

Analyse protéomique : Principes et exemple

Principes de l'analyse protéomique :

Partie I

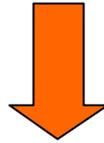
- 1 Généralités et rappels sur les protéines
- 2 Extraction des protéines
- 3 Séparation des protéines
- 4 Méthode de détection des protéines : E2D
- 5 Analyse Assistée par Ordinateur
- 6 Identification des protéines par spectrométrie de masse
- 7 Recherche dans les bases de données
- 8 Les domaines d'application
- 9 Conclusion

Généralités et rappels sur les protéines



Définition

PROTEINES



Deux étymologies :
-*prôtos* = premier
-*Protée* = change de forme à volonté

**Déf : polymère biologique
issu de la traduction du code génétique
porté par l'ADN**

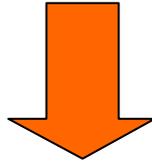
Cours CNAM, Rennes le 15/01/2009 Laurent Marché



Selon le Petit Robert le terme de protéine apparaît vers 1838. En grec, le terme *prôtos* veut dire premier, ce qui est à l'origine des choses. Par ailleurs, dans la mythologie grecque, une divinité de la mer se nomme Protée (*Prôteus*) et peut prendre des formes variées. La définition biologique est quant à elle plus rationnelle. Une protéine est un polymère d'acides aminées issu de la traduction, par la machinerie cellulaire, de la traduction du code génétique porté par l'ADN.

Définition

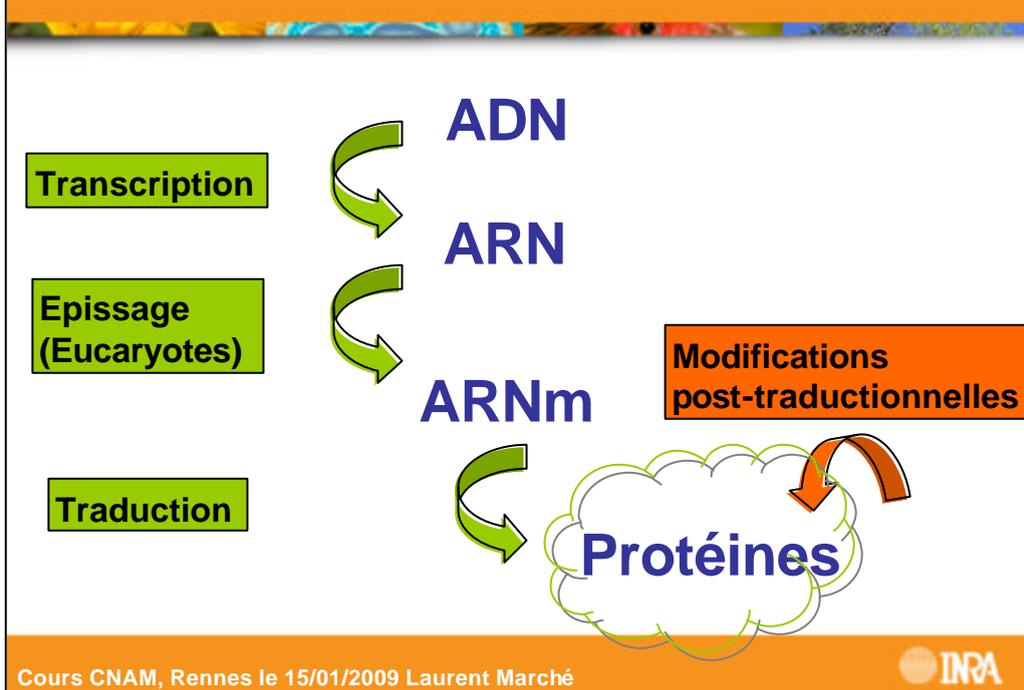
- Quantitatif : 50 % du poids sec
- Qualitatif : structure et fonctionnement



**Molécules indispensables
aux êtres vivants**

Les protéines représentent une forte proportion des constituants du vivant. On estime à environ 50 % du poids sec des êtres vivants. Du point de vue qualitatif, les protéines sont des molécules qui sont impliquées aussi bien dans la structure que dans le fonctionnement des organismes.

Origine : le code génétique



Cours CNAM, Rennes le 15/01/2009 Laurent Marché



Rappel :

Les protéines sont issues de la transcription du code génétique porté par l'ADN qui donne alors un ARN. Cet ARN, chez les eucaryotes subit l'épissage. Les séquences non codantes (introns) sont excisées et les séquences codantes (exons) sont réunies et forment alors l'ARNm messager. Chez les procaryotes l'ARN transcrit est directement traduit il n'y a pas d'épissage.

C'est l'ARNm qui porte l'information qui va être traduite en protéine. La traduction se fait par la reconnaissance de séquence dite : codon, composées de trois bases successives. Le code génétique fait correspondre à chaque codon un acide aminé. Cette traduction s'effectue par l'intermédiaire des ARNt dit de transfert, qui portent une séquence anticodon et un acide aminé. C'est la séquence anticodon qui reconnaît le codon. Les protéines seront formées par l'enchaînement des acides aminés suivant l'ordre de la séquence des codons.

Par la suite, les protéines pourront selon le cas subir des modifications post-traductionnelles, c'est l'étape de maturation des protéines. Ces modifications peuvent être de différents types : glycosylation, phosphorylation, mise en place de ponts di-sulfures par exemple.

Structure

- **Atomique : C, H, O, N, (S,P)**
- **Moléculaire : macromolécules constituées par assemblage d'acides-aminés (de 50 à plusieurs milliers).** (E. Fischer, 1852-1919)

Cours CNAM, Rennes le 15/01/2009 Laurent Marché

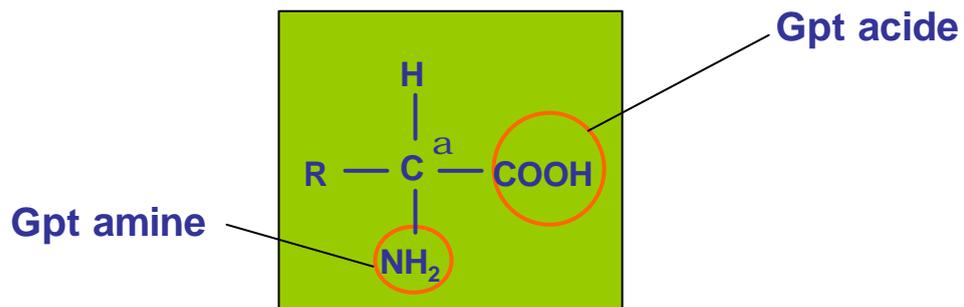


Au niveau atomique, les protéines sont toutes constituées de C, H, O, N, certaines contiennent aussi du S et du P. Les protéines peuvent contenir jusqu'à 16 % d'azote, à la différence des glucides et des lipides qui eux n'en contiennent pas.

Les protéines sont des macromolécules constituée par l'assemblage de sous unités que sont les acides-aminés (50 à plusieurs milliers).

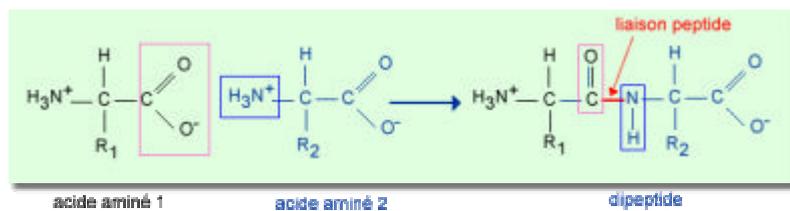
Structure

- Les acides-aminés :
- environ 20 dans le vivant



Il existe 20 acides aminés utilisés par l'ensemble des êtres vivants pour produire leurs protéines. Les acides aminés sont tous (à l'exception de la proline et de l'hydroxyproline hétérocyclique) construits à partir de la même base (cf-diapo). Les acides aminés portent une fonction acide et une fonction amine.

La liaison peptidique



Obtention de longues chaînes

Réf dessin : Jean-Luc Zimmermann
Éléments de chimie organique, 2^e édition, 2005

Lorsqu'ils sont liés entre eux dans un peptide, un polypeptide ou une protéine on parle alors de **résidu**. Ils ont alors perdu un H en NH₂ et OH en COOH. Dans toute la chaîne primaire les acides aminés sont reliés par une liaison peptidique.

Structure

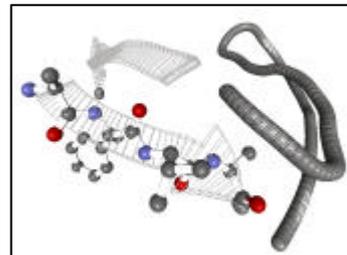
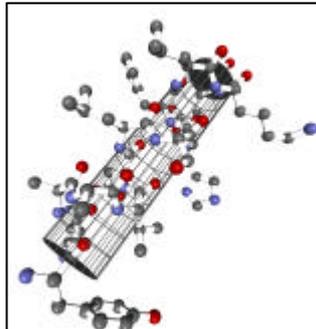
▪ Niveaux d'organisations

- **PRIMAIRE** enchaînement des acides-aminés



- **SECONDAIRE** deux types de repliements dans l'espace

- Feuillet β (Chaîne antiparallèle ou parallèle)
- Hélice α



Reference: Lee, C. H., Saksela, K., Mirza, U. A., Chait, B. T., Kurlyan, J.: Crystal structure of the conserved core of HIV-1 Nef complexed with a Src family SH3 domain. Cell 85 pp. 931 (1996)

Cours CNAM, Rennes le 15/01/2009 Laurent Marché



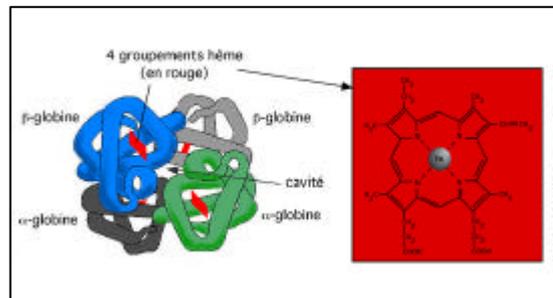
La structure primaire des protéines est celle constituée par l'ordre de succession des acides aminés.

Structure secondaire : Les structures secondaires des protéines sont de deux types. La structure en hélice alpha. Cette structure est issue de liaisons hydrogène intra-chaîne. La molécule est enroulée en hélice, il y a 3,7 résidus par tour de spire.

La seconde structure rencontrée est celle en feuillet beta : structure en zig-zag. On parle aussi de structure en chaîne anti-parallèle ou parallèle. Il y a mise en place de liaisons hydrogène situées dans les plans du feuillet plissé entre des groupements $-CO$ et $-NH$ de deux liaisons peptidiques.

Structure

- **TERTIAIRE** repliement de la chaîne polypeptidique par liaisons de type : pont-disulfure, hydrogène, ionique, hydrophobe
 - Acquisition des propriétés biologiques
- **QUATERNAIRE** agrégation de sous unités polypeptidiques entre elles.



Cours CNAM, Rennes le 15/01/2009 Laurent Marché



La structure tertiaire est acquise ensuite par la protéine. C'est cette structure qui va permettre l'acquisition des propriétés biologique de la protéine dans le cas ou elle en possède une. Les chaînes polypeptidiques ayant déjà leur structures secondaire vont se replier. Selon le type de résidus portés par les chaînes et les propriétés biochimiques de ces derniers, il aura mise en place de liaisons "interchaînes" de différents types : ioniques, hydrogène, pont-disulfure ou hydrophobes.

Pour certaines protéines, il peut aussi exister une structure quaternaire. Cette structure est en générale formée à partir de l'assemblage de sous-unités protéiques ayant déjà mises en place leur structure tertiaire. Pour certaines protéines il peut aussi y avoir assemblage avec un ou des éléments qui ne sont pas de nature protéique, on parle alors de groupement prosthétique. Le cas le plus connu est celui de l'atome de Fer au sein de l'hémoglobine.

Classification

2 grands groupes de protéines



▪ **Holoprotéines**



▪ **Hétéroprotéines**

Comme on vient de le voir les protéines peuvent être organisées avec d'autres types de molécules.

On parle d'holoprotéines lorsque les protéines sont uniquement composées d'acides-aminés et d'hétéroprotéines lorsqu'elles sont associées avec un groupement prosthétique de nature : glucidique, lipidique, nucléique, ou autre.

Familles/Fonction

- **Enzymes**
- **Structure**
- **Défense**
- **Régulation**
- **Transport**
- **Contractile (mouvement)**
- **Stockage**

Au fur et à mesure de la découverte des différentes protéines et de leurs fonctions, on a pu définir de grandes familles de protéines selon leur fonction biologique.

Les enzymes sont un groupe très large qui comprend toutes les protéines qui interviennent dans les réactions biochimiques des organismes. Elles sont impliquées dans les différentes réactions du métabolisme et du catabolisme. Les protéines de structure sont aussi parmi les plus nombreuses et vont participer à l'édification structurale des organes et des cellules. On peut citer les protéines membranaires, les kératines qui composent les phanères (ongles, sabots, cheveux, plumes...) des mammifères ou des oiseaux par exemple.

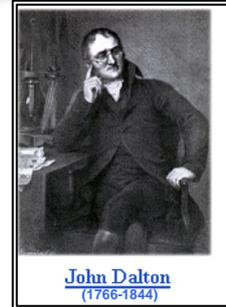
Le système de défense repose assez souvent dans le monde vivant sur des protéines. Chez certaines bactéries ce sont plutôt des petits peptides qui ont des propriétés anti-bactériennes, chez les insectes il existe un système glyco-protéique. Chez les animaux supérieurs, les anti-corps composés d'une chaîne lourde et légère sont des composés protéiques complexes ayant une capacité spécifique de reconnaissance des agents étrangers à l'organisme.

Dans les systèmes de régulation on note la présence de protéines qui sont alors classées parmi les hormones.

Au sein d'un organisme certaines protéines sont spécialisées dans le transport d'autres molécules. Elles peuvent transporter des glucides, des lipides, des acides nucléiques ou d'autres protéines. Le transport peut être effectué à distance ou simplement trans-membranaire pour permettre le

Le Dalton

- La masse molaire des protéines est exprimée en Daltons (Da) plutôt qu'en $\text{g}\cdot\text{mol}^{-1}$.



1Da = 1/12^{ème} de la masse d'un atome de ^{12}C



1,66 10^{-27} kg

L'unité légale de la quantité de matière est la mole (symbole: mol). Une mole représente la quantité de matière d'un système contenant autant d'entités élémentaires qu'il y a d'atomes dans 12 g de carbone 12. (Ce nombre est appelé nombre d'Avogadro = $6,022 \times 10^{23}$)

Pour les protéines on préfère parler en Daltons plutôt qu'en $\text{g}\cdot\text{mol}^{-1}$. Un Dalton correspond à $1,66 \cdot 10^{-27}$ kg.

Le Dalton

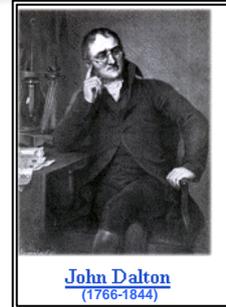
Protéine de masse molaire
 $82\,000\text{ g}\cdot\text{mol}^{-1}$



On écrit $82\,000\text{ Da}$ soit 82 kDa



$1\text{ mole} = 82\text{ kg}$



John Dalton
(1766-1844)

Un exemple : pour une protéine de masse molaire $82\,000\text{ g}\cdot\text{mol}^{-1}$, on préfère parler de $82\,000\text{ Da}$ soit 82 kDa . Une mole de cette protéine aura une masse de 82 kg .

