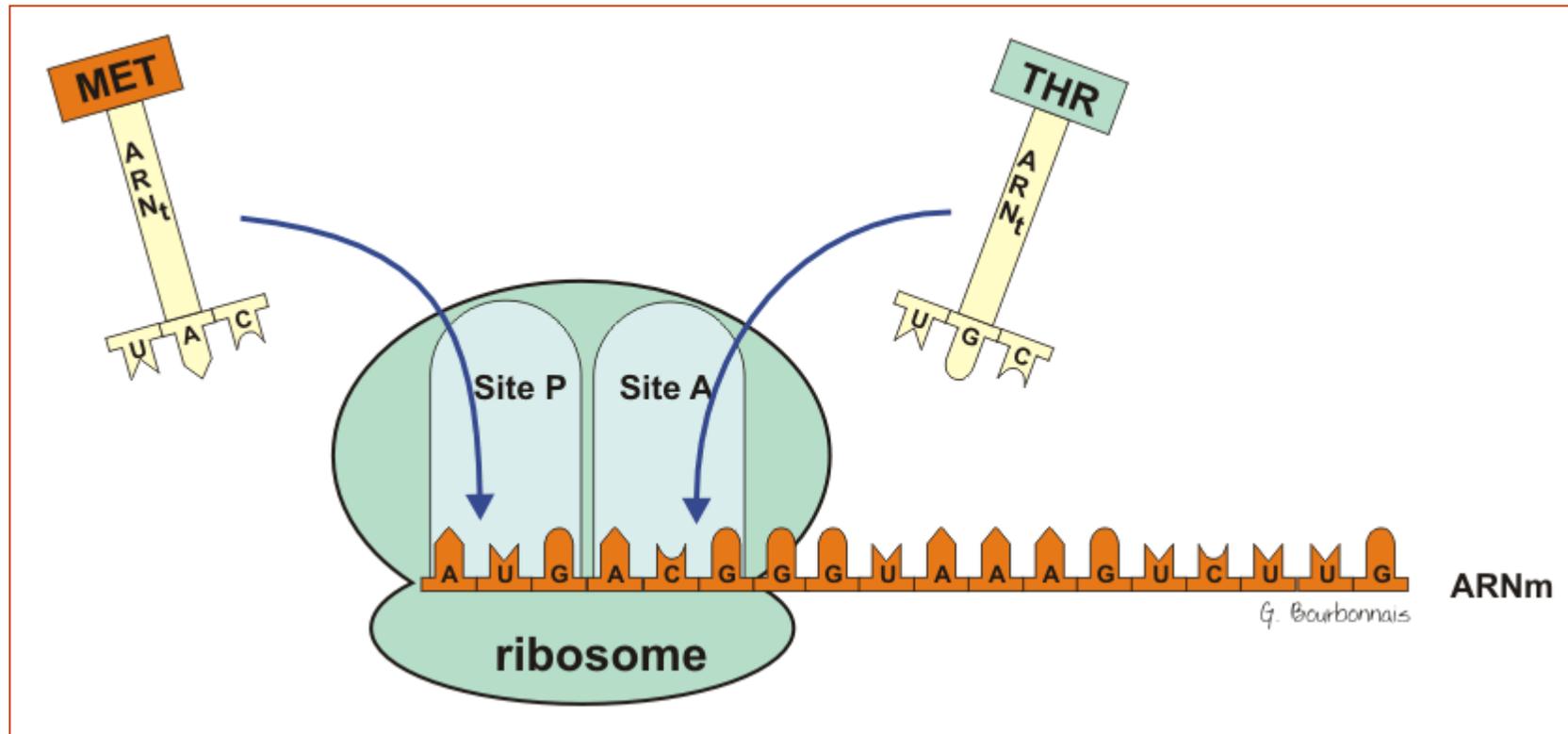
- On parle de **peptide** pour désigner des chaînes d'acides aminés dont la longueur n'excède pas **20 à 30 acides aminés**. On parle de **polypeptide** pour désigner des chaînes d'acides aminés qui possèdent **plus de 30 acides aminés**. On parle de **protéine** pour désigner la **molécule fonctionnelle**. Une protéine peut être formée d'une ou de plusieurs chaînes polypeptidiques.

Les acteurs de la synthèse protéique

Pour synthétiser une protéine

- ADN : Brin codant
- ARNm = messenger de l'information
- Ribosome = machine à assembler les acides aminés
- Acides aminés = pièces de construction
- ARNt (ARN de transfert) = molécules qui transportent les acides aminés du cytoplasme au ribosome où ils sont assemblés en protéine.

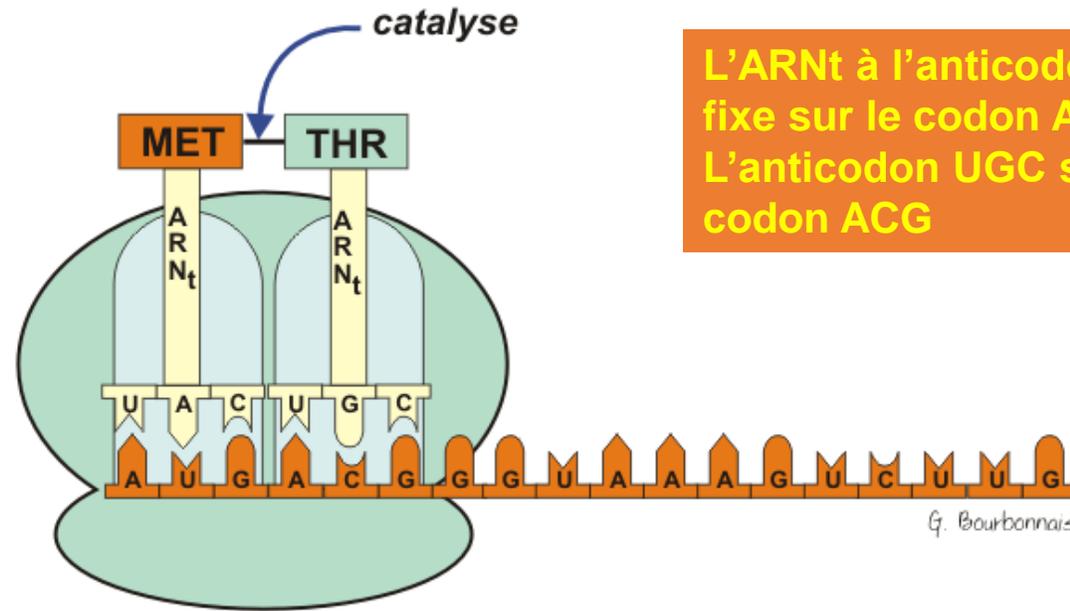
Mécanisme de la traduction



Le brin d'ARNm s'attache au ribosome.
En fait, il s'attache d'abord à la petite unité. C'est à ce moment que la grosse unité vient se fixer. Donc, les deux unités ne s'assemblent que lorsque l'ARNm se fixe à la petite unité.

Deux ARNt peuvent se fixer par leur anticodon sur l'ARNr au niveau du ribosome (un sur la zone appelée site P et l'autre sur la zone appelée site A).

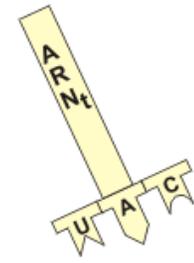
Liaison codon-anticodon de deux ARNt (il y a deux sites de liaison sur le ribosome).



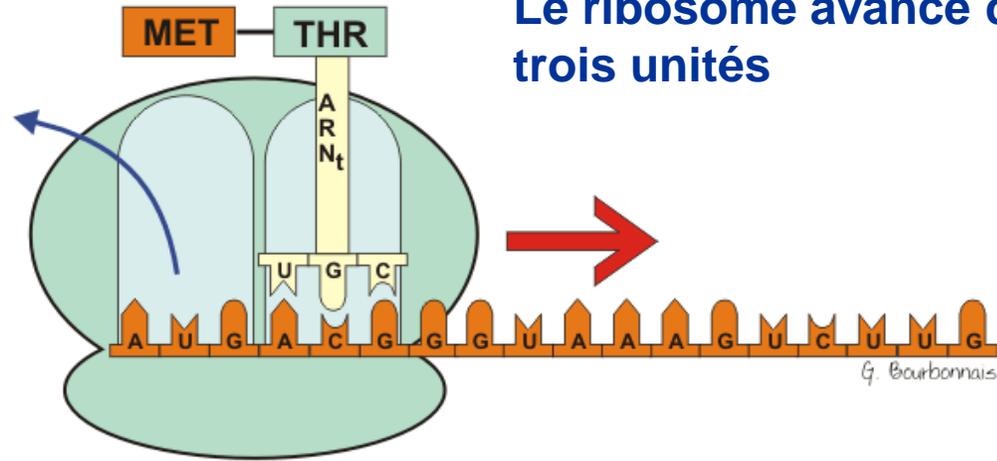
L'ARNt à l'anticodon UAC se fixe sur le codon AUG.
L'anticodon UGC se fixe sur le codon ACG

Chaque ARNt se fixe **par son anticodon** sur trois nucléotides de l'ARNm. Ces trois nucléotides de l'ARNm constituent ce qu'on appelle un **codon**.

Après leur fixation, les acides aminés qu'ils transportent sont reliés entre eux (le catalyseur est constitué d'une partie de l'ARN du ribosome et non d'une enzyme protéique).

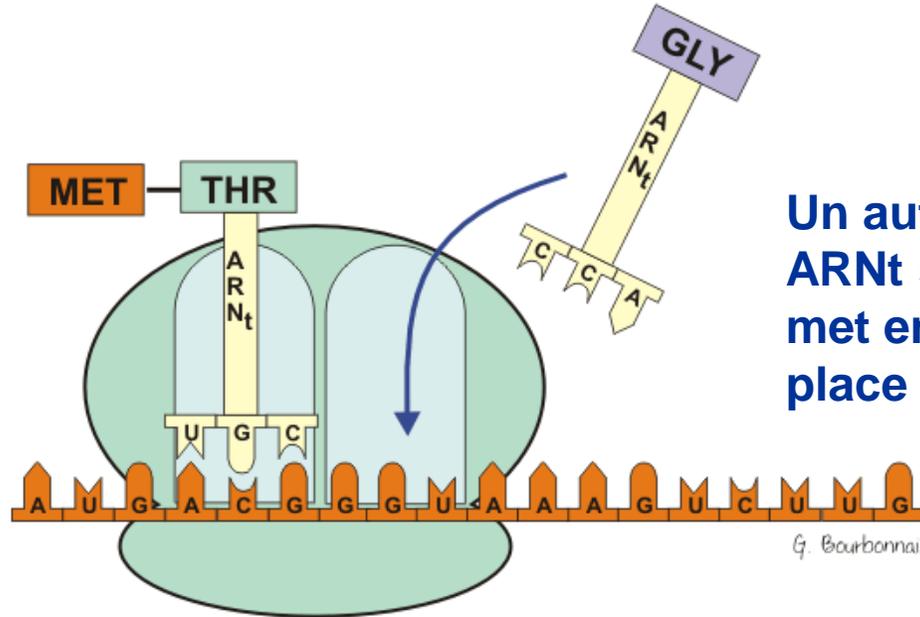


Le premier ARNt est retiré.