Fonction = Structure

- Structure II, III
- Séquence I [®] structure spatiale

Insuline humaine / bovine ¹ 1 a-a

Insuline humaine / porcine ¹ 3 a-a

 **Protéines homologues**

La fonction des protéines est directement reliée à leur structure secondaire ou tertiaire. La structure primaire des protéines va, elle, être impliquée indirectement car elle permet ou non la mise en place des structures secondaires et/ou tertiaires en fonction des propriétés des résidus en question.

Par exemple, pour les insulines, humaine, bovine et porcines, il existe des différences au niveau de leur structure primaire, cependant elles ont toutes les trois une même fonction. On parle alors de protéines homologues.

Propriétés

Dénaturation : altération de la structure 3D de la protéine



Perte de fonction (définitive ou non)



Structure primaire conservée

Agents dénaturants : Chaleur, pH, Urée, Détergents (SDS)...

Cours CNAM, Rennes le 15/01/2009 Laurent Marché



Si la fonction des protéines est directement liée à leur structure, tous les traitements qui vont modifier la structure des protéines va modifier leur fonction. Si la structure primaire de la protéine est touchée, comme c'est le cas dans les lyses enzymatiques par exemple, on dira alors que la protéine est détruite.

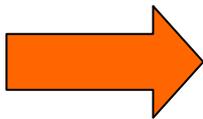
Cette dénaturation peut être passagère et la protéine peut alors retrouver sa fonction d'origine. Dans certains cas la perte de fonction est irréversible et la protéine a alors perdu sa structure définitivement.

Les agents dénaturant sont d'origine diverses : physique, mécanique, chimique, biologique.

Propriétés

Protéine = polymère chargé

Chaînes latérales et gpt -N et -C terminaux peuvent se charger selon le pH du milieu



le pI (point isoélectrique)

C'est le pH pour lequel la charge de la protéine devient nulle.

Le pH est la concentration en ions H^+ (H_3O^+) d'une solution. Les ions peuvent se fixer sur la protéine et faire ainsi varier sa charge.

Le pI est égal au pH pour lequel la protéine porte autant de charges + que -. La protéine est alors électriquement neutre. Elle n'est alors plus capable de migrer si elle est soumise à un champ électrique. (Propriété intéressante pour l'électrophorèse)

Analyse protéomique

- 1955 Première protéine séquencée par Sanger
 - 1958 Première protéine synthétique
 - 1975 Electrophorèse bidimensionnelle des protéines O'Farrell (J. of Biol.chem. 250:4007-402)
 - 1980 Gradient de pH immobilisé (reproductibilité)
 - 1990 Spectrométrie de masse appliquée aux protéines
- 1995 **Notion de protéome**

Cours CNAM, Rennes le 15/01/2009 Laurent Marché



l'Insuline est la première protéine séquencée entièrement par Sanger en 1955. C'est aussi la première produite de façon synthétique en 1958 puis par des moyens biotechnologique en 1979 sans la connaissance de la séquence nucléotidique.

-En 1975 Patrick O'Farrell met au point l'Electrophorèse bi-dimensionnelle (E2D) grâce à l'isoelectrofocalisation en gradient de pH. C'est le début de la migration des protéines en deux dimensions : selon leur pI et selon leur masse. Mais avec cette technique des dérives sont observées notamment pour les protéines basiques qui subissent une dérive cathodique et font que les premières dimensions de l'époque manque souvent de reproductibilité. En 1980 l'isoelectrofocalisation en gradient de pH immobilisé permet enfin la reproductibilité des gels.

-Au début des années 1990, ce sont les progrès en spectrométries de masse qui vont faciliter l'identification des protéines qui donnent un nouvel élan à l'analyse des protéines.

-1995 le terme protéome par analogie à génome apparaît.

-L'évolution des systèmes informatiques (l'expérience acquise pour le séquençage de génomes) bases de données, haut débit, mise en réseau des informations favorisent aussi le développement de la protéomique.

■ 1995 Notion de protéome



Déf. : Ensemble des protéines
exprimées par un organisme vivant.

Wilkins *et al.* (1995)

Biotechno. Gene. Eng. Env. 13: 1-50

-1995 le terme protéome par analogie avec celui de génome apparaît

-Le concept de protéome est assez récent puisqu'il a à peine dix ans
(Wilkins *et al* *Biotechnol. gene.Eng. Env.* 1995, 13, 19-50 et Wasinger *et al*
Electrophoresis 16 : 1090-1094 1995).

Analyse protéomique

- 2 types d'analyse protéomique



Analyse protéomique d'expression

- 2D
- Spectrométrie de masse
- Chromatographie

Il existe deux types d'analyse protéomique.

L'analyse protéomique d'expression qui s'intéresse à l'étude des protéines exprimées. Elle utilise principalement les techniques de E2D (présence/absence, quantification), la spectrométrie de masse (caractérisation de leur masse, de leur séquence, identification), la chromatographie (purification, identification)

- 2 types d'analyse protéomique



Analyse protéomique d'interaction

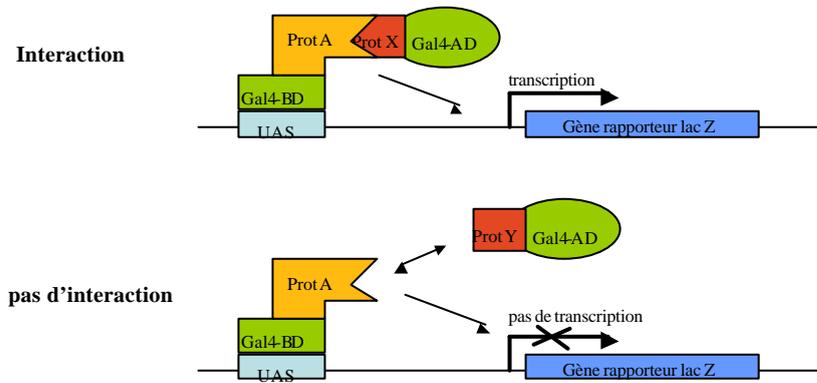
- Double-hybride
- TAP-TAG

L'analyse protéomique d'interaction s'intéresse quant à elle uniquement aux interactions entre les protéines : formation de complexes, auto-assemblage par exemple.

Les techniques utilisées pour étudier les interactions sont principalement le double-hybride et le TAP-TAG. Mais il en existe aussi d'autre comme le BIACORE, les puces à protéines permettant de réaliser des criblages et trouver rapidement les différentes interaction possibles.

Analyse protéomique d'interaction

Double-hybride



d'après Mathilde Allami, INRA Nantes

Cours CNAM, Rennes le 15/01/2009 Laurent Marché



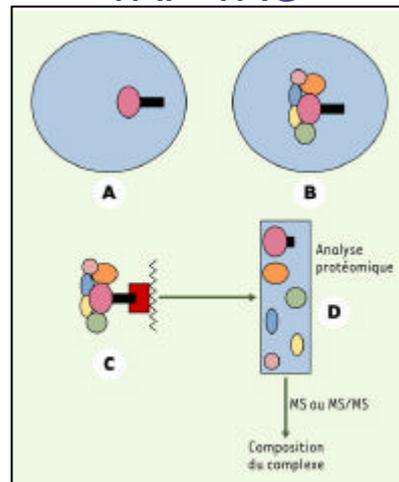
La technique du double hybride ne permet d'étudier les interactions que deux à deux.

Le gène rapporteur lac Z a besoin, pour être transcrit, que l'UAS (upstream activator sequence ou activateur de séquence en amont) soit actif. L'activation de l'UAS est soumise quant à elle à la mise en présence de protéines la GAL4-BD qui est un domaine d'accrochage et la GAL4-AD qui est un domaine activateur. L'activation de l'UAS ne peut se faire que lorsque les deux domaines sont en présence.

On va réaliser des constructions géniques qui permettent d'exprimer en fusion le domaine d'accrochage Gal4-BD et une protéine A. D'autres constructions qui permettent d'exprimer le domaine d'activation GAL4-AD et une autre protéine X ou Y dont on veut savoir si elles interagissent avec A. Les deux types de complexes sont produits en fusion par des plasmides (vecteurs) qu'on co-transformera deux à deux (A + X et A + Y) par exemple dans une levure ou une bactérie. Si la protéine X interagit avec A, alors il y a l'expression du gène rapporteur lac Z. La protéine Y n'interagit pas avec la protéine A et donc le gène rapporteur lac Z ne sera pas exprimé.

Analyse protéomique d'interaction

TAP-TAG



D'après Lescuyer et Rabilloud, 2004

Cours CNAM, Rennes le 15/01/2009 Laurent Marché



Le technique du TAP-TAG permet d'étudier des interactions entre une protéine et plusieurs autres à la fois.

A : On transfecte une bactérie ou une levure, par exemple, par une construction exprimant la protéine rose en fusion avec un module de purification (rectangle noir).

B : Cette protéine appât peut (ou non) former des complexes avec d'autres protéines de la cellule.

C : On récupère ce complexe grâce par exemple à une chromatographie (colonne d'affinité) qui accroche le module de purification et permet de séparer notre complexe des protéines des autres protéines.

D : La dernière étape consiste à éluer le complexe et à récupérer puis identifier par spectrométrie de masse les différents composants du mélange protéique.

L'analyse protéomique d'expression



Etude globale de l'expression protéique d'un système biologique (cellules, tissus, liquides physiologiques, organisme) à un moment donné.

Nous détaillerons dans ce cours l'analyse protéomique d'expression.

L'analyse protéomique d'expression permet de visualiser

- Les protéines solubles
- Les protéines en quantité suffisantes
- Les protéines présentes



? du protéome

Le protéome c'est l'ensemble des possibles :

- On ne peut étudier que les protéines en quantité suffisantes = seuil de détection de la méthode
- On ne peut étudier que les protéines présentes = c'est un instantané de l'expression du génome

PROTEOME ou TRANSCRIPTOME ?

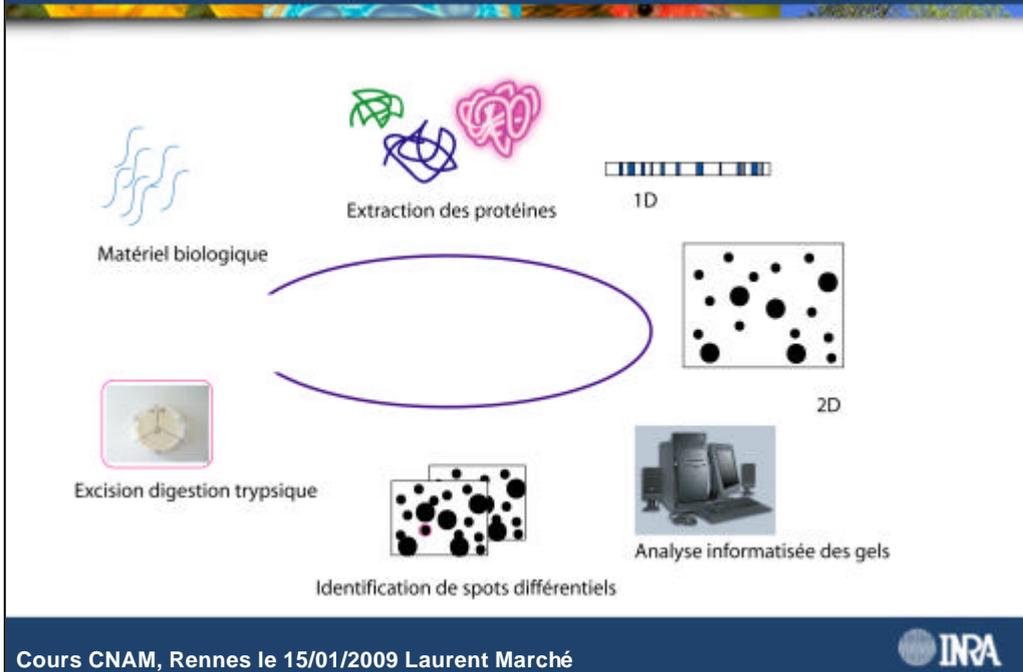
- Produit final des gènes
- Modification post-traductionnelles
- Faible corrélation quantitative
ARNm/Protéines

La question de savoir pourquoi étudier plutôt le protéome que le transcriptome (ARNm) peut effectivement se poser. En effet, beaucoup d'études existent sur le transcriptôme, cependant on peut parfois obtenir plus d'information par le protéome, ou en tous cas des informations complémentaires.

C'est par exemple le cas lorsque :

- On est en présence d'ARNm muet, qui ne sont présent dans la cellule mais jamais traduit en protéine (A. Vincent, 1980)
- On a une faible corrélation quantitative en la quantité d'ARNm présent et la quantité de protéine produite (NL. Anderson, 1997)
- Il existe une faible corrélation quantitative de la **variation** d'ARNm par rapport à la **variation** de quantité de protéine produite (D. Rousseau, 1992. S. Khlochbin, 1991)

Les différentes étapes de l'analyse protéomique



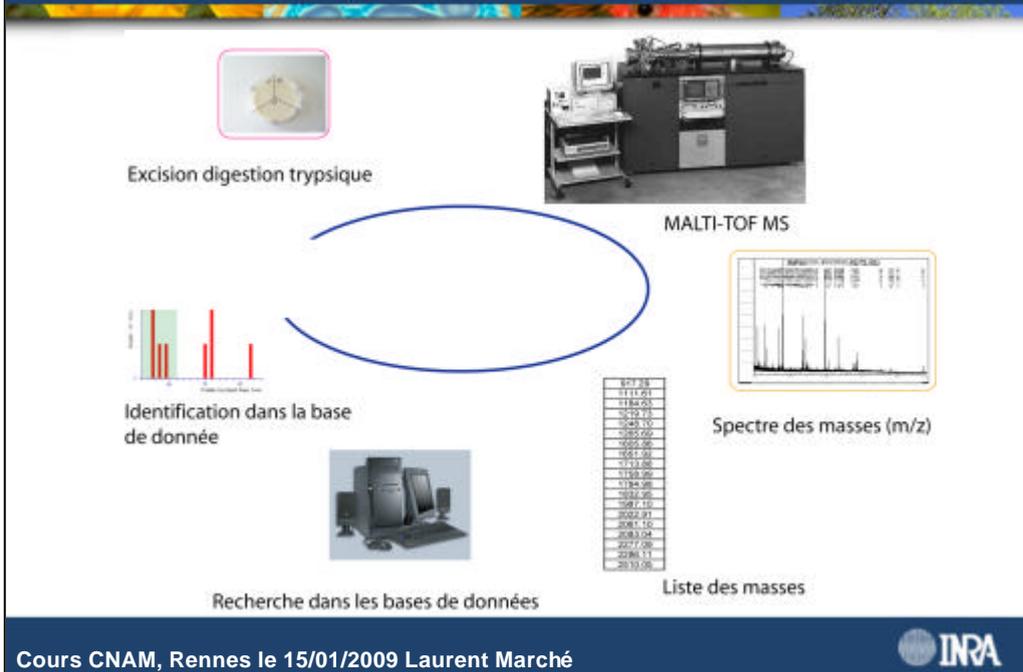
Cours CNAM, Rennes le 15/01/2009 Laurent Marché



Nous allons détailler dans la suite du cours les différentes étapes de l'analyse protéomique d'expression.

- 1/ l'extraction des protéines à partir du matériel biologique
- 2/ L'électrophorèse de première dimension
- 3/ L'électrophorèse de deuxième dimension
- 4/ L'analyse assistée par ordinateur des gels qui permet d'identifier les spots protéiques d'intérêt
- 5/ La récupération et la digestion des spots protéiques

Les différentes étapes de l'analyse protéomique



Cours CNAM, Rennes le 15/01/2009 Laurent Marché



- 6/ L'analyse au spectromètre de masse
- 7/L'identification des protéines

Extraction des protéines



Pas de protocole type



Adaptation à :

- **L'échantillon**
- **L'objectif de l'étude**

Il n'existe pas de protocole type pour l'extraction des protéines. Les protocoles d'extraction des protéines varient selon : - **le matériel biologique** et **l'objectif fixé** pour l'étude à réaliser.