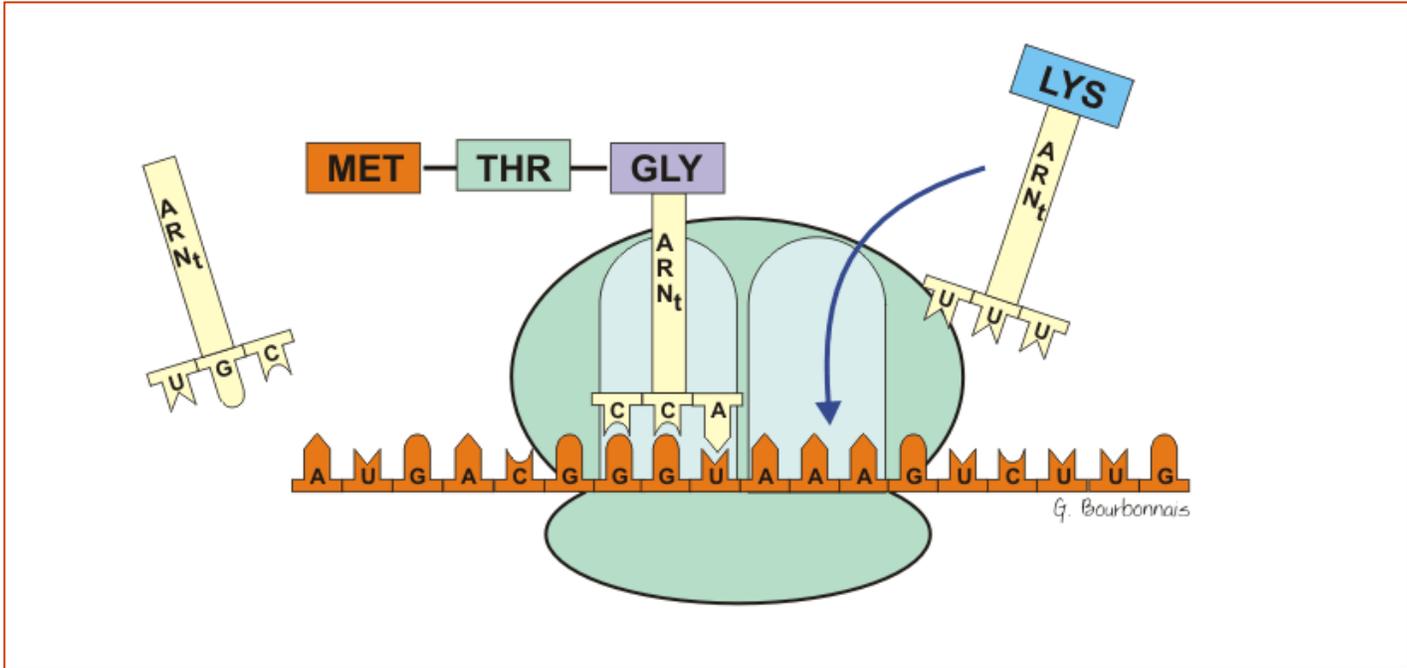
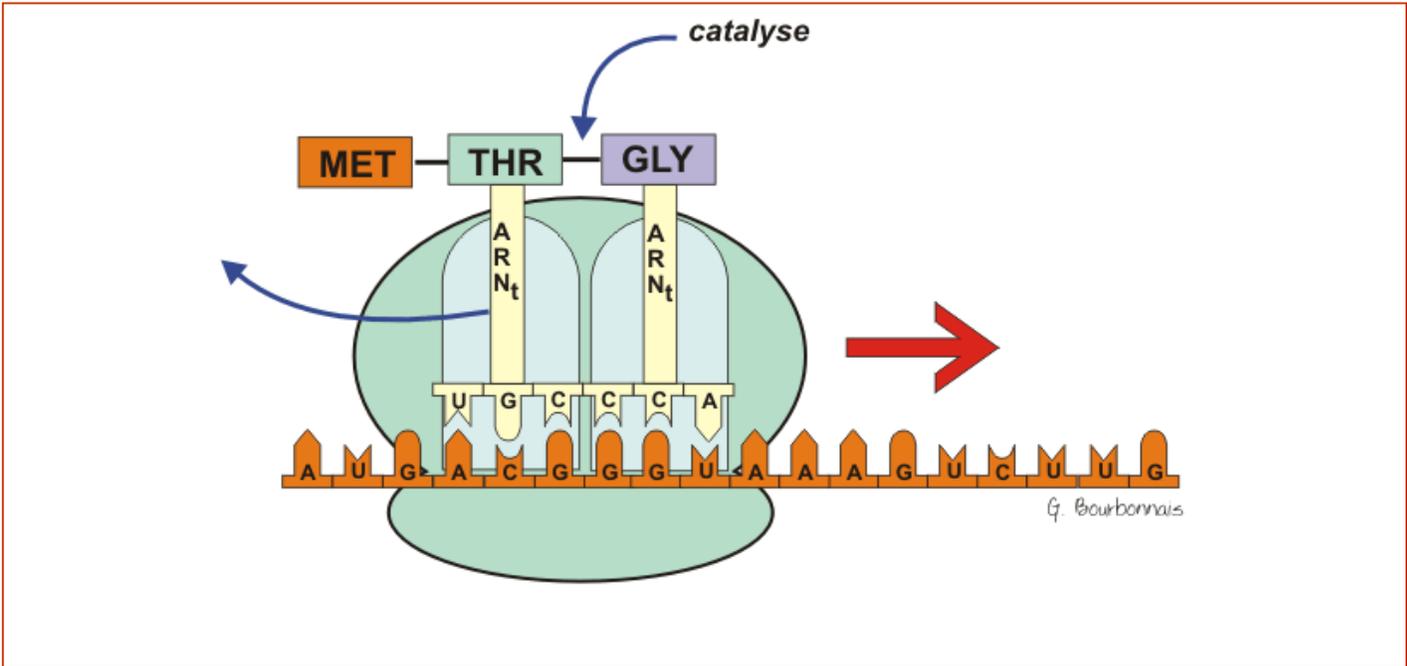
Le ribosome avance de trois unités

G. Bourbonnais



Un autre ARNt se met en place

G. Bourbonnais



- **Le code génétique**

- Ce code établit la correspondance entre trois nucléotides successifs appelés triplet ou codon de l'ARN messager et un des vingt acides aminés entrant dans la composition des protéines. Le code génétique est universel : il est le même chez tous les êtres vivants. Un codon correspond à un seul acide aminé. Mais le code génétique est dit dégénéré car il est redondant : un acide aminé peut être codé par plusieurs codons.

Caractéristiques du code génétique:

- 1. Les codons sont des triplets de nucléotides** et ils codent pour un acide aminé.
- 2. La séquence du gène et la séquence de la protéine codée sont colinéaires**, c'est-à-dire que la longueur du gène et la longueur de la structure primaire de la protéine finale sont proportionnelles.
- 3. Le code génétique est universel.** En effet chaque acide aminé dispose d'un ou plusieurs codons et ceci au niveau d'une multitude d'organismes vivants procaryote et eucaryote.
- 4. Le code génétique est redondant (ou dégénéré).** Plusieurs codons codent pour un même acide-aminé : on trouve 64 codons et 20 acides aminés. Souvent se sont les deux premiers nucléotides du codon qui définissent l'acide aminé, la redondance est donc due au troisième nucléotide du codon.

5. Le code génétique est non-chevauchant. Les nucléotides d'un codon ne participent qu'au code d'un seul acide aminé, ainsi le prochain acide-aminé sera codé par le prochain codon présent sur l'ARNm. On parle du **cadre de lecture** (ou **reading frame**).

6. Le code possède un système de ponctuation. Le codon d'initiation est le codon AUG (GUG pour ma mitochondrie) et les codons de terminaison sont les codons UAA (ocre), UAG (ambre) et UGA (opale). Le codon UGA (opale) n'est pas présent au niveau de la mitochondrie.

Code = Redondant, univoque et **UNIVERSEL**. **c'est le même pour tous les êtres vivants** sauf quelques rares exceptions:

PROTENOMIQUE

Pourquoi étudier les protéines

- Les protéines sont plus proches du phénotype final que l'ADN ou l'ARN
- Une grande partie de la régulation se fait au niveau des protéines
- Une grande partie des connaissances sur les maladies humaines provient de l'étude des protéines
- Les protéines sont des cibles directes des médicaments

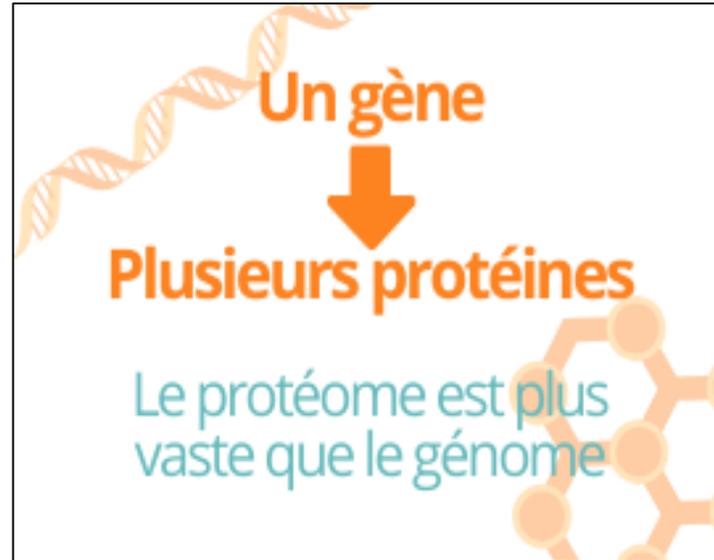
Définition

- ❑ La protéomique consiste à **étudier l'ensemble des protéines** d'un organisme, d'un fluide biologique, d'un tissu, d'une cellule ou même d'un compartiment cellulaire. Cet ensemble de protéines est nommé "**protéome**«
- ❑ Le protéome est **une entité dynamique et complexe**. Au sein de chaque cellule, le contenu de protéines se **modifie en permanence** en fonction des conditions intra ou extra cellulaires. De plus, par le biais de réarrangements, **un même gène peut donner naissance à plusieurs protéines**. Le protéome contient **donc un nombre beaucoup plus important de protéines que le génome ne contient de gènes**.

Définition

Etudier l'ensemble des protéines d'un échantillon biologique :

le protéome



- diagnostic
- pronostic
- traitement



La protéomique participe à la mise en place de la **médecine de précision**

- La protéomique, c'est l'histoire de la chenille et du papillon. Ces deux organismes apparemment si différents ont exactement le même génome. Ce qui les distingue, ce sont les produits finaux d'expression de leurs gènes, c'est à dire leurs protéines. Cet exemple montre à quel point il est nécessaire, pour comprendre un organisme, de s'intéresser à ses protéines et pas seulement à son génome.