- **Objectifs :**
 - **Solubiliser un maximum de protéines**
 - **Récupérer uniquement des protéines**
 - **Préserver l'intégrité des protéines**

Les objectifs de l'extraction des protéines sont listés ci-dessus.

Il faut garder à l'esprit que :

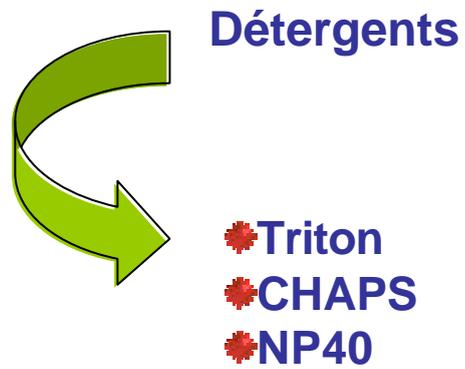
- 1- il manquera toujours une fraction ou une autre selon le protocole utilisé, et notamment les protéines non solubles.
- 2- il faut éviter de récupérer les agents sources d'artefacts pour la suite de l'étude
- 3- toutes modifications du type lyse par protéase ou cassure des protéines qui changeraient les résultats de la 2D doivent être évités voire supprimés autant que faire se peut.

▪ Principes :

- **Déstructurations des membranes**
- **Dénaturation des protéines pour les rendre solubles**
- **Protection contre la carbamylation**
- **Limitation de la précipitation**
- **Élimination des substances sources d'artefacts**

Les principes mis en œuvre pour l'extraction des protéines sont :

- 1- déstructuration des membranes et/ou parois cellulaires pour récupérer les protéines
- 2- permettre l'individualisation et la solubilisation des polypeptides : rupture des ponts disulfures et des liaisons non covalentes
- 3- l'urée en solution est en équilibre avec le cyanate d'ammonium qui provoque la carbamylation des protéines. Cette modification provient quand l'iso-cyanate, un produit de dénaturation de l'urée, modifie de façon covalente les résidus lysine induisant alors un changement de pI des protéines.



Pour déstructurer les membranes on utilise des détergents qui vont modifier l'équilibre de la double couche lipidique et ainsi permettre la libération des protéines.



Agents dénaturants

- ✿ **DTT** (*Dithiothréitol*)
- ✿ **SDS** (*Sodium dodecyl sulfate*)
- ✿ **Urée ou Thiourée**

Les agents dénaturants sont eux utilisés afin de rompre les liaisons et ponts disulfures afin d'individualiser les protéines.



Ampholytes

✿ **ex. Biolyte™ 3-10**

La présence de l'urée dans le milieu peut provoquer une carbamylation des protéines qui n'est pas souhaitable. On ajoute donc des ampholytes qui vont subir la carbamylation en lieu et place des protéines.

L'extraction des protéines

▪ Substances sources d'artefacts à éliminer

Substances à éliminer	Elimination
Lipides	détergent
Acides nucléiques	ultracentrifugation, Dnase, Rnase
Sels	limiter les intrants
Protéases	méthode dure (plantes), ou inhibiteur (ex.Leupeptine)

Cours CNAM, Rennes le 15/01/2009 Laurent Marché



- On peut de façon générale utiliser des Kits qui permettent d'éliminer les différentes substances, cependant on ne sait pas trop ce qui, au niveau protéique, est éliminé aussi.
- Sels : une partie sera éliminée lors de la 1D, sinon l'utilisation de kit est préconisée.
- Méthode dure = Acide trichloroacétique et SDS concentré (12 à 15% au lieu de 8) à chaud. On fait bouillir ce qui provoque une dénaturation ultrarapide qui ne donne pas le temps aux protéases d'agir puisqu'elles sont elles aussi rapidement dénaturées et perdent alors leur fonction protéase.

■ Protocole nématodes

- β -mercaptoethanol
- Biolyte 3-10
- Urée
- Tris-HCl (pH7,4)

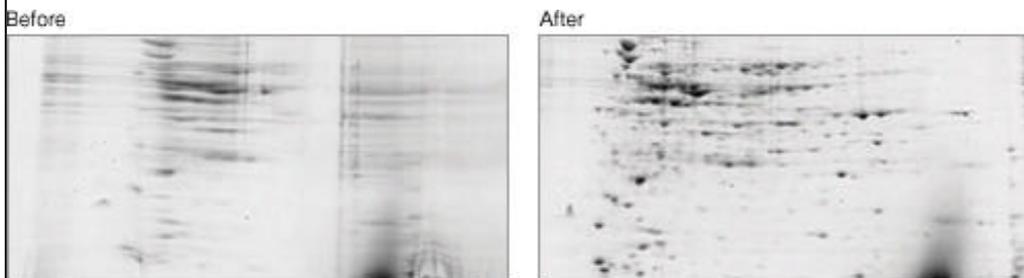
■ Protocole levures

- β -mercaptoethanol
- Tris HCl pH 8
- SDS
- NaCl, NaF
- Inhibiteur de protéases :
Leupeptine , pepstatine ,
soybean trips inhibitor...
- Dnase, Rnase
(*D., Schieltz, 2000*)

Comparaison de 2 protocoles un léger et un plus lourd à cause du matériel étudié qui n'est pas de même nature : des nématodes et des levures.

- Nématodes : peu de lipides et aussi peu de cellules (entre 600 et 1000)
- Culture de levures : vacuoles riches en protéases, richesse en acides nucléiques

L'extraction des protéines



**Gel E2D d'extrait protéique
D'*Escherichia coli* avec un excès de sels (NaCl 1M)**

Cours CNAM, Rennes le 15/01/2009 Laurent Marché

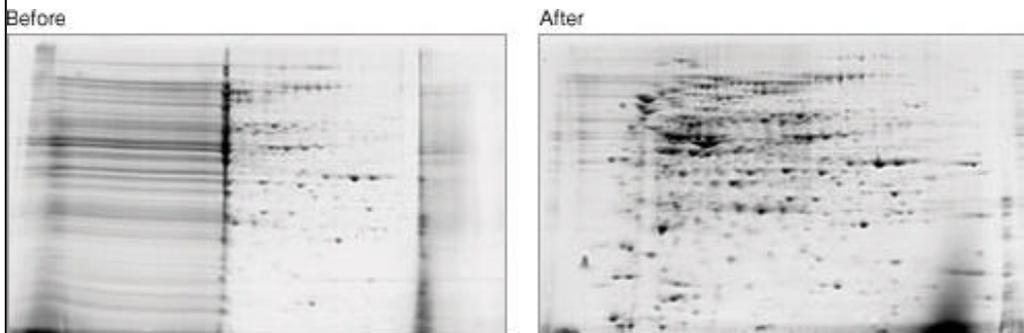


Elimination des sels par kit ready-prep 2D cleanup.

Extrait de *E.coli* avec NaCl 1M. Strip pH 3-10

On peut voir l'électrophorèse avant ou après le nettoyage de l'échantillon par le kit.

L'extraction des protéines



Gel E2D d'*E.coli* présentant un excès de SDS

Cours CNAM, Rennes le 15/01/2009 Laurent Marché



On peut voir les même effets avec un excès de détergent, de lipides ou de composés phénoliques.

La encore on visualise avant l'élimination de l'excès de SDS et après.

L'extraction des protéines

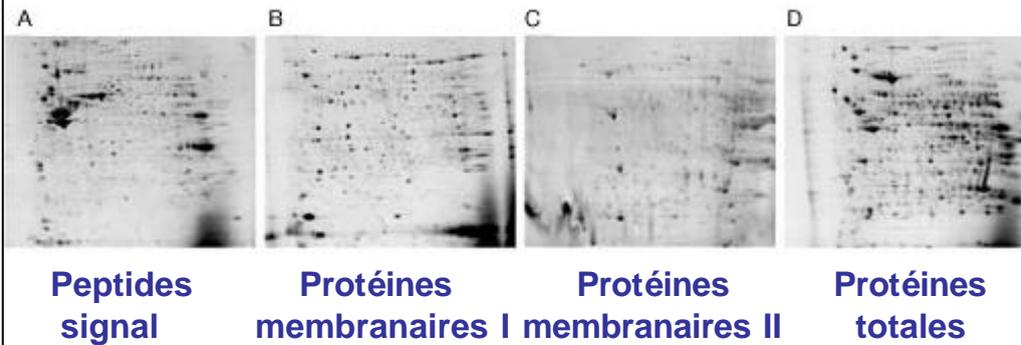
- **Extrait protéique total**
- **Extrait protéique partiel**
 - **Protocole d'extraction spécifique**
 - **LC (affinité, échange d'ions)**

Les protocoles d'extractions varient aussi en fonction de l'objectif de l'étude

On peut ne vouloir travailler que sur les protéines membranaires ou encore uniquement les protéines acides ou basiques.

L'extraction des protéines

Gels 2 D pH 3-10 Foie de souris



Cours CNAM, Rennes le 15/01/2009 Laurent Marché



On peut voir ici sur ces différents gels les dépôts issus d'un même échantillon de foie de souris mais extrait selon différents protocoles permettant de ne récupérer que la fraction souhaitée des protéines.

Légende : Differences in 2-D patterns obtained using ReadyPrep signal (A), membrane I (B), membrane II (C), and total protein (D) kits. Mouse liver samples were extracted using the recommended protocols for each kit. Purified protein (~450 μ g) was loaded onto 17 (A, B, C) or 24 cm (D), pH 3? 10 nonlinear ReadyStrip IPG strips. Overall spot patterns differ for A, B, and C, even though all three kits used isolate membrane proteins. Also observed is an enrichment of certain proteins in the membrane fractions that is not readily visible in the total protein extract (D).

Séparation des Protéines



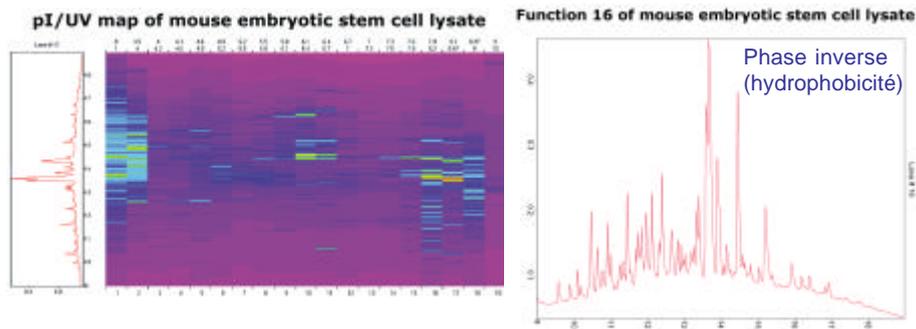
Techniques séparatives

- **Chromatographie 1 D et 2D**
- **Puces à Protéines**
- **Electrophorèse bidimensionnelle (E2D)**

La technique de l'E2D que nous allons détailler par la suite est une technique séparative au même titre que la chromatographie ou encore les puces à protéines. Les informations qu'elle donne portent sur le pI des protéines et leur masses.

Techniques séparatives

■ Chromatographie 1D et 2D



+ Haut débit / - Quantité de protéines (1 à 5 mg)

On peut travailler avec des colonnes de chromatographies qui permettent de séparer les protéines selon différentes caractéristiques physico-chimiques. On peut par abus de langage parler de deux dimensions.

- On réalise une première chromatographie qui sépare les protéines selon leur pI (échange d'ion) (1D) (avec 0,3 unités pH d'écart). Les protéines sont récupérées ensuite dans chaque collecteur pour réaliser une deuxième chromatographie.
- La deuxième chromatographie peut être, par exemple, une phase inverse (non poreuse) séparation selon polarité et/ou hydrophobicité (donc pas selon leur taille comme en 2D SDS PAGE).

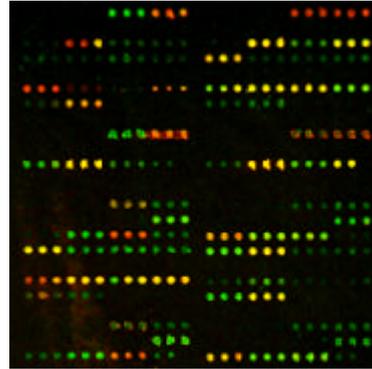
• AVANTAGE : haut débit

• INCOVENIENT il faut une grande quantité de protéine (exemple chez Beckman coulter ils préconisent entre 1 et 5 mg de protéine contre 2D SDS PAGE entre 50 et 300 µg)

Techniques séparatives

■ Puces à protéines

**Identifier en parallèle
plusieurs protéines
au sein d'un mélange**



+ Haut débit / - Protéines connues

Légende : Partie d'une puce à protéine GENEPIX 400A scan "allumée" avec 110 anticorps différent avec approximativement 9 réplicats.

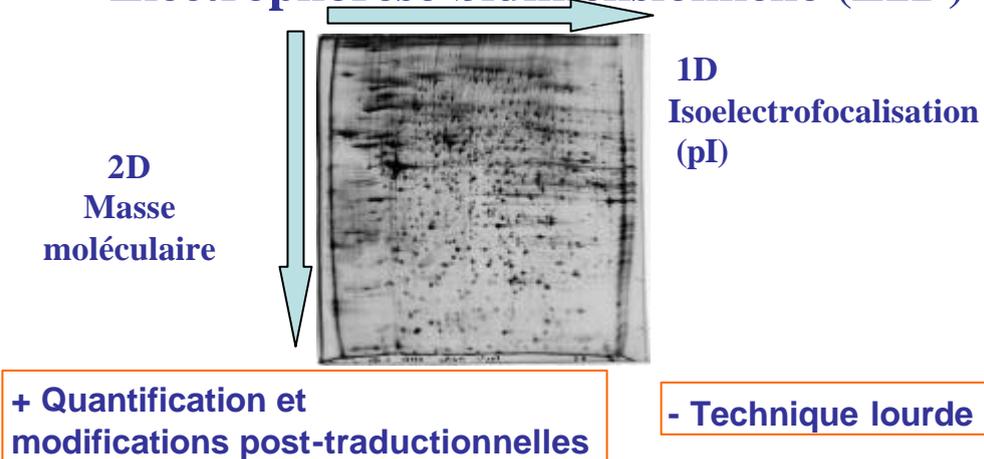
Une puce à protéine permet idéalement d'identifier en parallèle plusieurs protéines au sein d'un mélange. L'identification se fait en utilisant la réaction antigène-anticorps entre des protéines à identifier et des anticorps spécifiques fixés à la surface de la puce.

AVANTAGE : haut débit

INCONVENIENT : on identifie ce qu'on connaît, en effet il faut déjà posséder des anti-corps spécifiques et donc ce sont des anti-corps de protéines déjà connues.

Techniques séparatives

■ Electrophorèse bidimensionnelle (E2D)



La technique de l'électrophorèse bidimensionnelle permet elle de travailler aussi sur un mélange de protéines.

Elle permet :

- Une séparation par différences de pI (1D) et masse moléculaire (2D)
- Aujourd'hui utilisée comme une technique séparative en vue du passage au spectromètre de masse pour identification
- Avantage : quantification, identification de protéines inconnues, et des modifications post-traductionnelles

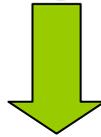
Electrophorèse bidimensionnelle



▪La première dimension

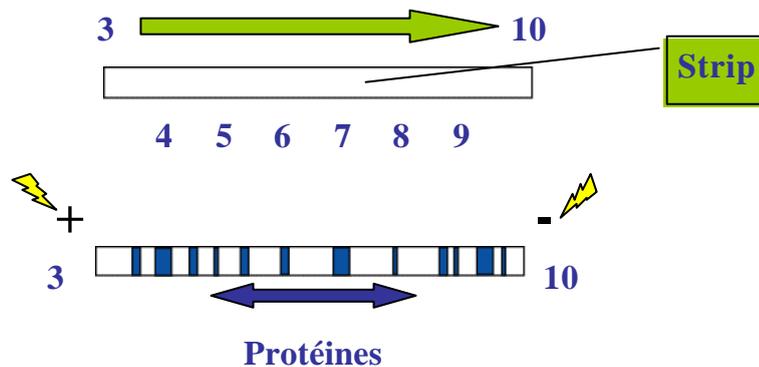
Principe de l'isoélectrofocalisation :

La charge des protéines s'annule quand le pH du milieu est égal à leur pI.



Les protéines soumises à un champ électrique s'immobilisent à $\text{pH} = \text{pI}$

La première dimension

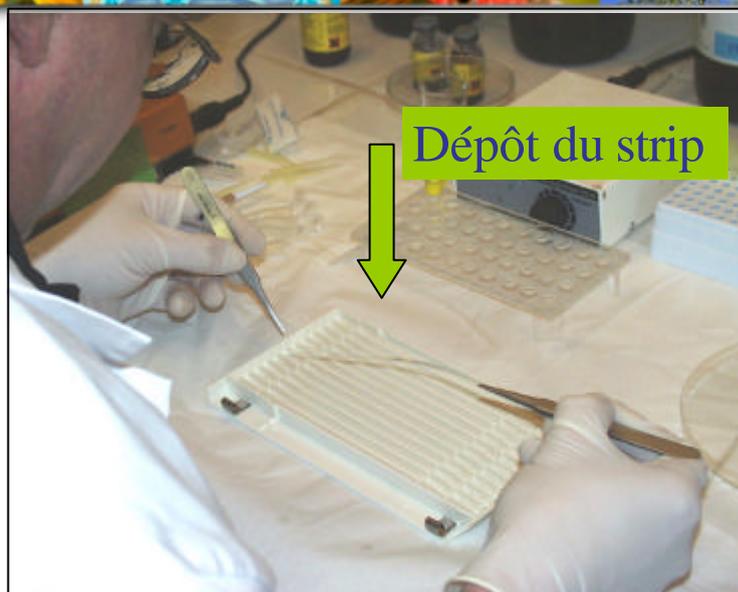


On utilise des bandelettes de gels appelées STRIP (ici ex 3-10). Le gel est un gel à gradient de pH immobilisé, c'est-à-dire qu'on a coulé un mélange d'acryl et d'ampholyte qui vont co-polymériser. C'est pourquoi les ampholytes deviennent immobilisés (immobiline). Le pH varie ainsi dans la bandelette de gel progressivement de gauche à droite sur la figure du pH 3 au pH 10. Sous l'effet d'une différence de potentiel électrique il y a création d'un champ électrique. Les protéines soumise à ce champ vont migrer en passant dans le gradient de pH et s'immobiliser lorsque pH est égal à leur pI. On peut donc séparer les protéines selon leur pI.

La première dimension : migration des protéines selon leur pI

- **Mise en présence strip à gradient de pH avec échantillon protéique**
- **Réhydratation strip : passive-active**
- **Focalisation isoélectrique (IEF)**

La séparation des protéines

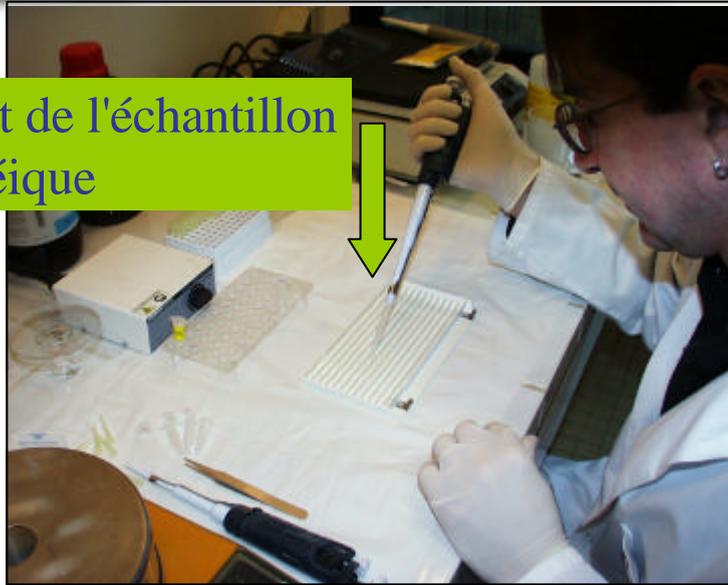


Cours CNAM, Rennes le 15/01/2009 Laurent Marché



La séparation des protéines

Dépôt de l'échantillon
protéique



Cours CNAM, Rennes le 15/01/2009 Laurent Marché



- Ici on dépose entre 50 et 150 μg de protéines
- On peut déposer jusqu'à 300 μg environ

La séparation des protéines

Hydratation
passive
10 min

Suivi par

Hydratation
active
50V / 10 h



Cours CNAM, Rennes le 15/01/2009 Laurent Marché



Les différentes phases de l'IEF :

- Hydratation passage des protéines dans les strips
- Etape 1 : dite de conditionnement les excès de sels sont éliminés en migrant vers les électrodes.
- Etape suivantes sont celle dit de focalisation proprement dite.

La séparation des protéines

IEF

Étape 1
200 V / 1h

Étape 2
200-1000V
en 1h
1000 à 10000V
en 6 h



La séparation des protéines

IEF

Étape 3

pH 3-6 60 000V

pH 5-8 80 000V

pH3-10 100 000V

(ex. 4h30 en 3-6
et 9 h en 7-10)



Electrophorèse bidimensionnelle



La deuxième dimension

La séparation des protéines

Equilibration :

- Ionisation des protéines par le SDS
- Elimination des composants artefactuels en 2D. (Urée, thiourée, amp holytes, détergents)



Cours CNAM, Rennes le 15/01/2009 Laurent Marché



Ici l'étape d'équilibration des strips après l'isoélectrofocalisation pour préparer les protéines à subir le deuxième dimension.

- Ionisation par le SDS : permet de charger toutes les protéines de façon négative en rendant leur charge native négligeable. Elles peuvent alors migrer selon leur MM. (charge constante par unité de masse)
- Rupture ponts disulfure par le DTT (conditions dénaturantes)
- La migration sera bien fonction de la MM. Il existe une relation linéaire entre la MM et la migration relative

La séparation des protéines

Equilibration :

- Dénaturation des protéines par DTT et SDS

- Iodoacétamide

- éviter les trainées sur les gels due au DTT

- éviter le retour de ponts disulfure

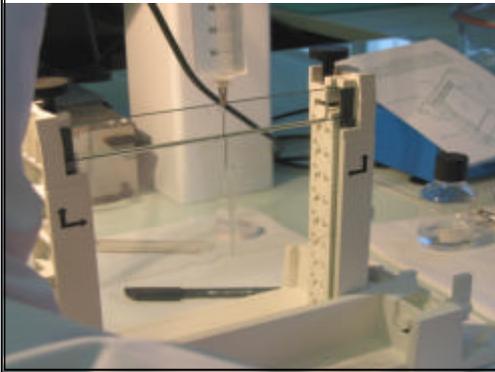


Cours CNAM, Rennes le 15/01/2009 Laurent Marché



Les objectifs de l'équilibration sont la dénaturation des protéines et la mise en place de charges de même nature sur toutes les protéines. Et d'autres part, avec l'ajout de l'iodoacétamide, d'obtenir une meilleure netteté des gels et d'éviter le retour des ponts di-Sulfures.

La séparation des protéines



**Gels acrylamide +
pipérazine diacrylamide
(10%)
20 x 20 cm**



**Vérification de la
jonction strip/gel**

Cours CNAM, Rennes le 15/01/2009 Laurent Marché



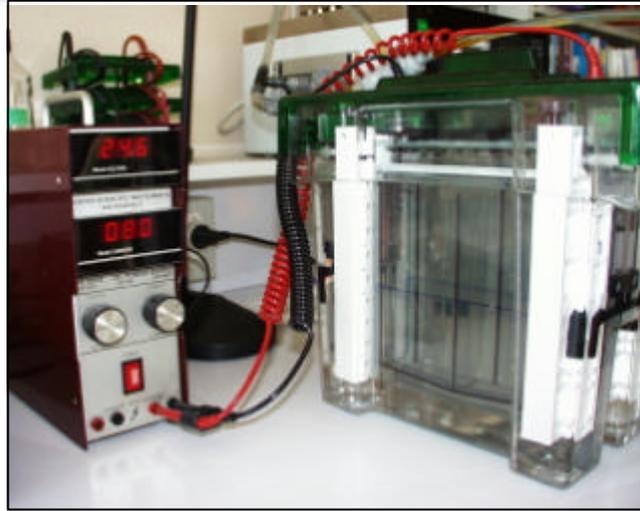
On ajout de l'agarose après s'être assuré de la bonne jointure entre le strip et le gel. En effet, il ne faut pas de phase liquide pour que le passage se fasse correctement entre le strip et le gel.

La séparation des protéines

- **Electrophorèse**

**10 mA/gel
20 min**

**20 mA/gel
5 h**



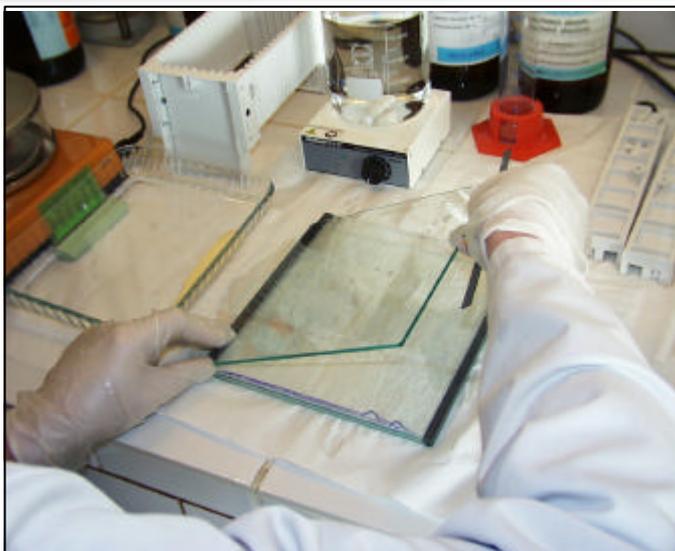
Cours CNAM, Rennes le 15/01/2009 Laurent Marché



- Tampon de migration ici = Tris-glycine
- Au début un faible ampérage est appliqué pour laisser le temps au SDS d'éluer les protéines et permettre aussi le passage des protéines du strip au gel.

La séparation des protéines

Démoulage des gels



Cours CNAM, Rennes le 15/01/2009 Laurent Marché



Avant la coloration le gel est démoulé

Méthodes de détection des protéines

-Exemple du nitrate d'argent

Méthodes de détection des protéines

Type de détection	Limite de sensibilité	Quantification	Spectrométrie de masse
Nitrate d'argent	200 pg	++	Oui (sauf si glutéraldéhyde)
Bleu de coomassie	100 ng	+++	Oui
Coloration fluorescente	400 pg	++++	Oui
Marquage fluorescent	250 pg	++	Oui
Radioélément	1 pg	+++	Non
Phosphore imager	0,2 pg	++++	Non

Cours CNAM, Rennes le 15/01/2009 Laurent Marché



Le nitrate d'argent AgNO_3 :

- bien pour la sensibilité, on peut détecter de très faibles quantités de protéines
- peut s'avérer insuffisant pour quantification car l' AgNO_3 permet la quantification mais de façon linéaire uniquement sur une petite gamme de concentration, il y a très vite saturation.

▪ Coloration fluorescente : ex sypro ruby

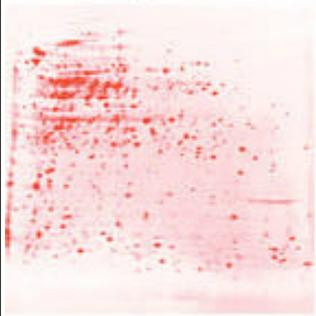
▪ Marquage fluo : Cy dyes ou monobromobimane (avant IEF), DIGE (coloration différentielle) Cydies

Avantage en fluo pas de bruit de fond car le marquage est fait avant le gel.

- Phosphoimager : écran de phosphore (cassette) capture rayon bêta, gamma et X provenant de gels radio marqués (autoradiographie). Puis image latente lu par balayage d'un laser. (bonne quantification car chaque point est évalué)

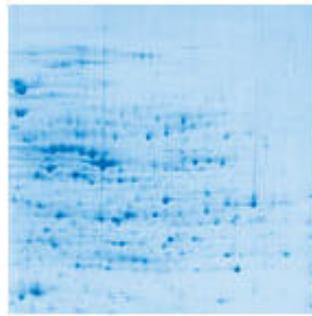
Méthodes de détection des protéines

SYPRO Ruby



400 pg

Coomassie Blue



100 ng

Silver Stain



200 pg

Cours CNAM, Rennes le 15/01/2009 Laurent Marché

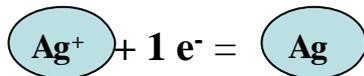


Comparaison des différents seuils de sensibilités pour différentes méthodes de décoloration : SYPRO Ruby, Bleu de Coomassie, et Nitrate d'argent

Méthodes de détection des protéines

Exemple de la coloration au nitrate d'argent

Principe : hypothèse fixation-réduction
des ions Ag^+ en ions Ag .



Amine

Cys

Met

Il y aurait fixation sur les groupements $-\text{NH}_2$ des protéines et préférentiellement sur les Methionine et Cystéine des protéines. Les ions Ag^+ se fixeront préférentiellement à ces endroits : groupements amines, et aussi sur les résidus soufrés Methionine et cystéine

Fixation



**Protéines rendues insolubles
(Ethanol, Acide acétique)**

**Elimination des composés sources
d'artefacts (glycine, Tris, SDS, ampholytes)**

Durée des bains 1 h 30 au total

Sensibilisation



**Liaison entre Tétrathionate ($K_2O_6S_4$),
Acétate de potassium (KCH_3COO) et les
protéines**

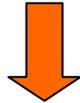
L'objectif est de limiter le bruit de fond en rendant l'accrochage des ions AG préférentiellement sur les protéines.