Définition

- La protéomique est l'analyse à grande échelle des protéines, de leurs domaines et/ou d'autres caractéristiques.
- Les domaines concernés sont les suivants :
 - Profilage des protéines-
 - Modification des protéines-
 - Interactions protéine-protéine
 - Protéomique structurelle

Les principaux objectifs de la protéomique

Les principaux objectifs de la protéomique sont :

- **d'identifier, de quantifier et caractériser finement les protéines présentes** dans un échantillon biologique à un instant T
- d'obtenir des données fonctionnelles : localisation, modifications, identification de protéines partenaires, sites de liaison de ligands...

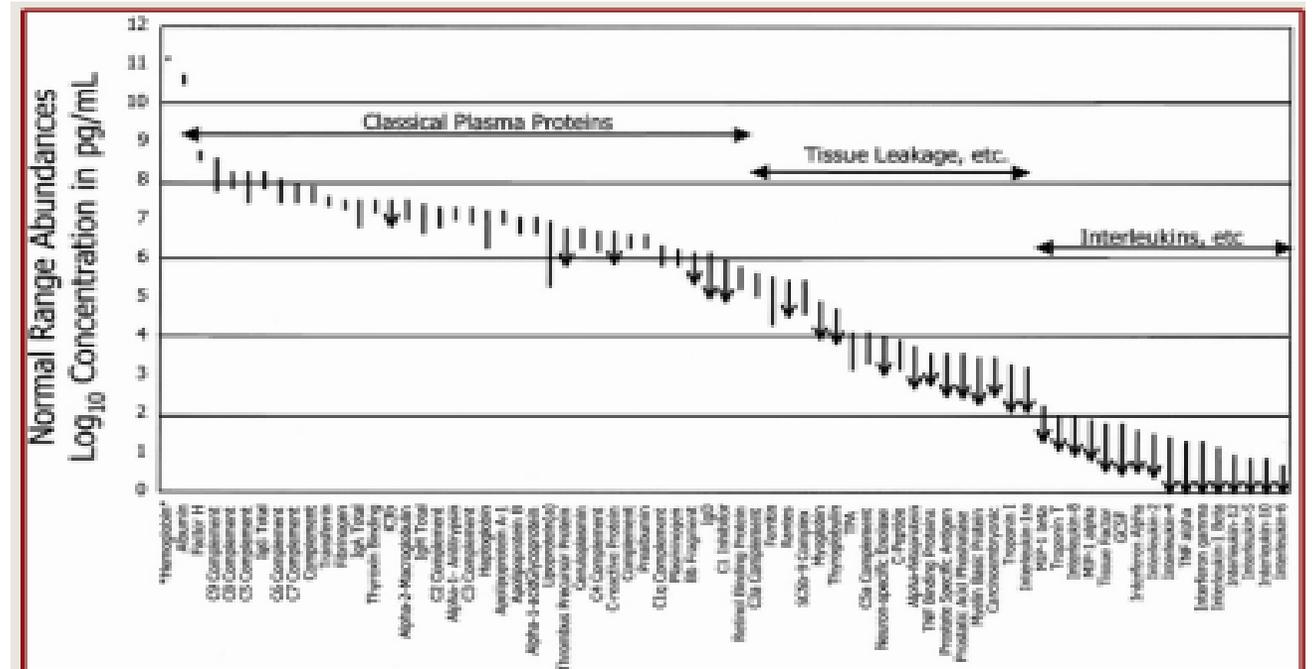
Ces données permettent de **mieux comprendre les mécanismes moléculaires impliqués dans les grandes fonctions cellulaires** et la physiologie des organismes vivants.

Human Proteome Project (HPP)

- Un vaste projet international de protéomique, lancé en 2011 par l'organisation mondiale HUPO (*Human Proteome Organisation*).
- Ce projet implique environ 50 équipes de recherche réparties dans 25 pays et consiste à établir une base de données permettant de décrire les protéines correspondant aux **19 800 gènes prédits** comme codants chez l'Homme.
- En 2019, 11% de ces protéines "prédites" n'ont pas encore été identifiées : leur existence reste donc toujours à confirmer. On les appelle les **protéines "manquantes"**.
- L'accomplissement de ce projet devrait permettre d'améliorer la connaissance de la biologie humaine et fournir de nouvelles informations utiles pour améliorer la prise en charge médicale des patients (diagnostic, choix thérapeutique...).

Profilage ou caractérisation des protéines

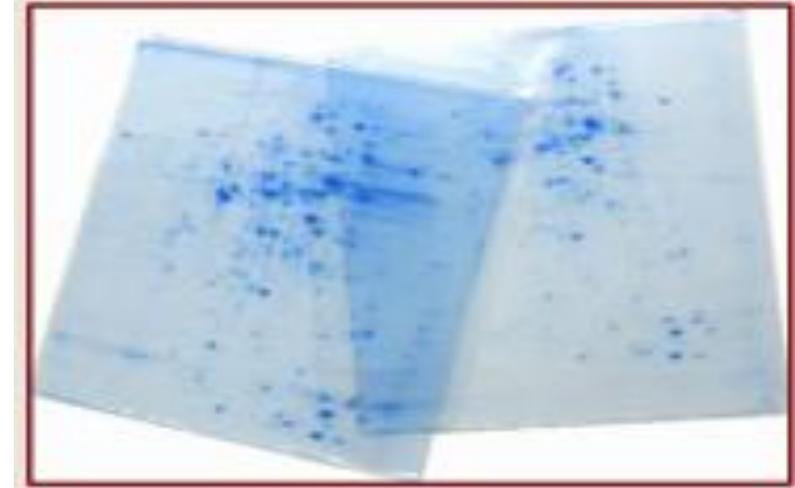
- Déterminer les niveaux de protéines
 - Dans des conditions différentes (santé vs maladie)
- Niveaux absolus et relatifs
- Défi :
 - Les protéines se trouvent à des niveaux très différents dans les cellules et les tissus (plus de 10 ordres de grandeur)



Le profilage des protéines

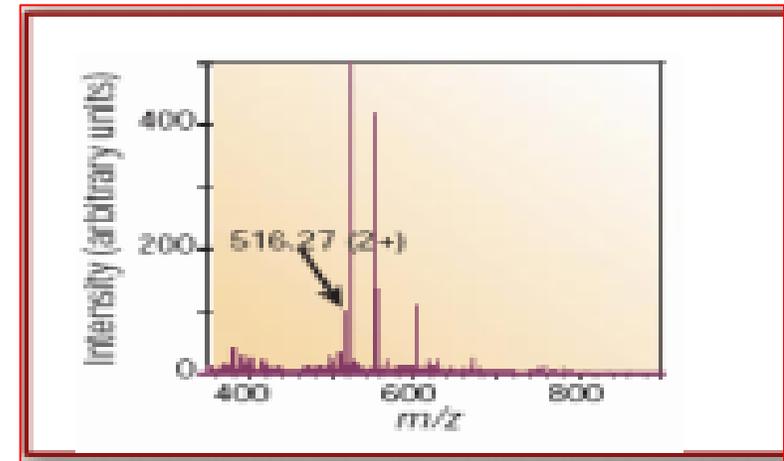
Traditionnellement: 2-Dimensional gels

- ✓ ~5000 spots;
- ✓ difficultés de migration;
- ✓ obtenir des spots qui doivent être identifiés



Actuellement : Spectrométrie de masse

- ✓ Très sensible
- ✓ Détecter les **atamoles**
- ✓ Très haute résolution



La spectrométrie de masse

Définition

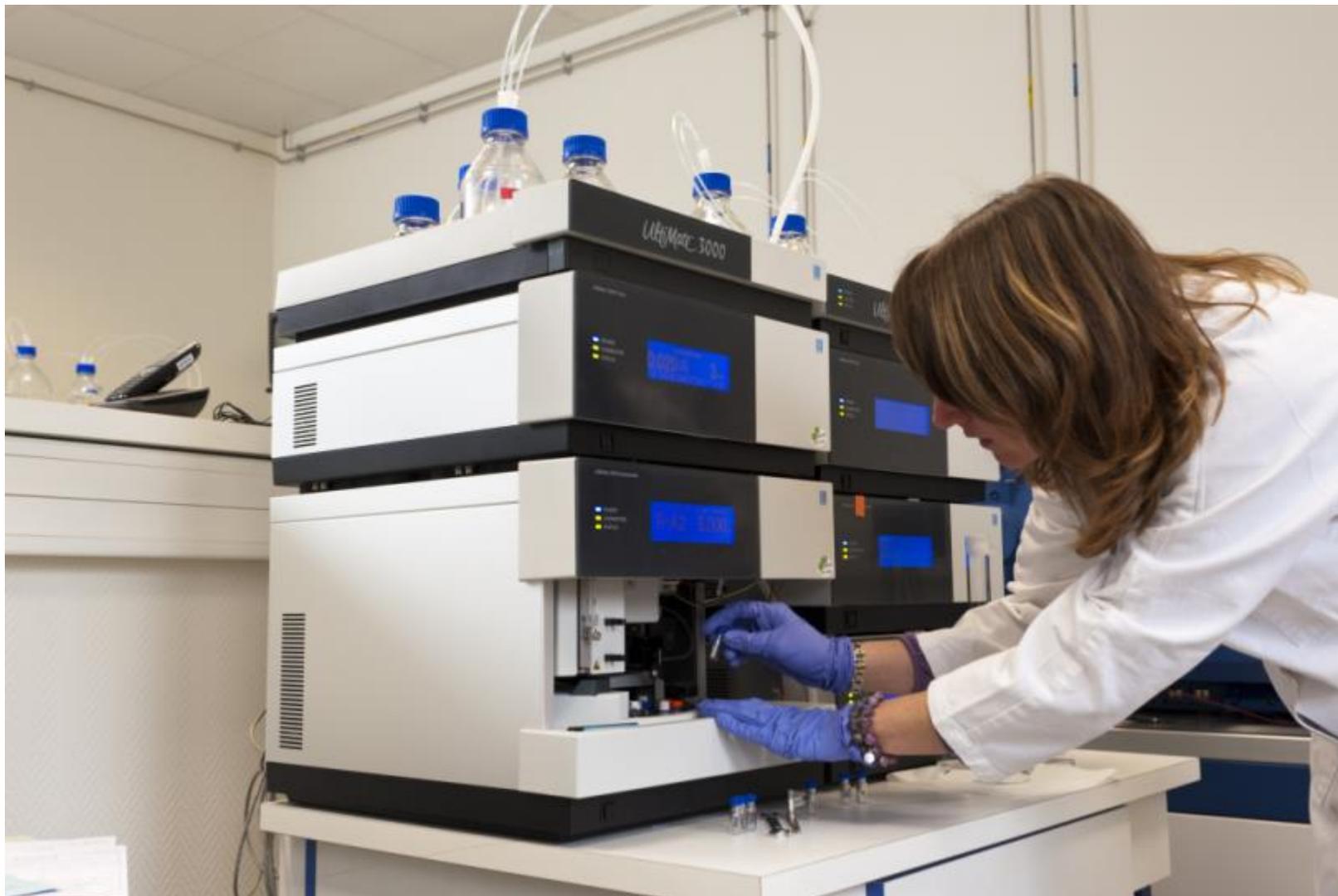
- La spectrométrie de masse consiste à **identifier des molécules en fonction de la mesure précise de leur masse.**
- Prix Nobel de chimie, attribué en 2002 à John Fenn et Koichi Tanaka
- Analyse des échantillons biologiques extrêmement complexes, pouvant contenir des milliers de protéines comme les fluides biologiques : plasma, urine.
- Caractériser des échantillons biologiques restreints et précieux correspondant à quelques cellules, voire une seule cellule, et contenant des quantités infimes de protéines.