Imprégnation



Ajout AgNO₃

Développement (ou réduction)



**Formaldéhyde, carbonate de potassium,
thiosulfate**

Réduction des ions Ag^+ en Ag sensible (couleur noire)

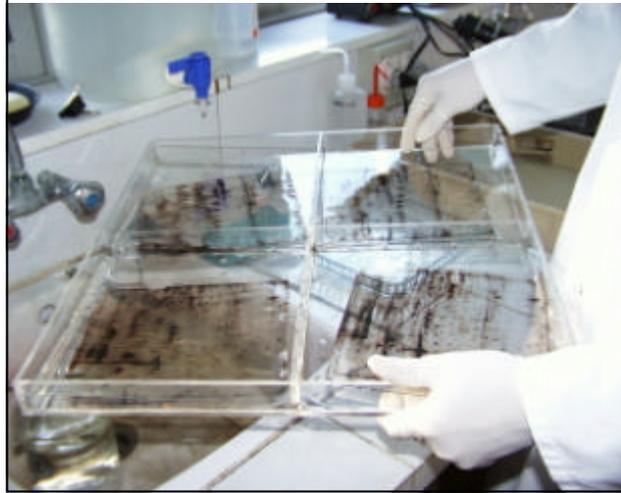
Méthodes de détection des protéines

- **Arrêt de la réaction**

- Tris

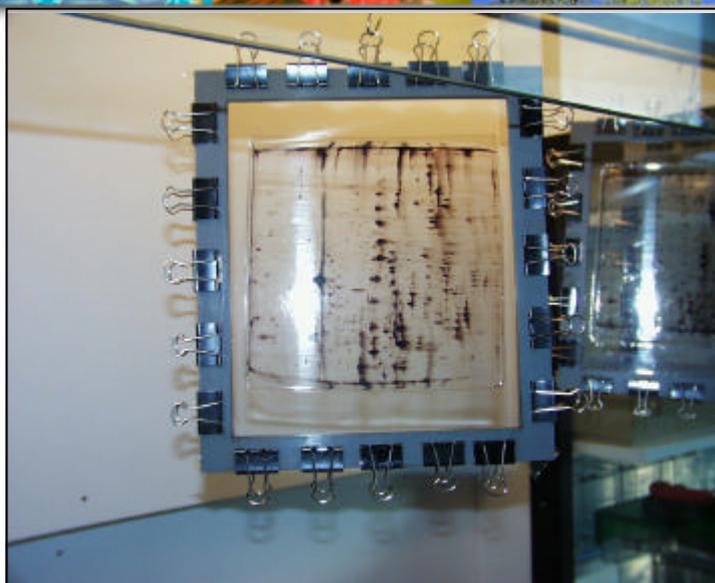
- Acide acétique

- **Rinçage**



Méthodes de détection des protéines

▪ Séchage des gels



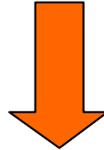
Cours CNAM, Rennes le 15/01/2009 Laurent Marché



Analyse d'images assistée par ordinateur



POURQUOI ?



- **Nombre de gels par expériences**
- **Nombre de spots par gels
(souvent entre 800 et 2000)**

Lors d'étude de protéomique il faut en général réaliser environ 4 gels pour un échantillon. Pour certaines expériences, s'il y a multiplication des échantillons leur analyse devient très vite trop lourde.

L'utilisation de l'analyse d'image informatisée est vite rendue inévitable.

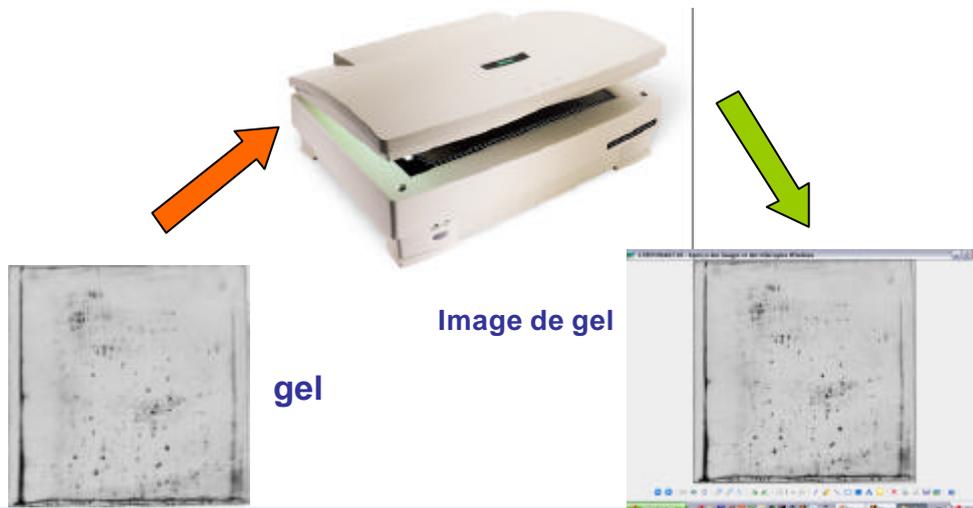
Logiciels

- Kepler
- PDQuest
- Phoretix2D
- **Imagemaster 2D Platinum V (Melanie)**

Différents logiciels existent. Imagemaster 2D platinum est le plus utilisé.

Analyse d'images assistée par ordinateur

Scanner densitomètre

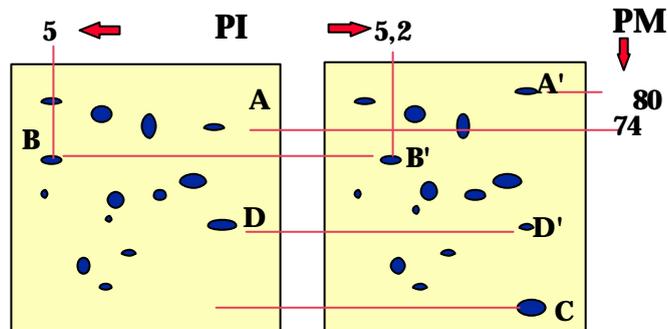


Cours CNAM, Rennes le 15/01/2009 Laurent Marché



On peut aussi acquérir les images grâce à une camera CCD, une densitomètre laser, ou un phosphoimager

Que cherche-t-on à comparer ?



Sur les gels on cherche à comparer des différences.

Ces différences peuvent être de différents ordres :

A-A' variation de poids moléculaire

B-B' variation de point isoélectrique

C-C' protéines spécifiques (présente dans un échantillon mais pas dans l'autre)

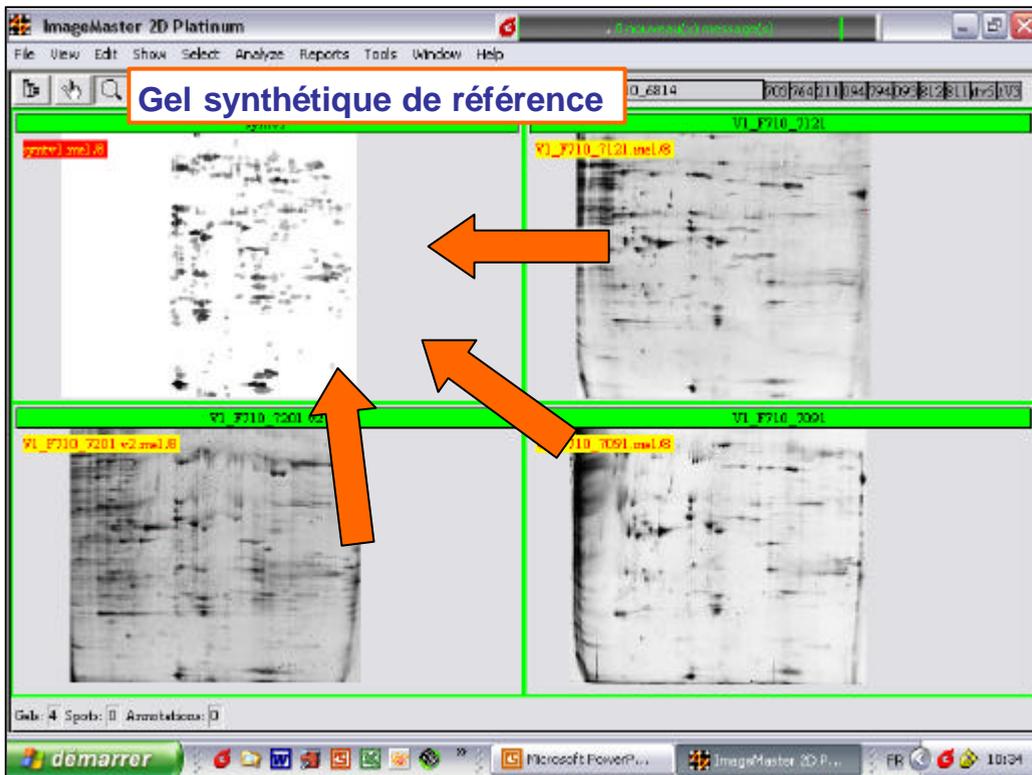
D-D' variation d'expression entre deux échantillons



Les gels scannés sont introduits dans une expérience au sein du logiciel.

On peut calibrer l'intensité pour compenser les différences de colorations (ceci ne touche pas les images réels mais juste leur affichage. Si on veut faire de l'analyse quantitative il est préférable de travailler avec des gels qui sont très semblables en ce qui concerne la coloration.

On choisit un gel de référence (copie) ou on peut aussi travailler avec un gel synthétique qui est la somme de tous les gels qui doivent être analysés.



Gels de synthétique = somme des spots présents dans tous les gels

Paramètres de détection des spots protéiques

**Smooth : nombre de passage avant
détection**

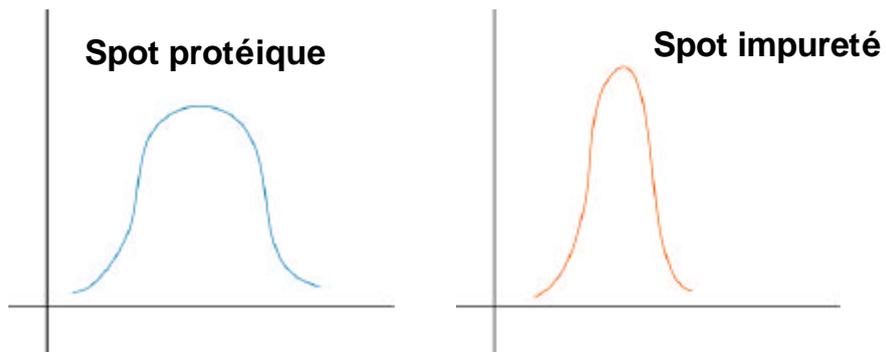


séparation des spots proches

Ce paramètre est régi par un algorithme qui permet d'isoler des spots très proches.

Paramètres de détection des spots protéiques

Saliency : Seuil de "courbure" des spots

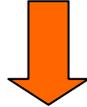


Courbure du spot : Paramètre très variable entre les gels

Un spot protéique vu sur un gel à une base large le bord n'est pas net contrairement aux impuretés qui ont un bord en général très tranché. C'est ce que représente les deux courbes du dessus.

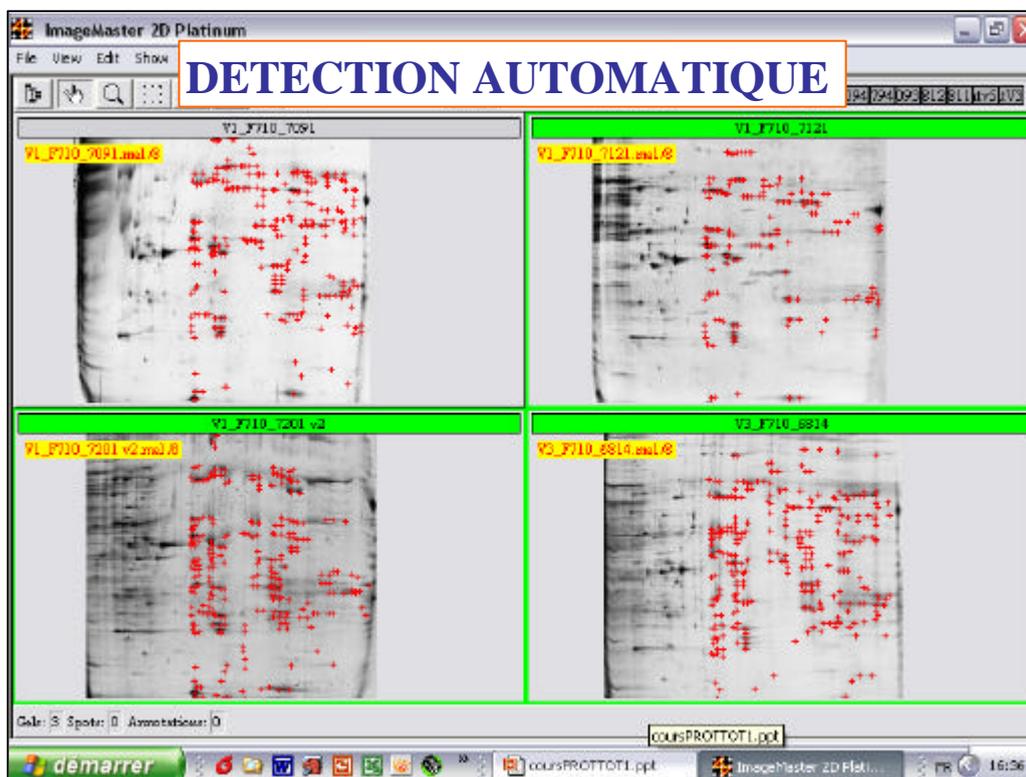
Paramètres de détection des spots protéiques

Aire minimale



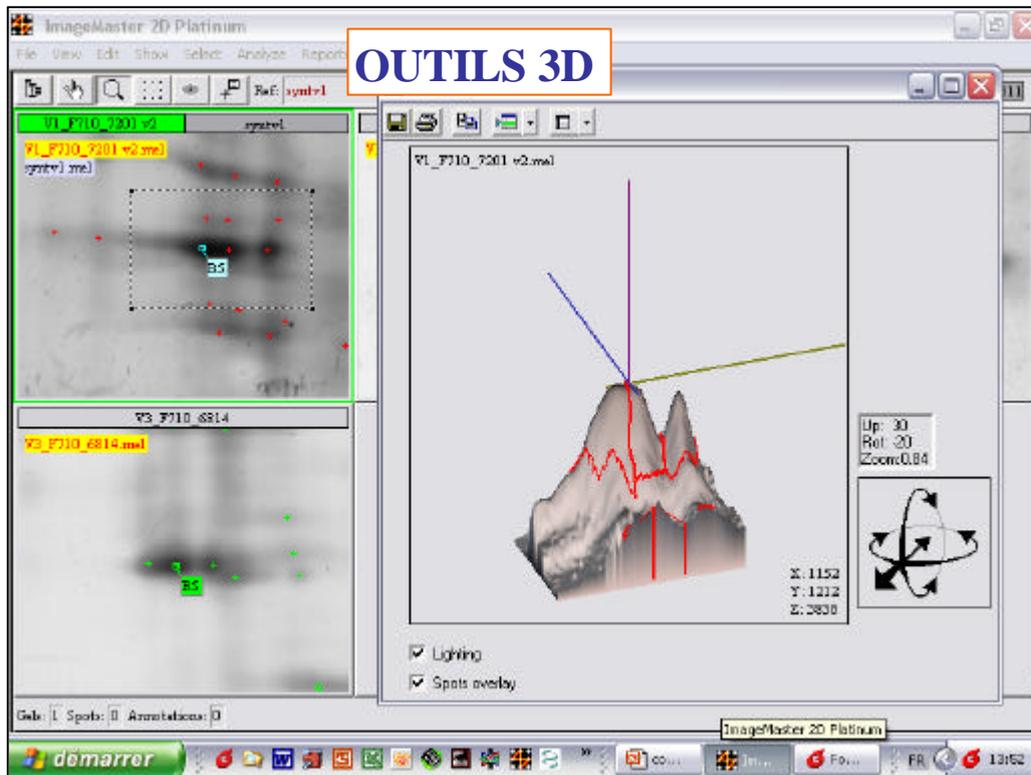
**Elimination des taches trop petites
(grains d'argent et impuretés)**

On peut définir une aire minimale de spots en dessous de laquelle le logiciel ne prend pas en compte la tache afin de ne pas introduire d'artefacts dues à des impuretés.



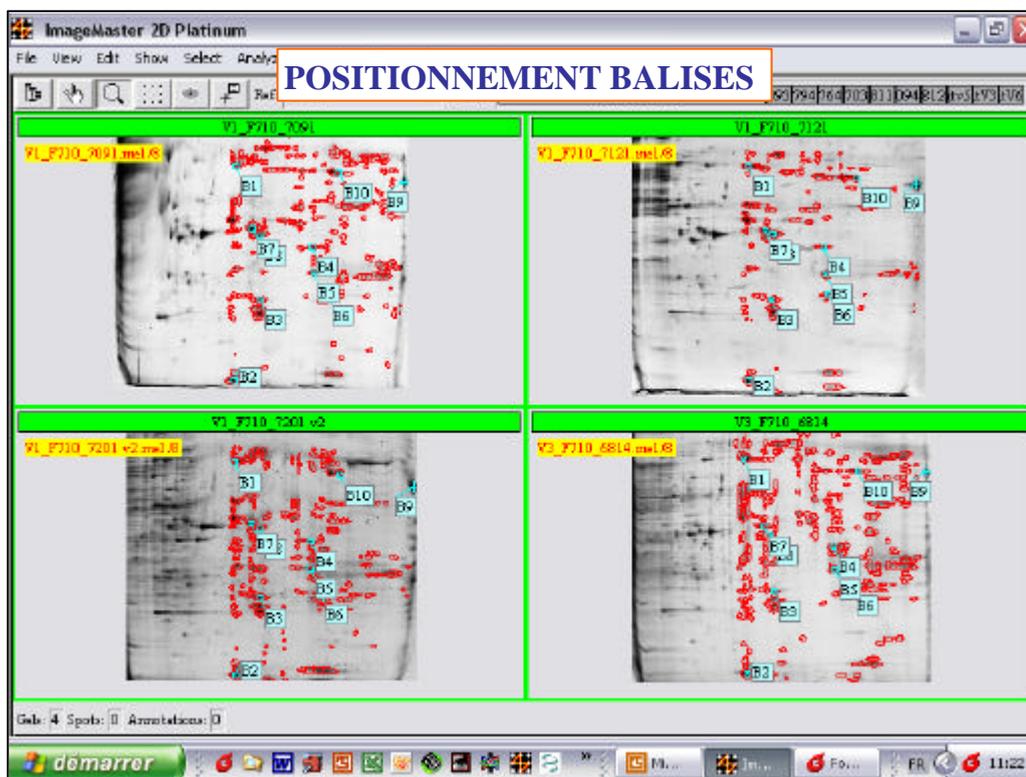
Gels après détection automatique et nettoyage

Lourde étape de nettoyage et de vérification en se référant aux gels vrais, en effet suite à cette étape il faut vérifier qu'il n'y a pas de spots omis par le logiciel ou un spot qui est compté deux fois car il est un peu large.

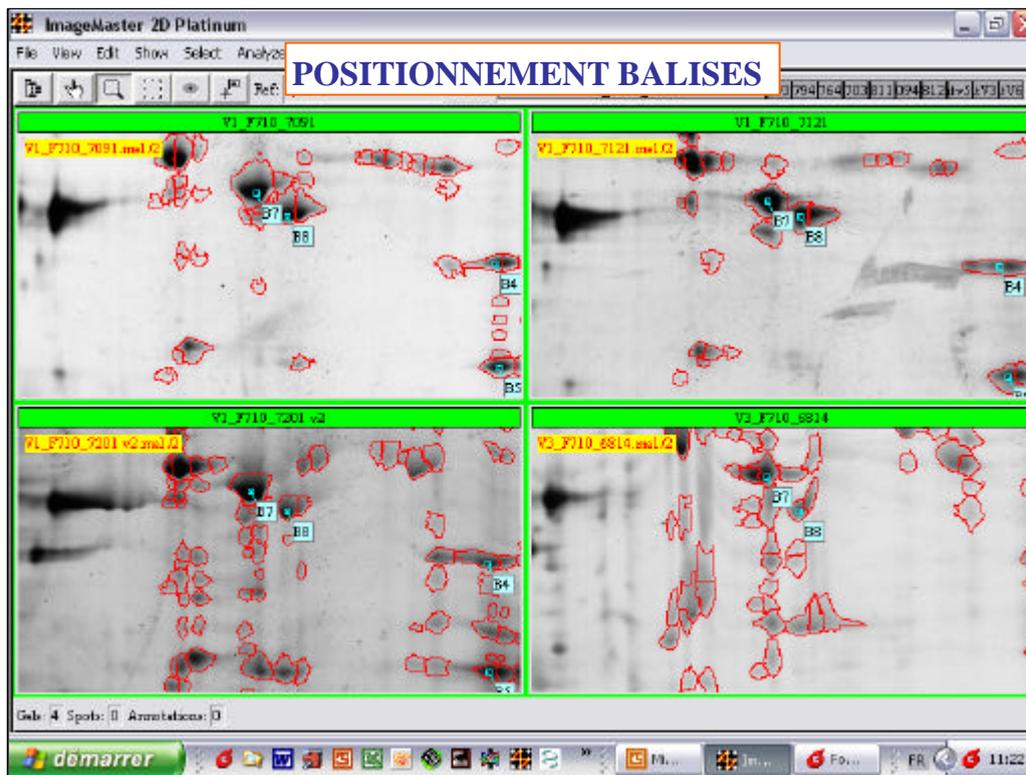


Outils d'aide à la décision : Vision 3D

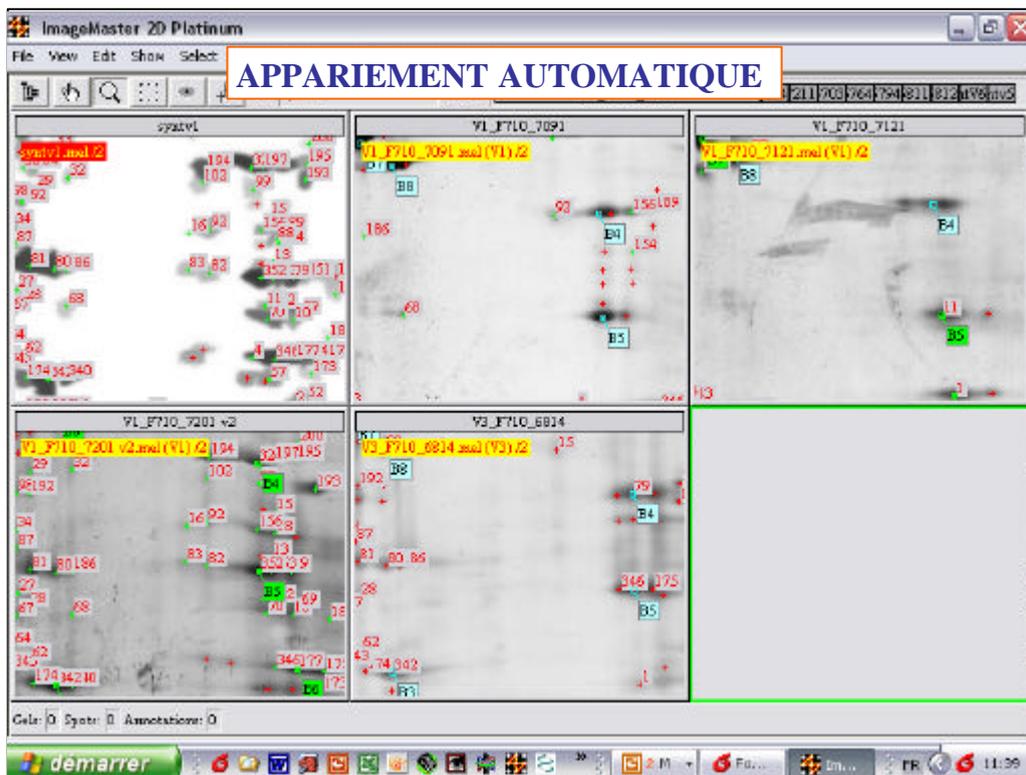
Permet de vérifier la détection des spots et de séparer ou réunir des spots proches.



Les balises sont des spots qui sont présent partout sur quasiment tous les gels.



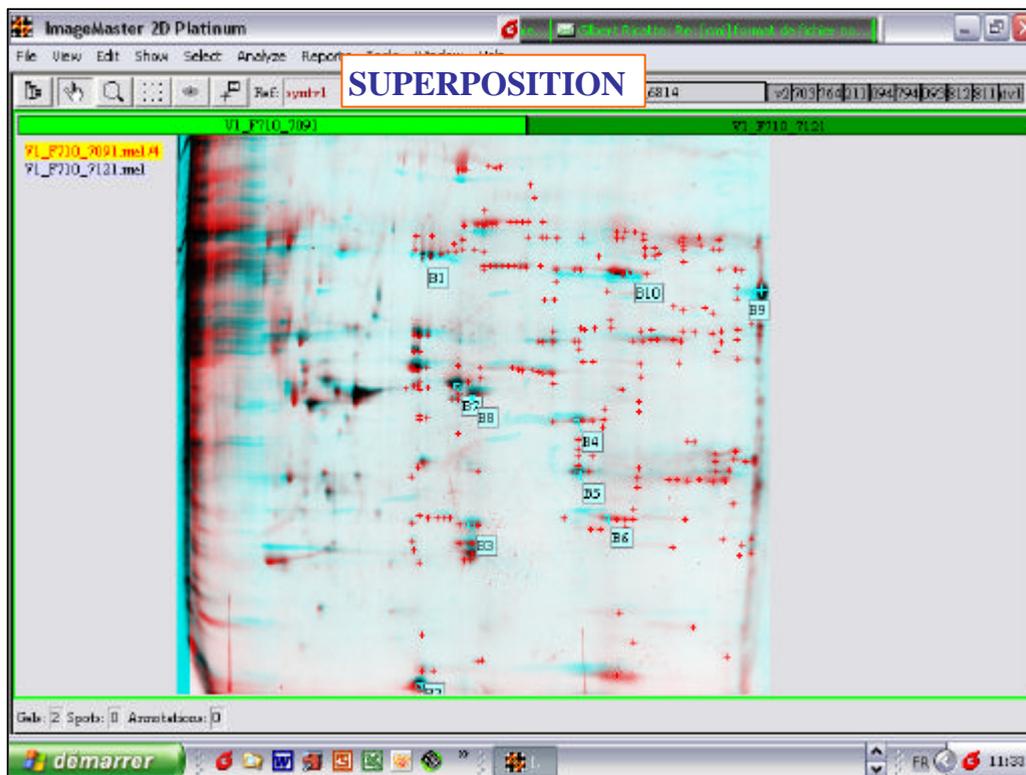
Ces spots agissent comme des repères pour le logiciel qui va ensuite identifier chaque spots avec une référence unique grâce à la triangulation.



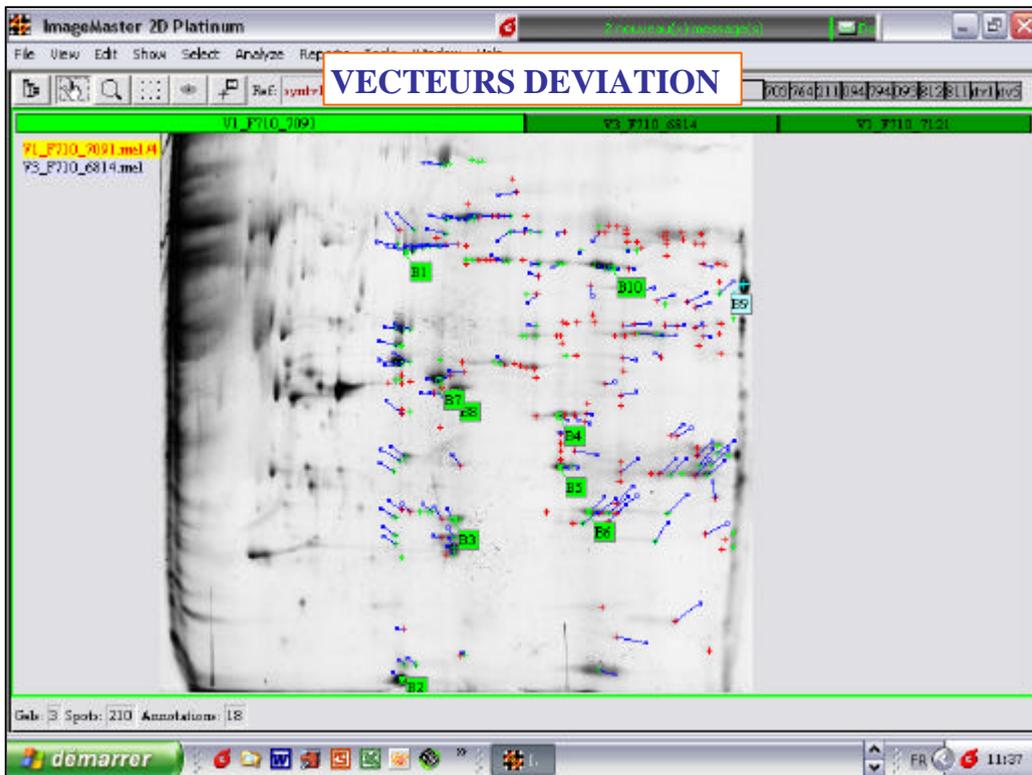
Une fois les spots identifiés de manière unique sur chaque gel, ils peuvent être appariés entre eux. C'est à dire que le logiciel indique que tel spots doit être le même sur le gel A et B.

Appariement automatique : l'algorithme procède par triangulation de proche en proche à partir des balises

On doit ensuite là aussi vérifier que ce sont le bon spot qui ont été appariés.



La vérification peut se faire en superposant les gels. On peut ainsi visualiser les décalage entre spots.



La vérification peut aussi se faire en utilisant les vecteurs de déviations qui permettent eux aussi de repérer les pairs qui ne sont pas correctes.

Ces vecteurs matérialisent sur un gel le décalage spatiale entre deux spots considérés comme ayant la même position sur un gel A et B. Lorsque se décalage est trop important ou qu'il n'est pas dans le même sens que celui pour les spots voisins, alors on peut considérer que le logiciel a fait une erreur.

Groups Report

Report on Groups (%Vol).
The chosen statistics are Mean (100%) and M.S.D.

Group ID	Mean (100%)	M.S.D.	Variation	synby1	Y1_F710_2091	Y1_F710_7121	Y1_F710_7201 v2
196	330	4.17090	1.25522	0.300948	2.62083	4.07376	6.12002
197	332	0.911350	0.236238	0.259214	0.616902	1.27489	0.845366
198	334	0.915871	0.322169	0.351786	0.567537	0.778937	1.43903
199	335	1.98373	0.472801	0.238340	1.27790	2.21208	2.56146
200	336	0.129297	0.099044	0.772133	0.0904203	0.140941	0.277022
201	337	0.113272	0.0371958	0.328059	0.0772099	0.174815	0.0930899
202	338	0.260059	0.0909675	0.349796	0.141763	0.212303	0.302037
203	339	1.23198	0.354170	0.287480	0.779917	1.11406	1.76326
204	340	0.632964	0.190398	0.300803	0.415570	0.904766	0.711564
205	341	0.319472	0.105808	0.331470	1.208870	0.220835	0.401675
206	342	1.43779	0.588767	0.409506	0.862714	0.992041	2.37355
207	343	0.504149	0.200752	0.660046	0.397167	0.103364	0.634222
208	344	0.218414	0.151527	0.693761	0.154413	-	0.368517
209	345	0.201739	0.0698845	0.346411	0.134947	0.141993	0.223183
210	346	0.517036	0.462167	0.893917	0.774812	1.14048	-
211	347	0.208885	0.210458	1.00753	0.381450	0.451050	-
212	349	0.0593329	0.0600734	1.01246	0.131964	-	0.105366
213	351	0.109488	0.129969	1.18706	0.119938	0.318015	-
214	352	0.380519	0.390548	1.00034	0.747016	-	0.775058
215	353	0.204105	0.207962	1.01899	0.351688	-	0.464734
216	354	0.0476281	0.0571284	1.20052	0.0605139	-	0.139236

Une fois les vérifications faites pour les pairs on peut éditer les rapports pour chaque spots. En orange l'identité d'un spots. En vert leur présence ou absence dans les gels. En bleu le spot n°340 est présent dans les trois familles de gels. En jaune, le spot 344 qui lui est absent dans la première famille.