La spectrométrie de masse

Principes

Pour réaliser une étude protéomique, il faut:

1. Digérer les protéines de l'échantillon à étudier grâce à une enzyme, afin d'obtenir des fragments protéiques (ou "peptides") qui sont solubles dans la solution qui est injectée dans le spectromètre de masse.
2. Fragmentation des peptides par la machine.
3. Mesure des masses de chaque peptide et des fragments.
4. Permettent d'identifier les peptides contenus dans l'échantillon, en comparant les données expérimentales aux données déjà existantes dans des banques.
5. Grâce aux performances des spectromètres de masse les plus récents, il est également possible d'étudier des protéines entières, sans avoir à les digérer préalablement.



Analyse de protéines par chromatographie Nano LC et spectromètre de masse Q-Trap au laboratoire Biologie à grande échelle (unité Inserm 1038, Grenoble) © Inserm/Patrice Latron

LTQ Orbitrap –performance specifications



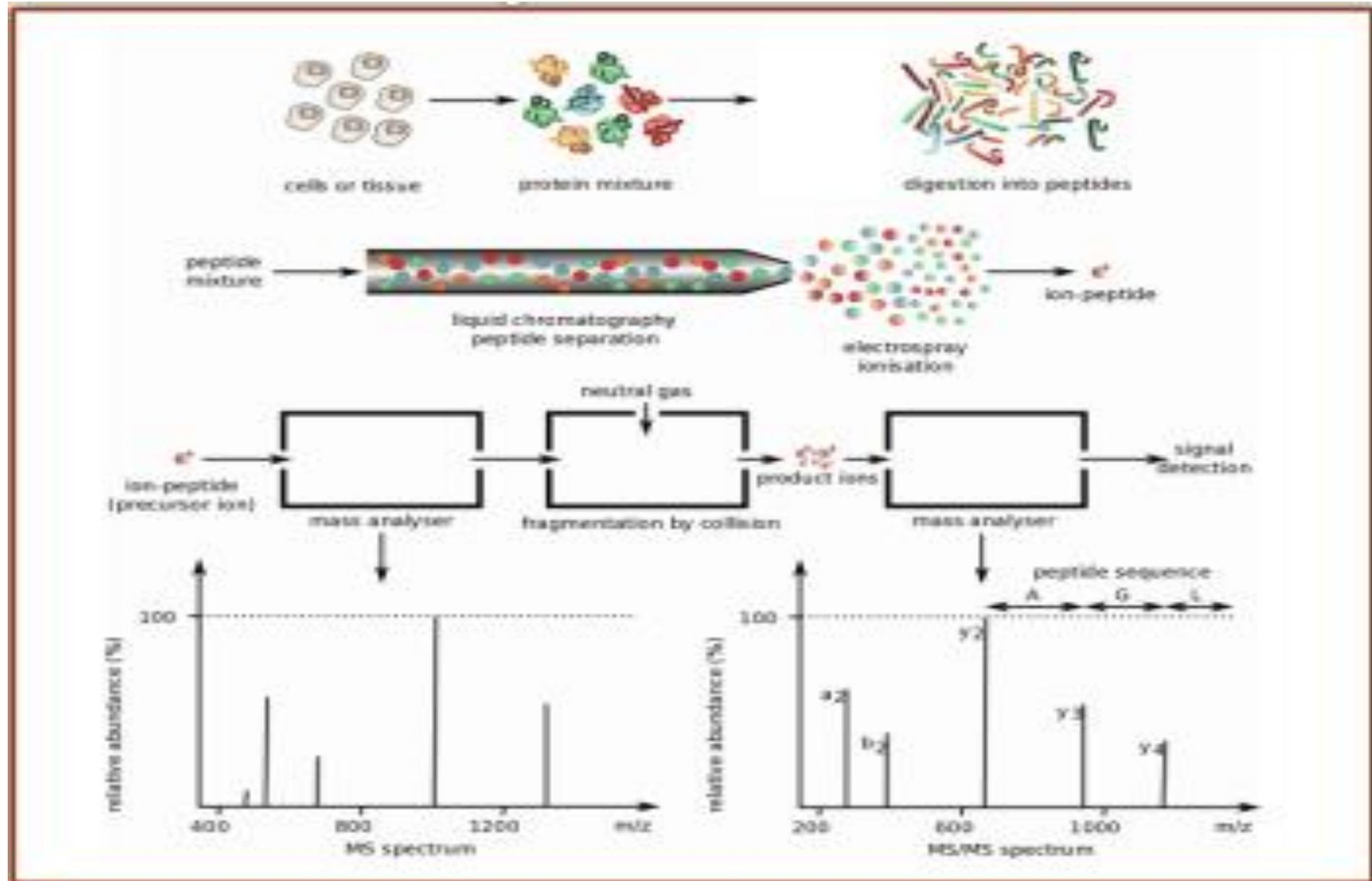
Une masse robuste et précise
sub ppm

Haute résolution
60 000 à m/z 400 avec 1 balayage par seconde
Résolution maximale >100 000

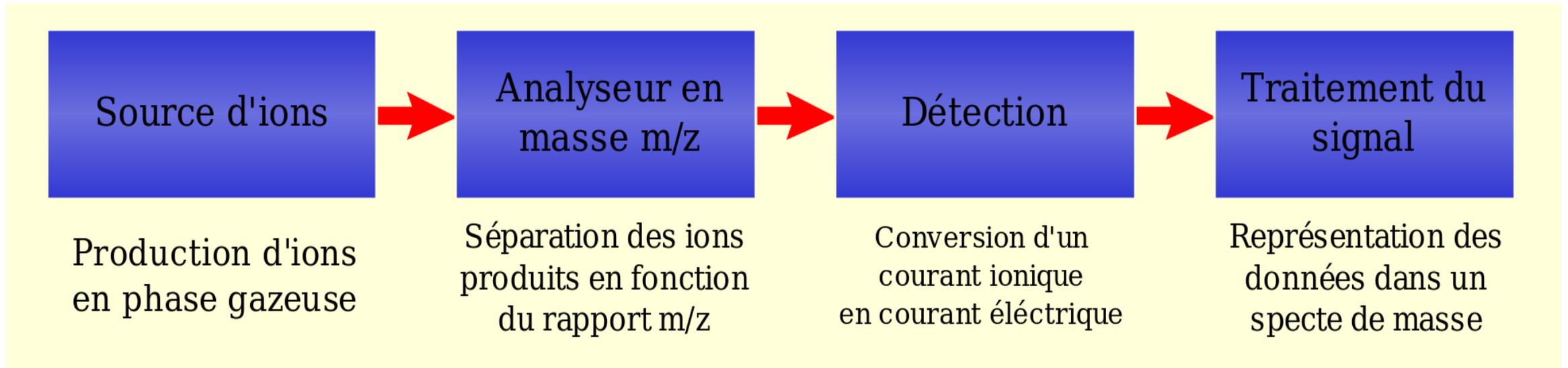
Gamme de masse
50-2000 ; 200-4000

Gamme dynamique élevée
>5000 dans un spectre
Mode de collision multiple CID, PQD, HCD, ETD

Séparation des protéines par spectrométrie de masse LC



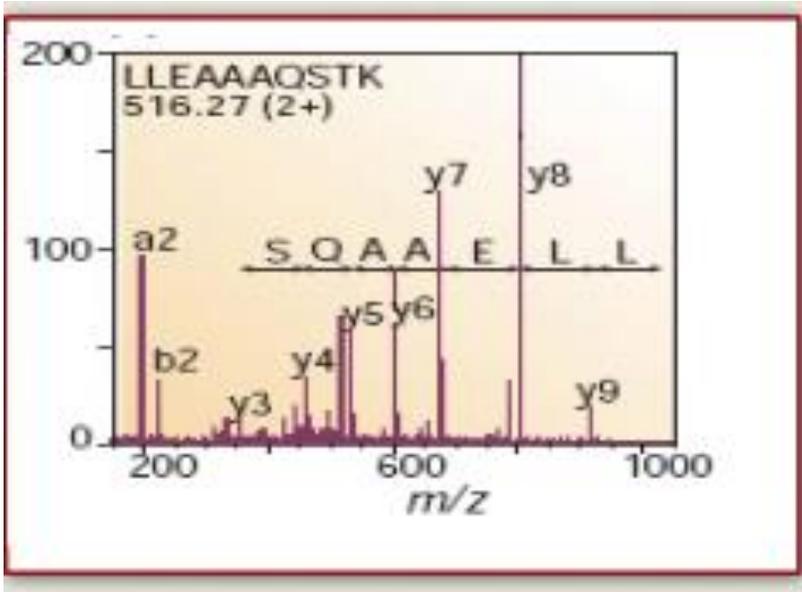
Séparation des protéines par spectrométrie de masse LC



Séparation des protéines par spectrométrie de masse LC

- Les données sont restituées sous une forme que l'on peut comparer à un puzzle. C'est aux scientifiques de **reconstituer ce puzzle pour retrouver l'identité des protéines qui étaient présentes dans l'échantillon.**
- Ce travail est bien sûr facilité par des logiciels informatiques de plus en plus performants et des bases de données de plus en plus riches.