Tests statistiques

-Test T (variable diffère entre classes)

**-Test Kolgomorov-smirnov
(deux échantillons sont de la même classe)**

Sur les résultats qui sont présence/absence ou quantification on peut réaliser des tests statistiques.

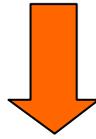
Test T montrer qu'un variable diffère ou non entre deux classes (pop)

Test Kolgomorov vérifier que deux échantillons sont bien de la même classe (pop)

Autres outils

Analyse heuristique

**l'algorithme procède par approximation
et propose la meilleur solution.**



**un dendrogramme présentant
le regroupement en classe de gels**

Ce regroupement en classe des gels permet par exemple de mettre en évidence l'existence de groupes ayant le même profil protéique ou un profil proche.

Autres outils

Analyse factorielle

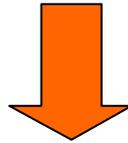
En cas d'expérience lourde, on réduit le nombre de données en un petit nombre de facteurs explicatifs.



- Comment les gels se regroupent.
- Quelles protéines caractérisent le mieux ce regroupement.

Résultats finaux

**Différentiel présence/absence
selon les classes**

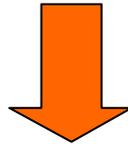


Liste de référence des spots

Cette classification permet lors d'une étude différentielle, par exemple deux pathotypes de ravageurs ayant des virulences différentes, de mettre en évidence des spots qui existent dans un pathotype et pas dans l'autre. On peut alors, si on a la chance d'identifier les protéines, trouver les protéines qui sont spécifiquement impliquées dans la pathogénicité.

Résultats finaux

Différentiel d'expression



**Liste de référence des spots ayant
des volumes différents selon les classes**

Ce sont les résultats obtenus lorsqu'on travaille en quantitatif. On peut alors dire que telle ou telle protéine est sur-exprimée ou sous-exprimée dans tel ou tel type d'échantillon.

Excision-digestion

Comme on vient de le voir les résultats ayant permis de repérer les spots protéiques d'intérêt, il faut maintenant les récupérer pour identifier les protéines.

- **Objectif : digestion des protéines pour obtenir des peptides.**



Utilisables pour l'identification au spectromètre de masse.

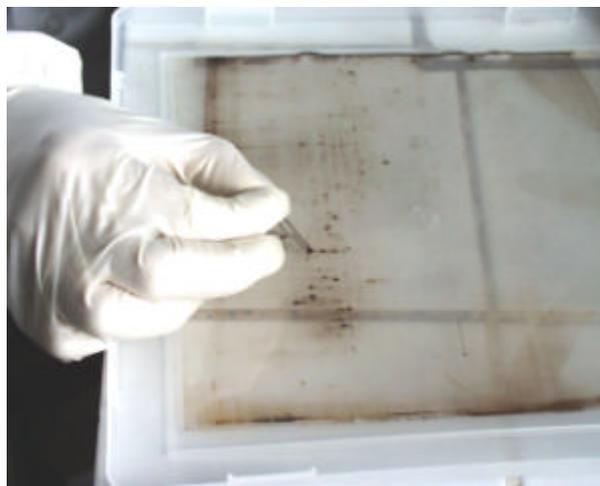
Principe : on digère par une protéase connue : trypsine , chymotrypsine.

La trypsine, par exemple, est une [endoprotéase](#) qui hydrolyse les liaisons peptidiques dans lesquelles un acide aminé basique (Lysine (Lys), Arginine (Arg)) engage sa fonction acide. Elle coupe en C-terminal de ces acides aminés. Elle coupe après les résidus Lys K et Arg R.

On obtient des peptides dont on évaluera la masse en spectrométrie. On obtiendra donc un profil de masse correspondant à l'enzyme utilisée pour la digestion.

Excision-digestion

- **Excision du spot d'intérêt repéré lors de l'analyse d'images**



Excision-digestion

- Digestion enzymatique
- **TRYPSINE 20 µg/ml**
37°C o/n en milieu
humide et pH acide
- **Coupe les protéines**
après K (Lys) et R
(Arg)



Identification des protéines



TECHNIQUES

- **Immuno-détection**
- **Microséquençage des acides aminés N-terminaux**

L'étape de détection était réalisé jusqu'à ces derniers temps avec des techniques plus ou moins lourdes

L'immuno-détection qui nécessite de posséder les anticorps ce qui implique la détection de protéines connues et donc limite fortement ce type d'outils.

Le micro séquençage des résidus N-terminaux est une technique assez lourde que tous les laboratoire ne possèdent pas, de plus elle n'est pas toujours possible pour toutes les protéines.

TECHNIQUES

Spectrométrie de masse

MS



**Empreintes
peptidiques
massiques**

MS/MS



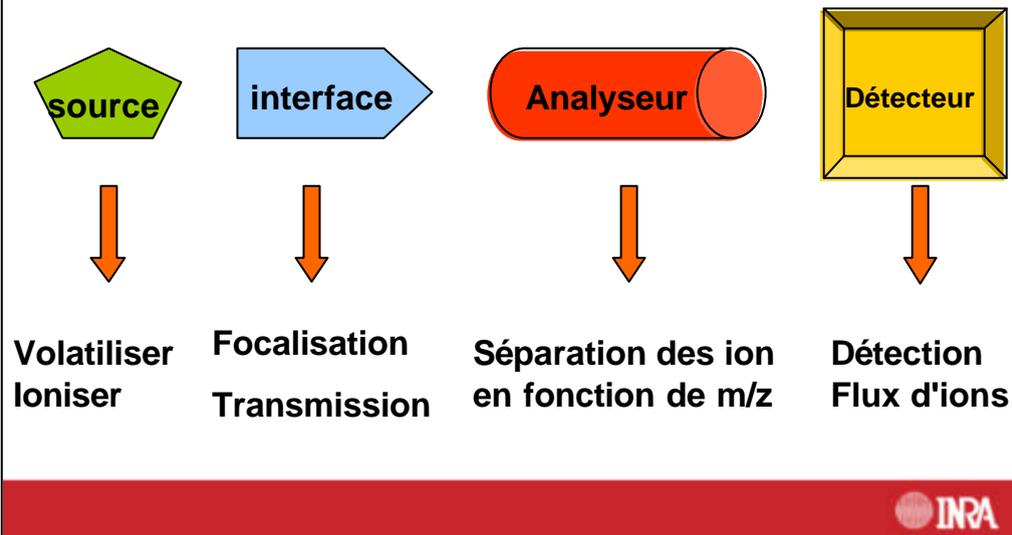
Séquences de protéines

En général en protéomique on commence par du MALDI (MS) puis si on obtient rien on passe par MS/MS. Les deux techniques donnent des résultats différents.

Dans le cas de la MS (mass spectrometry), on obtient en fait uniquement une empreinte de masse peptidiques alors que dans le cas de la MS/MS obtient des informations aussi sur les séquences.

Identification des protéines

Schéma d'un spectromètre de masse



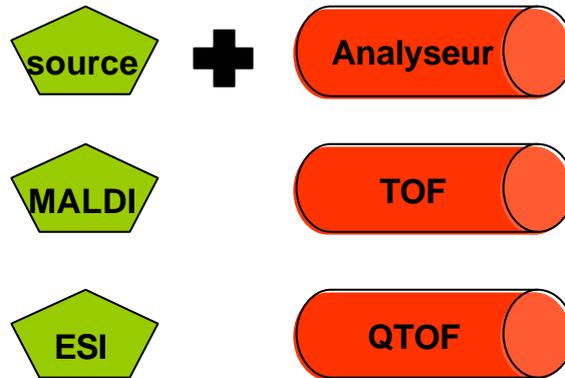
Voici les différents compartiments d'un spectromètre de masse classique.

On trouve tout d'abord une source qui permet de volatiliser et d'ioniser la protéine. Ensuite les peptides passent dans l'interface qui permet la focalisation du flux de peptides et la transmission de ceux-ci à l'analyseur.

L'analyseur est une partie importante car c'est en son sein que se réalise la séparation des ions en fonction de leur rapport m sur charge (m/z).

Enfin le flux de peptides percute un détecteur qui signale leur arrivée et le décalage de temps de vol entre les différents impacts.

Nom d'un spectromètre



Le nom d'un spectromètre de masse toujours donné par les deux éléments les plus importants le type de source et le type d'analyseur.

Deux exemples un MALDI-TOF ou un ESI-QTOF.

TECHNIQUES

Spectrométrie de masse

MS



- MALDI-TOF
- ESI-TOF
- Seldi-TOF

MS/MS

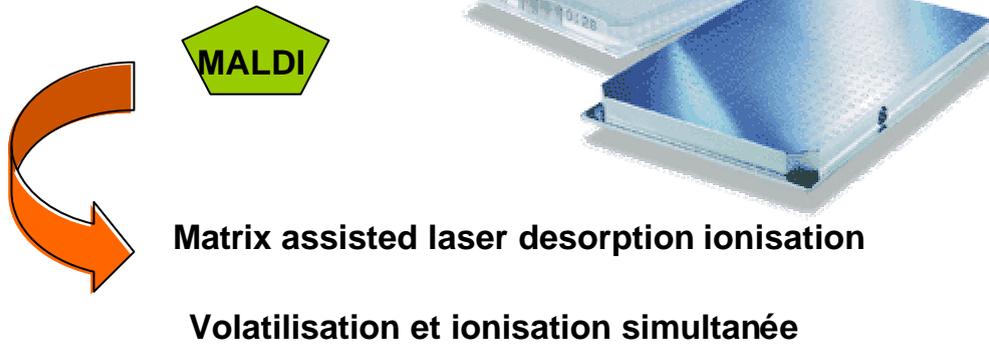


- LC-MS/MS
- ESI-QTOF

Les différents types de spectro ne sont pas utilisés pour les mêmes objectifs. La MS qui ne donne des infos que sur les masses des peptides est faite par MALDI-TOF, ESI-TOF, Seldi-TOF. La MS/MS qui renseigne sur les séquences des peptides issus de la digestion de la protéine est réalisée par, entre autres, des LC-MS/MS et ESI-QTOF

Identification des protéines

MALI-TOF MS



Pour la MALDI-TOF :

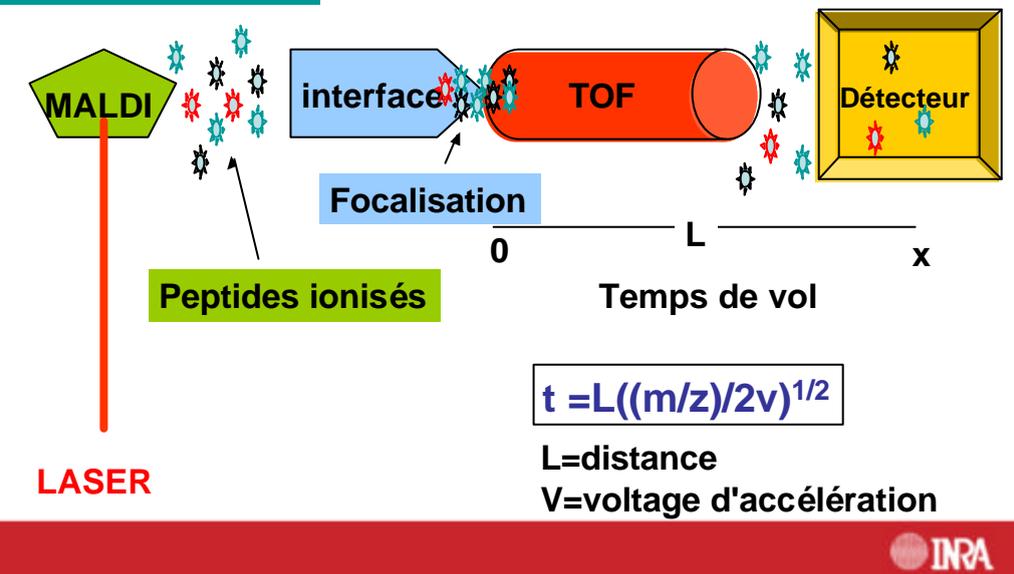
L'échantillon est mélangé à un matrice organique (acide 4, alpha-cyanocyanaminique), puis déposé pour cristallisation sur la cible.

La matrice est utilisée afin de capter le maximum de l'énergie du laser ce qui permet d'éviter l'altération des peptides.

La chaleur qui monte à plusieurs milliers de degrés au point d'impacte du laser (tir laser de qq nanosecondes) va créer un plasma contenant surtout des e- des ions H-, K+, O-. Qui vont se combiner avec les peptides de l'échantillon qui deviennent des ions. Le plasma se vaporise et les ions sont donc volatilisés et soumis à un champ électrique dans le vide pour se déplacer vers l'analyseur puis le détecteur

Identification des protéines

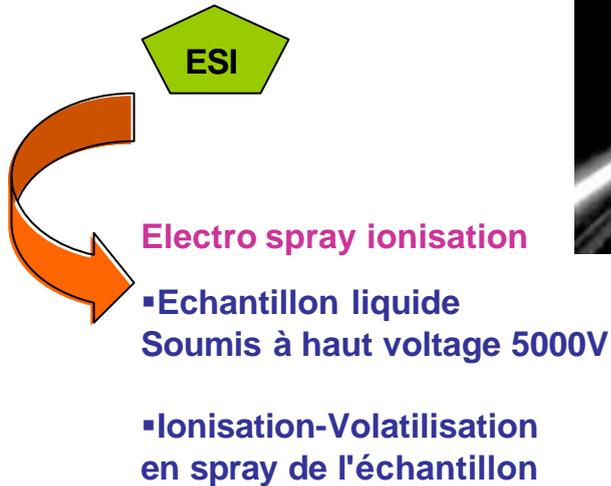
MALI-TOF MS



L'énergie apportée par le laser sur le mélange peptides+matrice va ioniser l'échantillon. Les peptides ionisés passent par l'interface pour organiser le flux. Les flux de peptides passe ensuite dans l'analyseur. Dans cet analyseur on applique une très forte tension qui va accélérer les peptides ionisés qui vont ensuite donc se déplacer vers le détecteur.

Leur vitesse est corrélée à leur masse et leur charge et leur temps de vol. Enfin, il percute le détecteur qui permet de révéler les différences de temps de vol entre les différents peptides ionisés. Le temps de vol d'un peptide étant relié à sa masse (cf formule ci-dessus), on peut en déduire la masse de chaque peptide composant le digesta de la protéine.