Identification des Protéines



peptides 1
peptides 2
peptides 3

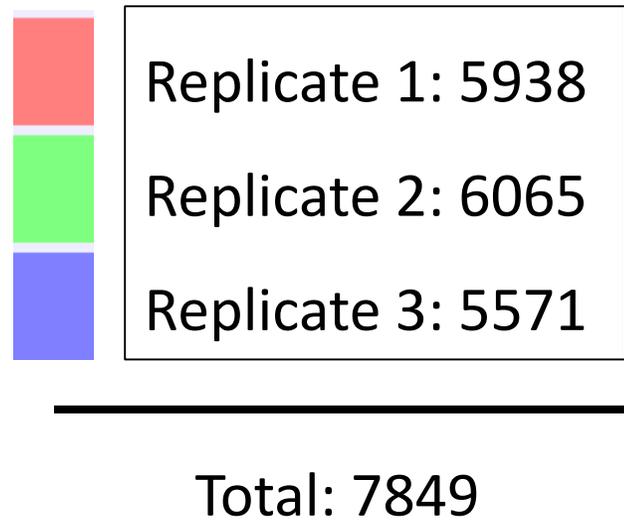


Protéine 1
Protéine 2
Protéine 3

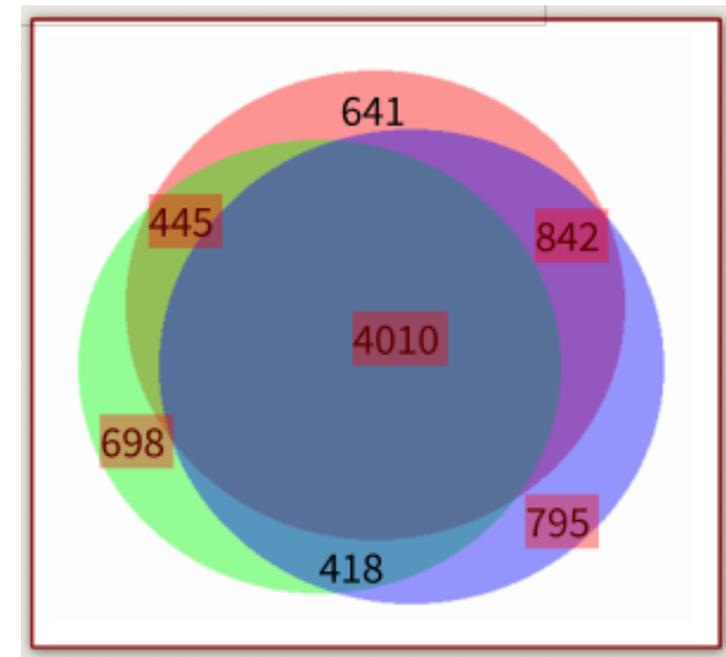
Faire correspondre chaque spectre à la base de données des spectres de peptides prévus sur la base de la séquence du génome

Utiliser toutes les informations sur les peptides pour déduire la liste finale des protéines

Identification totale de la protéine à partir de trois répliquats avec au moins un peptide unique



Pour chaque réplique : le taux de fausses découvertes est de 1% minimum 1 peptide unique atteint 98% de la protéine sont quantifiés



PBMCs/M. Snyder

Méthodes de préparation des échantillons : Très important

Destruction cellulaire :

Mécanique, ultrasons, pression, congélation-décongélation, osmotique et détergent

Solubilisation des protéines :

Chaotropes: 8M à base d'urée

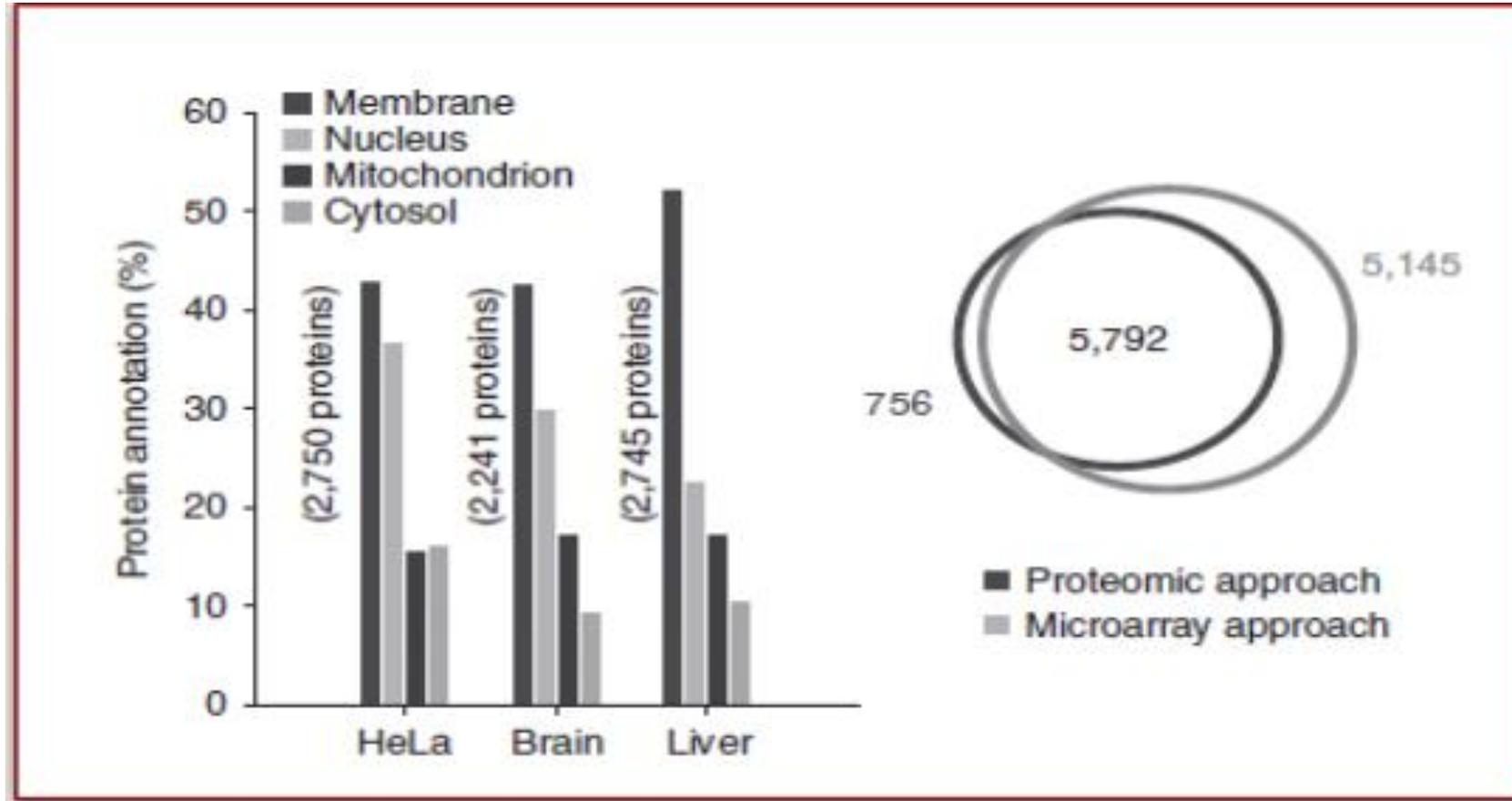
détergents : 4 % SDS

Méthodes d'enrichissement en protéines :

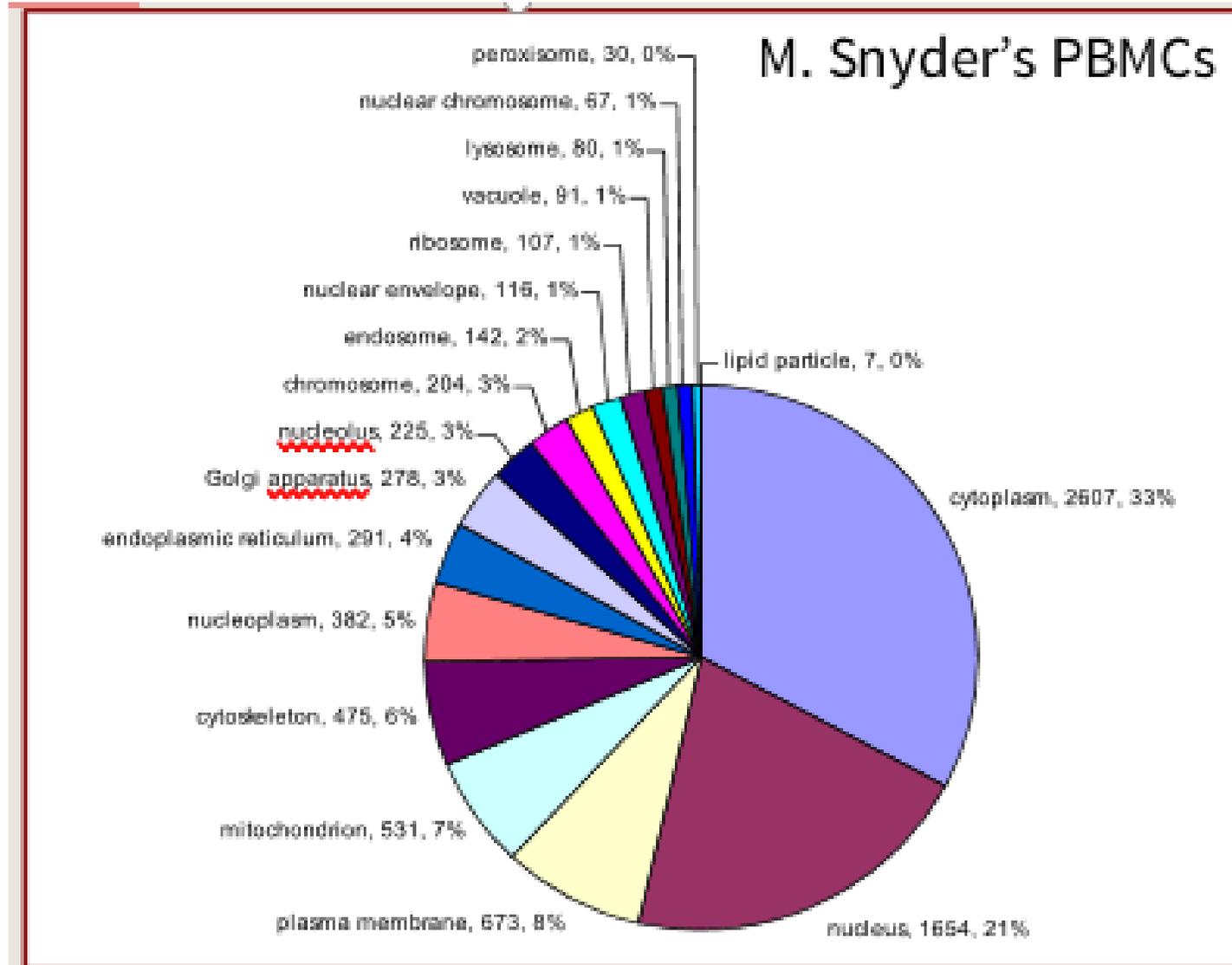
Précipitation de l'acétone

Précipitation du TCA Centrifugation

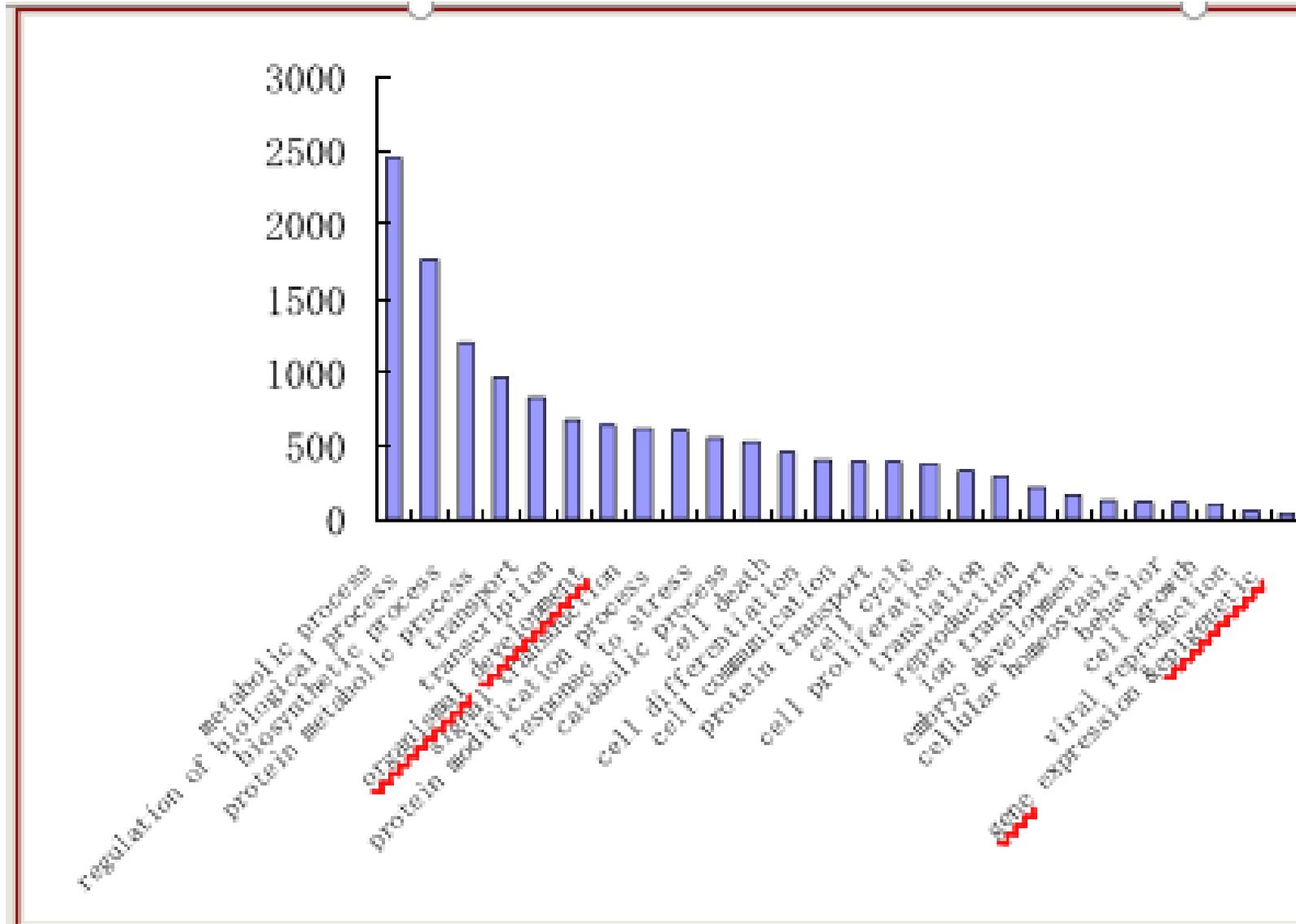
Annotation : Comparaison des résultats de gènes en ontologie



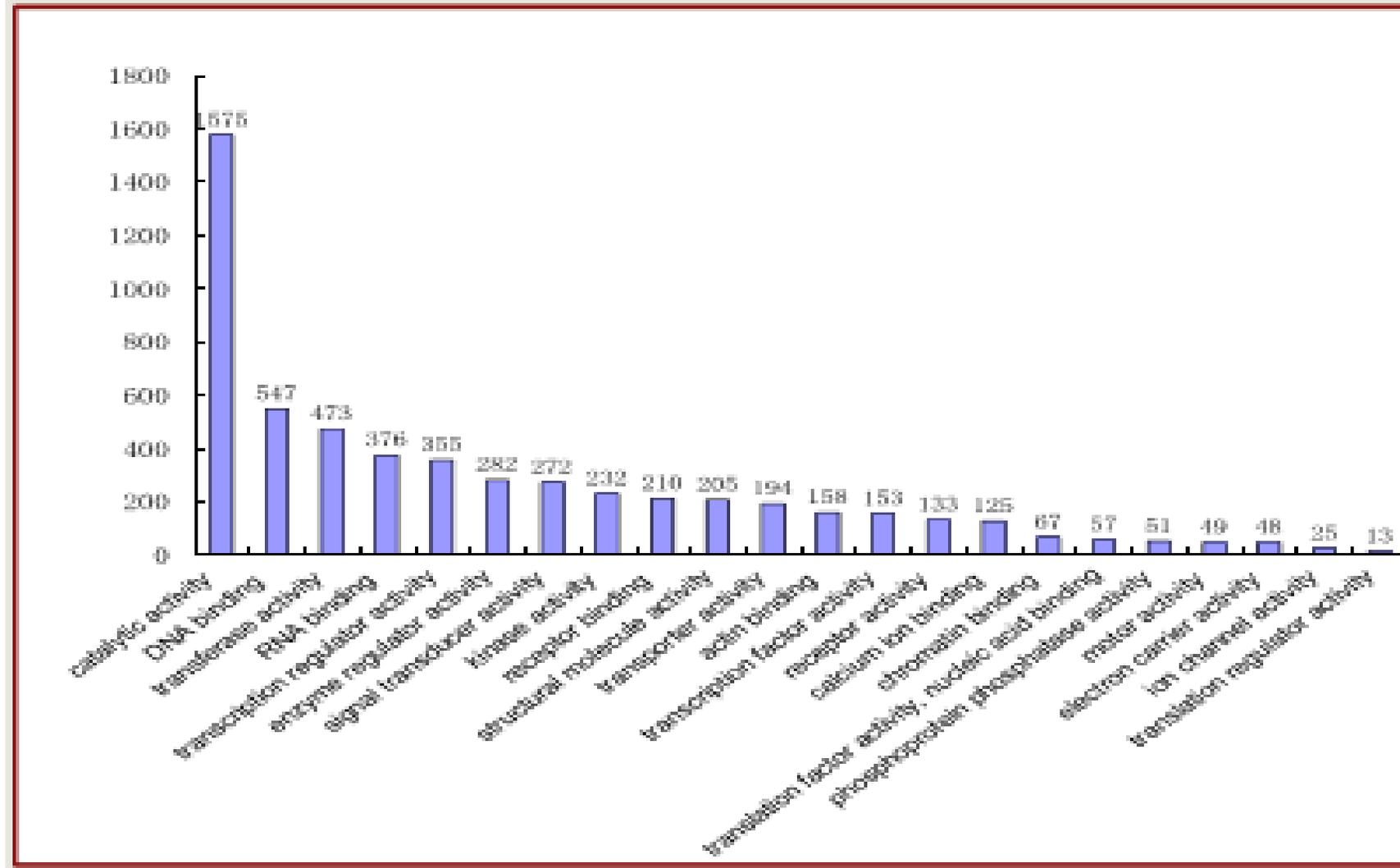
Nombre de protéines identifiées à partir de différents compartiments cellulaires



Nombre d'ID de protéines dans différents processus biologiques



Nombre d'ID de protéines ayant une fonction différente



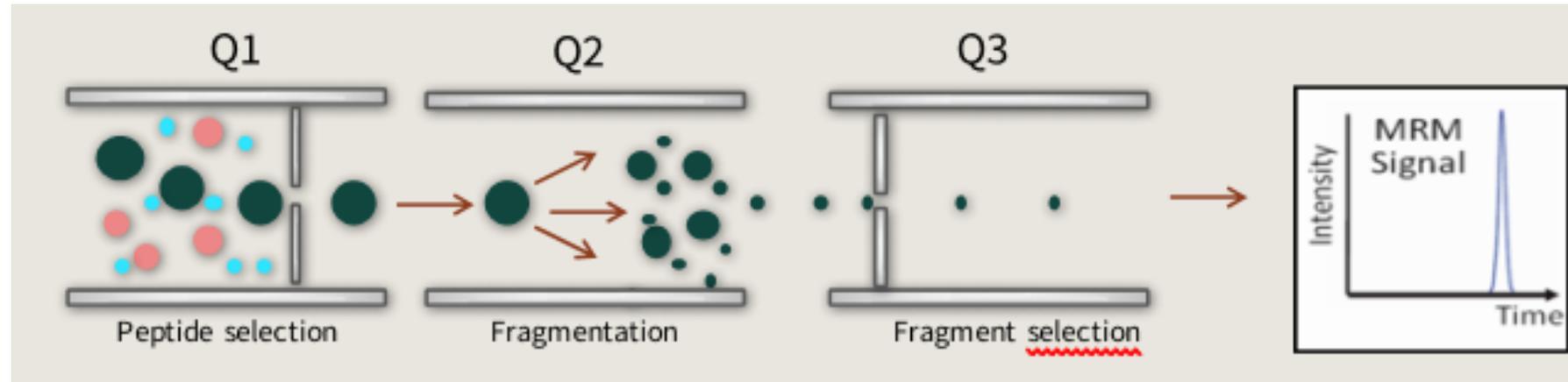
Approches Protéomiques

Protéomique non ciblée ou basée sur la découverte :

Séparer les peptides d'un mélange aléatoire de peptides ; sélectionner les pics et les fragmenter

Protéomique ciblée : Régler le spectromètre de masse pour rechercher directement les peptides d'intérêt

Spectrométrie de masse ciblée à l'aide d'un instrument triple quadrupole



Triple quadrupole instrument

Une gamme dynamique rapide et élevée

Sélection des fragments

Peut multiplexer : surveillance des réactions multiples (MRM)

Ont normalement une norme (voir module suivant)

- **La protéomique basée sur des découvertes non ciblées**

Pour

- ✓ Modérément rapide
- ✓ Gamme dynamique moyenne Non biaisée

Contre

- ✓ L'instrument doit avoir la capacité maximale pour traiter un échantillon non fractionné
- ✓ Toute étape de préparation d'un échantillon ajoute un bruit important

- **La protéomique ciblée**

Pour

- ✓ Rapide
- ✓ Plage dynamique très élevée
Donnée par l'hypothèse

Contre

- ✓ Impossible d'identifier les différences sur les protéines
- ✓ Une sensibilité et une dynamique insuffisantes pour traiter un échantillon non fractionné Compromis ...
- ✓ perte de confiance qualitative pour la sensibilité et la précision quantitative

Messages à retenir – Vocabulaire

- Protéomique non ciblée - Profilage général des protéines pour identifier les protéines présentes dans un mélange
- Protéomique ciblée - Analyse directe de protéines ou d'isoformes spécifiques dans un mélange complexe

Messages à retenir – Concepts

- La spectrométrie de masse est une méthode puissante pour l'analyse de lots de protéines dans un mélange complexe.
- L'utilisation d'instruments à haute résolution permet d'établir un profil général des protéines ~4-6K dans un mélange complexe
- La protéomique ciblée peut permettre l'identification directe de protéines spécifiques dans un mélange complexe.

Protéomique quantitative et mesure des niveaux de protéines dans le plasma

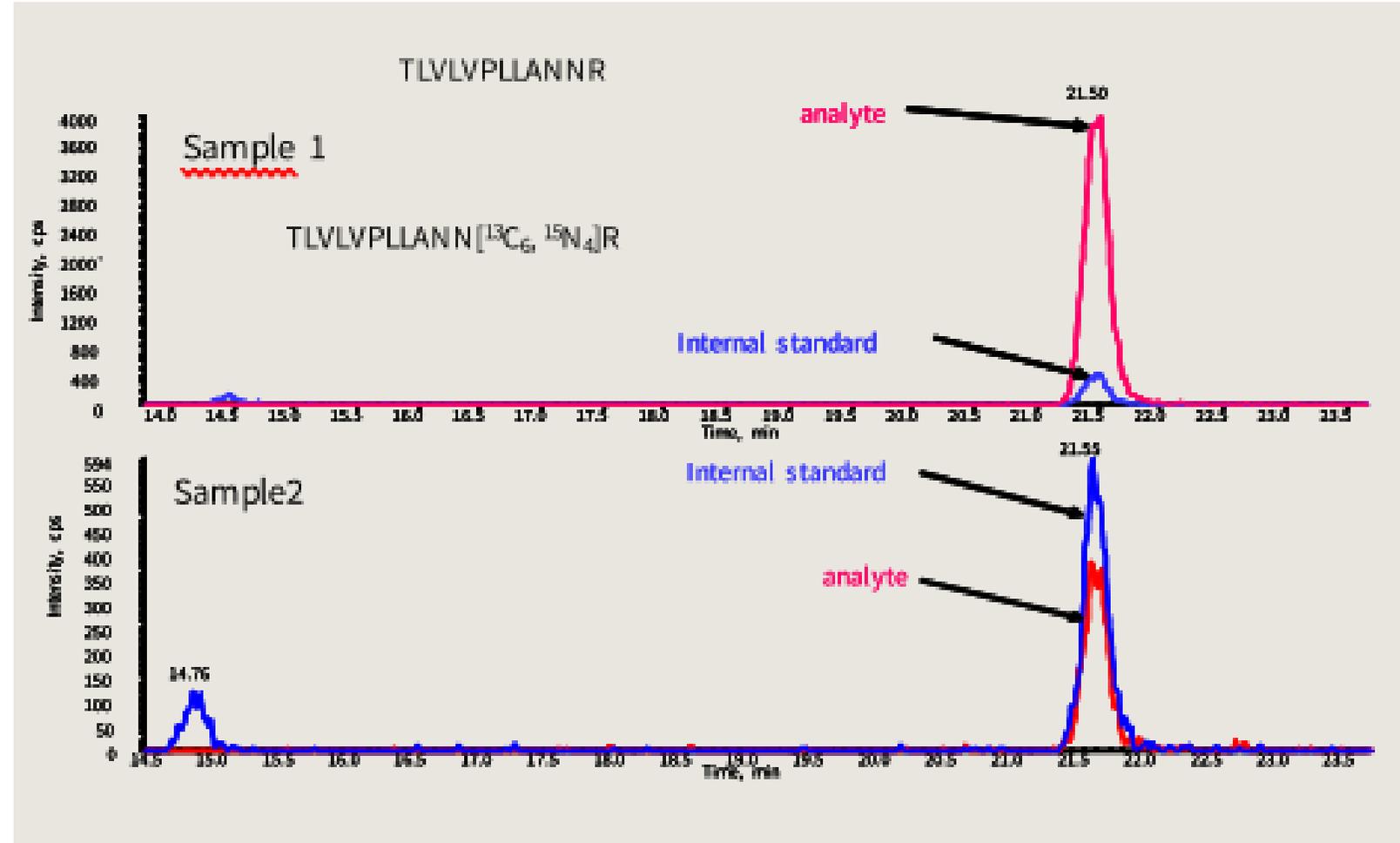
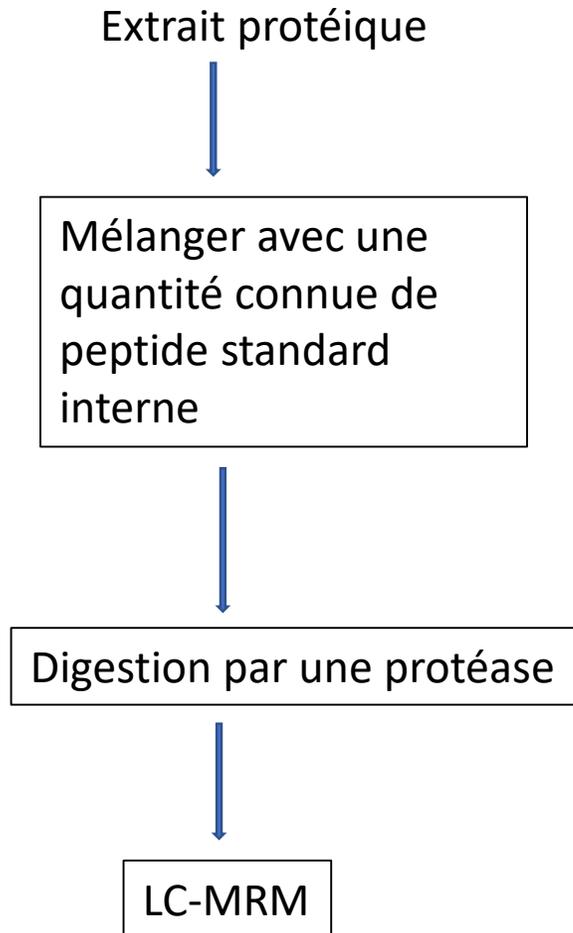
Objectifs

- Comment quantifier les niveaux de protéines
- Analyse des protéines dans le plasma

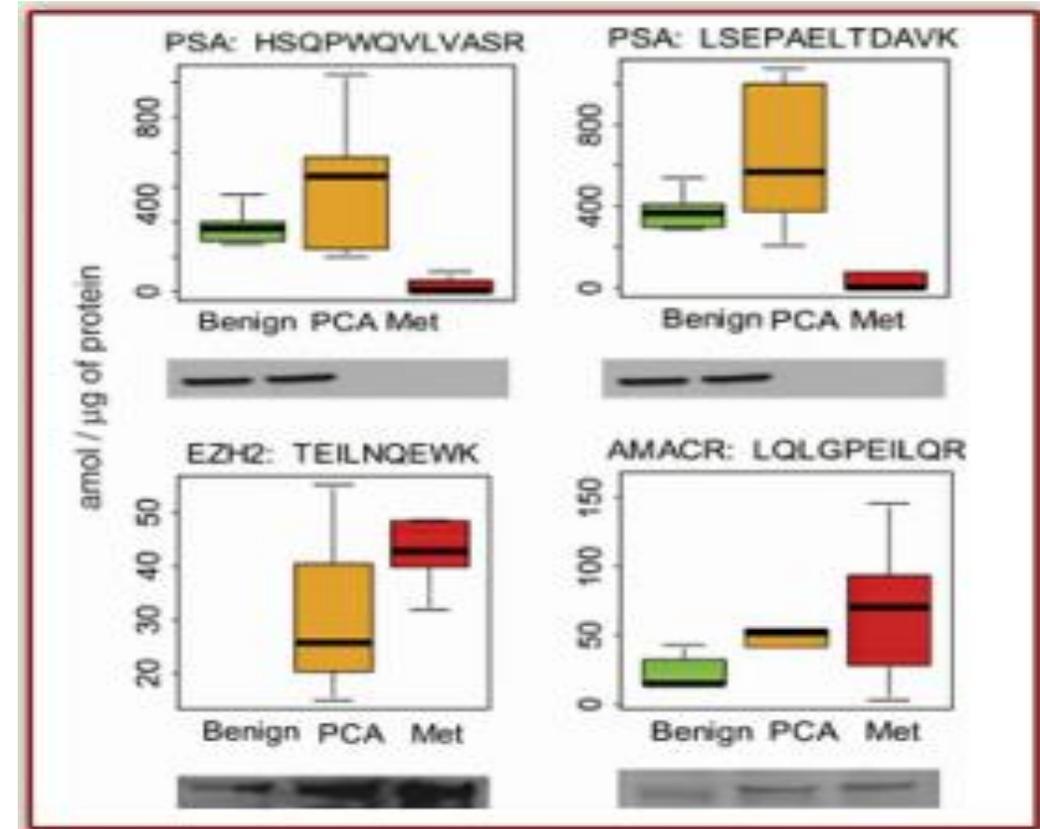
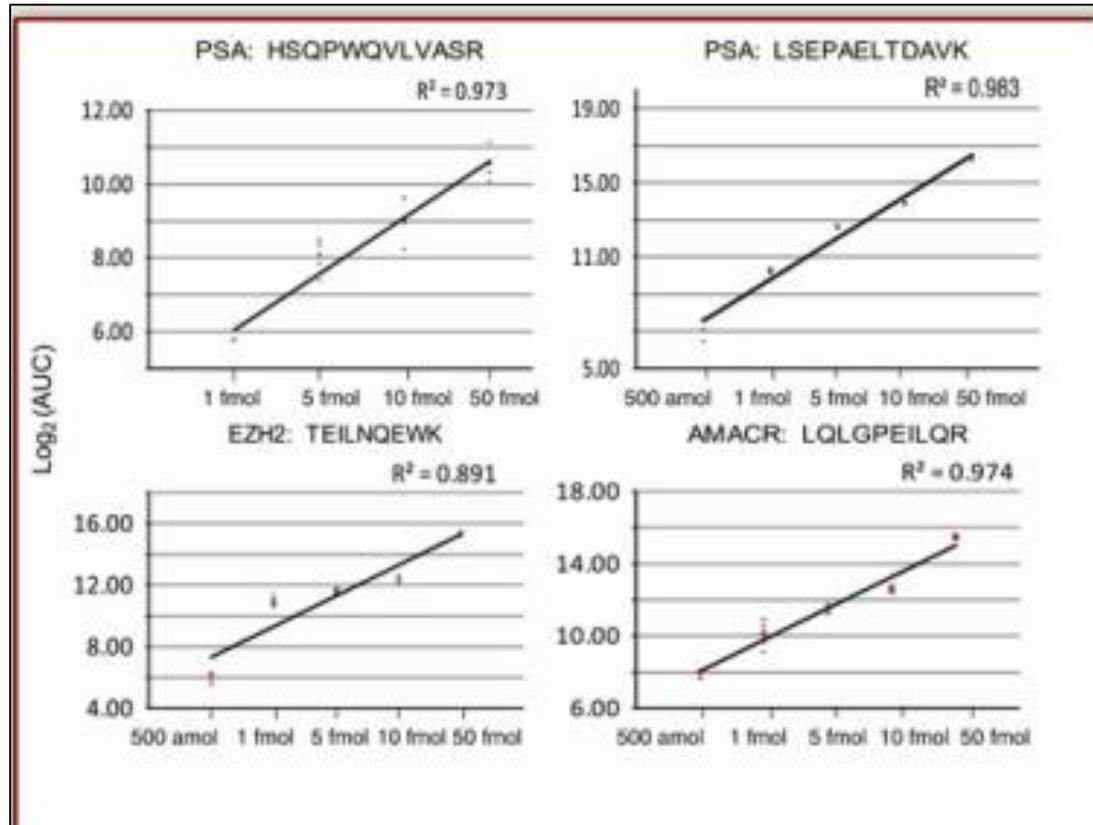
Protéomique quantitative : Plusieurs méthodes

- Absolue : Analyse avec peptides de référence
- Quantification relative:
 - Sans label
 - SILAC
 - Marquage isobare

Quantification absolue



Mesure des antigènes du cancer de la prostate dans les tumeurs de la prostate



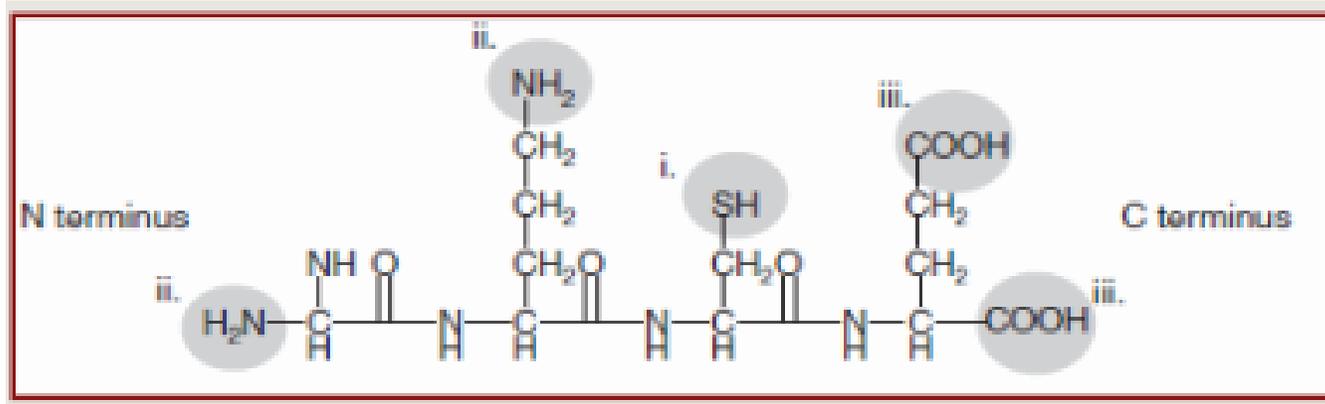
Yocum et al., Proteomics 2010 19:3506-14

Problème

- Besoin d'un peptide de référence (cher)
- La quantification relative est plus facile !

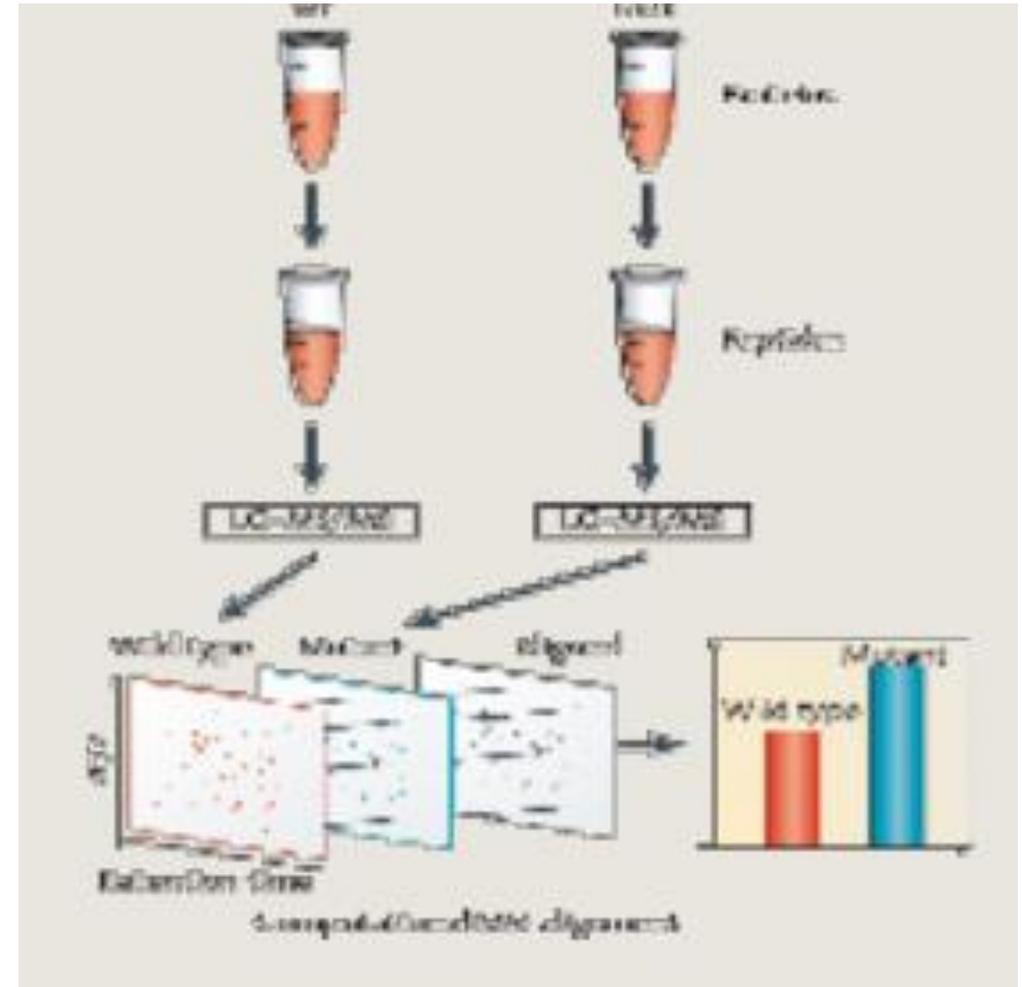
Méthodes de protéomique quantitative relative

- 1) Label Quantification libre
- 2) Étiquetage métabolique : SILAC (marquage des isotopes stables par les acides aminés en culture cellulaire)
- 3) Étiquetage des produits chimiques
 - i. Groupe sulfhydryle dirigé : ICAT
 - ii. Amine dirigé : TMT, ITRAQ
 - iii. Carboxyle dirigé : estérification du méthyle

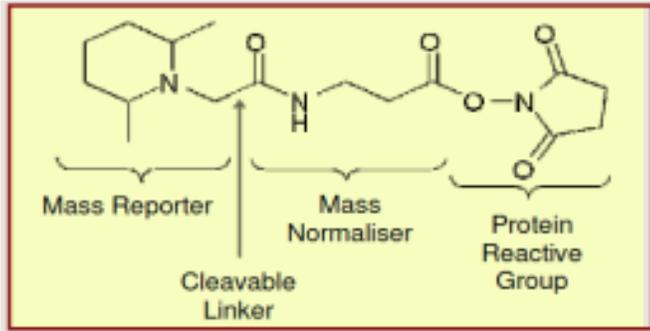


Quantification sans marquage

- Bon marché mais avec le plus d'erreurs de quantification
- S'appuie fortement sur les logiciels pour l'alignement des pics

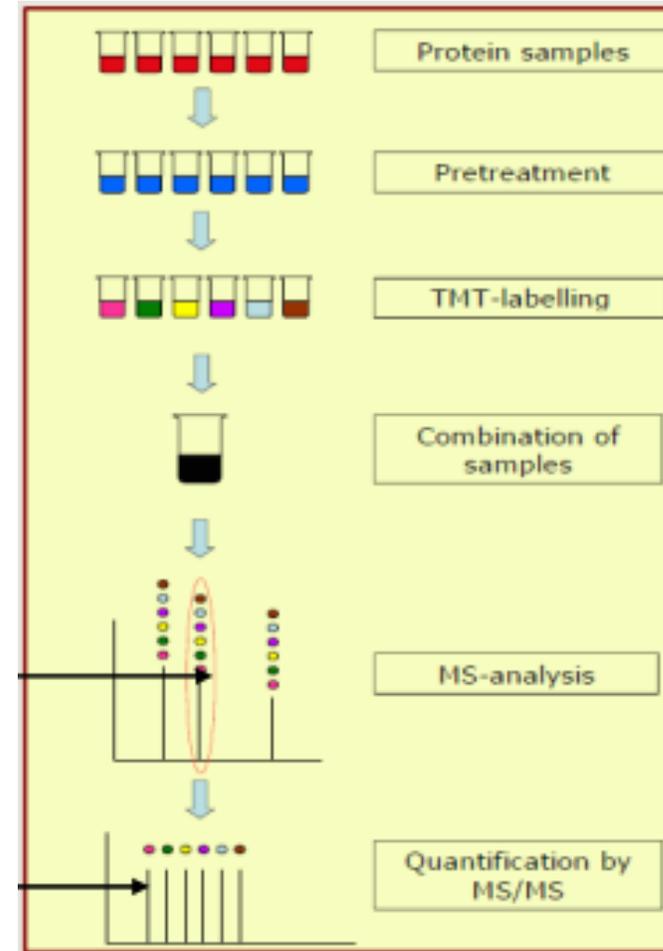


- Marquage chimiques (TMT 6plex)



Les peptides de différentes cellules montrent un pic

Les ions rapporteurs montrent une abondance différente



Avantages et inconvénients de l'étiquetage isobare des TMT

Avantages :

- 1) Les réactions avec le groupe amine primaire (N-terminal et lysines) peuvent être spécifiques et largement complètes.
- 2) Les peptides marqués de différents états sont isobares, de sorte que les spectres sont relativement simples.
- 3) Les informations de quantification ne sont révélées que dans les spectres de fragmentation.
- 4) Le multiplexage (jusqu'à 6 étiquettes isobariques) augmente le débit

Inconvénients

- Les peptides de co-élution de masse similaire génèrent les mêmes ions rapporteurs, ce qui complique la quantification

Messages à emporter - Vocabulaire

- SILAC - Marquage des isotopes stables par les acides aminés en culture cellulaire
- Étiquetage isobarique - Ajout de marqueurs de masse aux peptides qui permettent la quantification en MS2

Messages à emporter – Concepts

- La spectrométrie de masse est une méthode puissante pour la quantification des protéines
- L'utilisation de peptides de référence marqués aux isotopes permet la quantification absolue d'un nombre limité de protéines
- La quantification relative des niveaux de protéines peut être réalisée par :
 - Sans marquage (moins précis)
 - SILAC
 - Marquage TMT

- Dans l'avenir, un enjeu majeur est de **développer des outils mathématiques et informatiques permettant d'intégrer les données issues de ces approches protéomiques avec d'autres types de données** (imagerie, biochimie médicale, génomique, transcriptomique...) afin de générer des signatures moléculaires "composites" permettant de mieux caractériser les patients et les pathologies, avec pour objectif d'établir une médecine de précision.