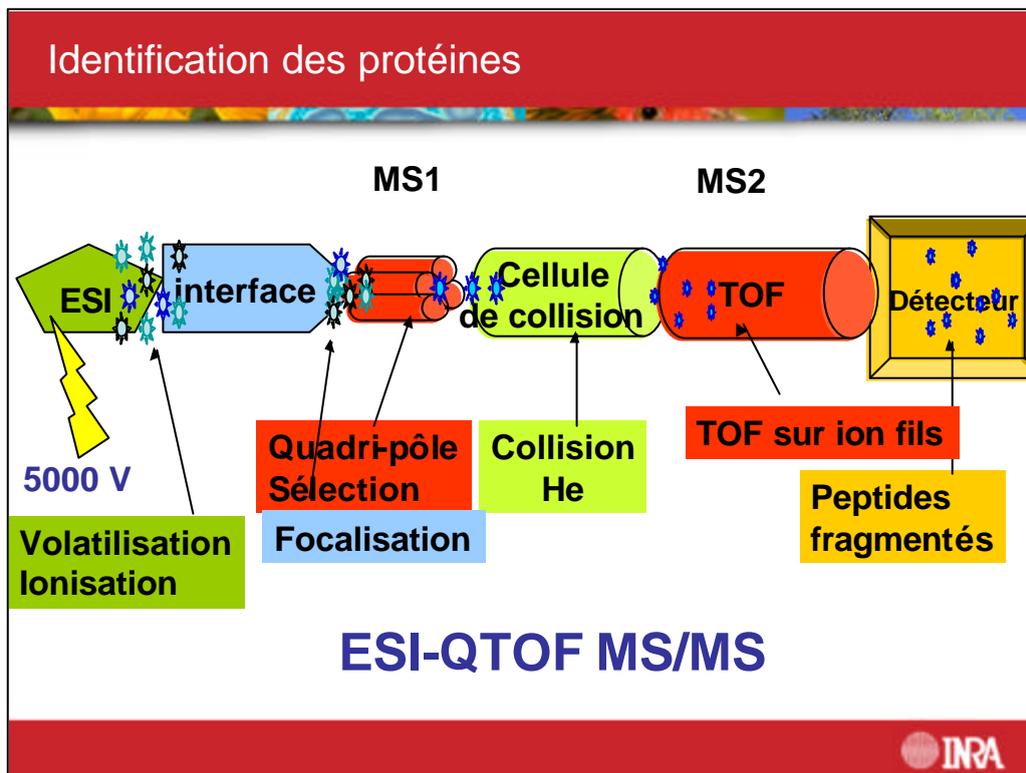
ESI-QTOF MS/MS



Dans le cas de spectromètre de masse dont la source est une source électrospray. L'échantillon qui contient les peptides est injecté grâce à une aiguille en verre recouverte d'or et de palladium et soumise à un très haut voltage.

Les peptides sont dans le même temps ionisés (différence de potentiel électrique) et en même temps volatilisés dans le spray. Ils sont donc prêts à être envoyés dans l'analyseur.

On peut coupler avec ce système un système de chromatographie liquide qui va permettre de faire une première séparation des peptides.



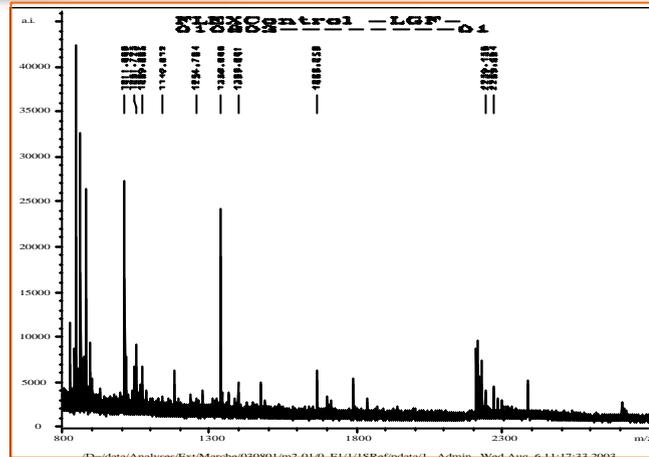
Le QTOF, schématisé ici, permet d'obtenir des séquences de peptides.

Après la volatilisation les peptides sont dirigés vers un quadri-pôle électrique. Cette chambre fait entrer en résonances les peptides qui sont détruits. On peut effectuer un balayage de fréquences de résonances, ceci permet de sélectionner les peptides. En effet, (cf vos cours de sciences physiques) le peptide qui résonne à la même fréquence que le quadri-pôle ne sera pas détruit et sera le seul à passer vers l'analyseur.

Ce peptide isolé passe alors dans une cellule de collision. Cette cellule est remplie d'hélium. Le peptide qui entre à très grande vitesse va se casser au grès des chocs avec les molécules d'He. Ensuite, les différents morceaux de peptides passent dans l'analyseur de temps de vol pour être séparés selon leur taille. Enfin, ils terminent leur course sur l'analyseur.

Identification des protéines

RESULTATS en MALDI-TOF



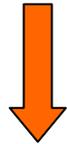
Spectre de masses



Les différents pics donnent la masse de chacun des peptides.

RESULTATS en MALDI-TOF MS

Empreinte de masses peptidiques



LISTE DE MASSE DE PEPTIDES



Comparaison avec banques de données

1066.5475
1148.6148
1203.7216
1219.7506
1235.2233
1282.7394
1316.7968
1442.8063
1446.8202
1482.8533
1526.8043
1651.9280
1713.8398
1764.9746
1824.95.95
1962.9613
1937.1038
2045.1148
2226.1487
2435.2555
2910.3334
2914.3803
3094.5123
3110.3581
3346.5564

On obtient donc une liste de masse des peptides constituant la protéine. Comme l'enzyme coupe toujours de la même manière, on peut comparer avec les listes de masses qui sont dans les banques.

Identification des protéines

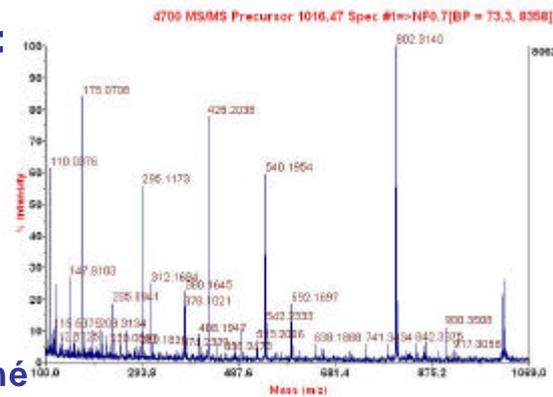
Résultats en QTOF



Liste des masses de fragmentation du peptide sélectionné

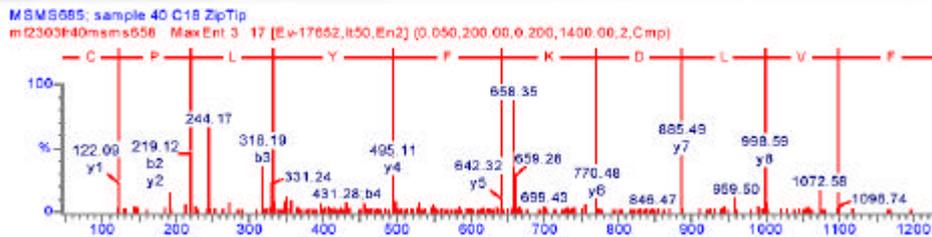


Comparaison avec banques de données



En Q-tof on peut, si on ne met pas en marche dès le début le quadri-pôle, laisser passer tous les peptides, dans ce cas les résultats sont équivalents à un MALDI-TOF.

Identification des protéines



Résultats en QTOF MS/MS



Reconstitution d'une séquence partielle



FVLDKFYLP

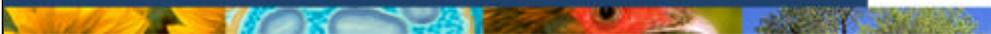
Comparaison avec banques de données



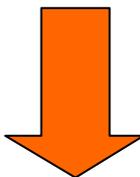
Si le quadri-pôle est en marche, on a alors sélection des peptides un à un. Le séquençage est réalisé par comparaison. On connaît la masse de chaque acide-aminé précisément. On peut donc par soustraction entre les différentes masses savoir quel acide-aminé manque et pas à pas reconstruire la séquence.

Remarque : on obtient rarement la séquence pleine longueur, on a plutôt des séquences de 15 à 20 acides-aminées à la suite, on parle de séquences TAG.

Recherche dans les bases de données protéiques



**Soumettre les données obtenues
en MS ou MS/MS aux bases de
données existantes.**



Identification du candidat.

Types de banques existantes

▪ Séquences protéiques, fonction, 3D,...

- Swissprot
- TrEMBL
- NCBI
- Uniprot

▪ Séquences nucléotidiques ou EST

- NCBI (Genbank)
- EBI (EMBL)

Ces différentes banques contiennent des informations de différentes catégories. Des séquences, mais aussi des informations sur la structure, la fonction des protéines. Il est aussi possible de travailler sur les banques de séquences ADN qui sont traduites en protéines.

Types d'interface pour l'interrogation des banques

- **MASCOT**
- **PEPIDENT**
- **PROTEINPROSPECTOR**
- **SRS**

Quelques types d'interfaces utilisés.

Il faut savoir que les différentes interfaces n'utilisent pas toutes les même algorithmes, il donc nécessaire de rechercher des concordances en utilisant plusieurs interfaces différentes car parfois les résultats obtenus sont différents. Si on obtient le même résultat avec toutes les interfaces, ce résultat est potentiellement plus sûr.

Matrix Science - Mascot - Peptide Mass Fingerprint - Mozilla

FORMULAIRE INTERROGATION

MASCOT Peptide Mass Fingerprint

Your name Email

Search title

Database

Taxonomy

Enzyme Allow up to missed cleavages

Fixed modifications: Biotin (N-term), Carbamidomethyl (C), Carbamyl (K), Carbamyl (N-term), Carboxymethyl (C)

Variable modifications: O18 (C-term), Oxidation (M), Oxidation (HW), PEO-Biotin (C), Phospho (ST)

Protein mass kDa Peptide tol. ppm

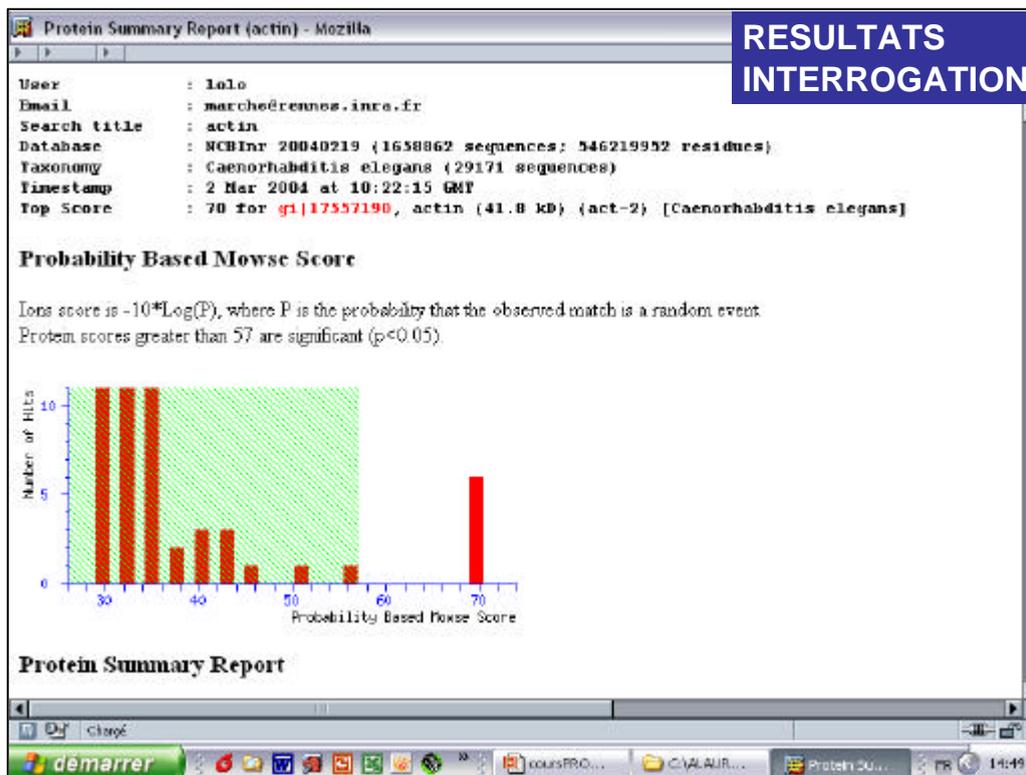
Mass values MH⁺ M_r Monoisotopic Average

Data file

Query 1011.691062
 NB Contents of this field are ignored if a data file is specified.
 1055.677798
 1082.131683
 1111.673235
 1126.610276
 1248.702865
 1314.853541
 4000.666004

Overview Report top hits

Une vue du type de formulaire qu'on doit renseigner pour interroger une banque. Différents éléments permettent d'affiner la recherche. On peut donner des informations d'ordre taxonomique sur l'échantillon. Le type d'enzyme de digestion utilisé ici la trypsine doit être indiqué. On peut renseigner les différentes modifications qu'on sait être intervenues sur l'échantillon protéique. En bas on indique la liste des masses des peptides obtenus après digestion de la protéine.



Les résultats sont classés par ordre de score. En effet, l'algorithme calcule une probabilité de ressemblance entre nos valeurs et celle des banques. Il donne donc une probabilité que notre échantillon soit la même protéine que la protéine de la banque. Ici la probabilité que notre protéine soit une actine de *Caenorhabditis elegans* est de 70 %.

RESULTATS INTERROGATION

Re-Search All Search Unmatched

1.	gi 17557190	Mass: 42093	Score: 70	Expect: 0.0032	Peptides matched: 11
	actin (41.8 kD) [act-3] [Caenorhabditis elegans]				
	gi 17568985	Mass: 42093	Score: 70	Expect: 0.0032	Peptides matched: 11
	actin (41.8 kD) [act-4] [Caenorhabditis elegans]				
	gi 283529	Mass: 42121	Score: 70	Expect: 0.0032	Peptides matched: 11
	actin 2 - Caenorhabditis elegans				
	gi 5620	Mass: 42024	Score: 69	Expect: 0.0038	Peptides matched: 11
	actin [Caenorhabditis elegans]				
	gi 17563820	Mass: 42111	Score: 69	Expect: 0.0038	Peptides matched: 11
	actin (41.8 kD) [act-3] [Caenorhabditis elegans]				
	gi 14278147	Mass: 41980	Score: 69	Expect: 0.004	Peptides matched: 11
	Chain 1, Crystal Structure Of Caenorhabditis Elegans Mg-Actin Complexed With Human Gel				
	gi 17568987	Mass: 37483	Score: 51	Expect: 0.22	Peptides matched: 9
	actin [act-4] [Caenorhabditis elegans]				
<hr/>					
2.	gi 7495955	Mass: 89255	Score: 55	Expect: 0.092	Peptides matched: 11
	hypothetical protein C14A4.2 - Caenorhabditis elegans				
	gi 17531775	Mass: 48356	Score: 42	Expect: 1.9	Peptides matched: 7
	mitochondrial ribosomal protein S22 (48.3 kD) (2K975) [Caenorhabditis elegans]				
<hr/>					
3.	gi 17509621	Mass: 71734	Score: 44	Expect: 1.1	Peptides matched: 9
	some crlo protein arginine n-methyltransferase (71.3 kD) (J0638) [Caenorhabditis elegans]				

Windows taskbar: démarrer, coursPRO..., C:\VALAUR..., Mascot Se..., 14:50

On obtient plus d'information ensuite sur notre probabilité. Ici la masse de notre échantillon, la probabilité et enfin le nombre de masse de peptides qui correspondent aux masses enregistrées dans la banque.

RESULTATS
INTERROGATION

Taxonomy: [Caenorhabditis elegans](#)

Links to retrieve other entries containing this sequence from NCBI Entrez:

[gi|3879474](#) from [Caenorhabditis elegans](#)

[gi|39591934](#) from [Caenorhabditis briggsae](#)

[gi|7507079](#) from [Caenorhabditis elegans](#)

[gi|21264389](#) from [Caenorhabditis elegans](#)

Fixed modifications: Carbamidomethyl (C)

Variable modifications: Oxidation (N)

Cleavage by Trypsin: cuts C-term side of KR unless next residue is P

Number of mass values searched: **40**

Number of mass values matched: **11**

Sequence Coverage: **33%**

Matched peptides shown in **Bold Red**

```

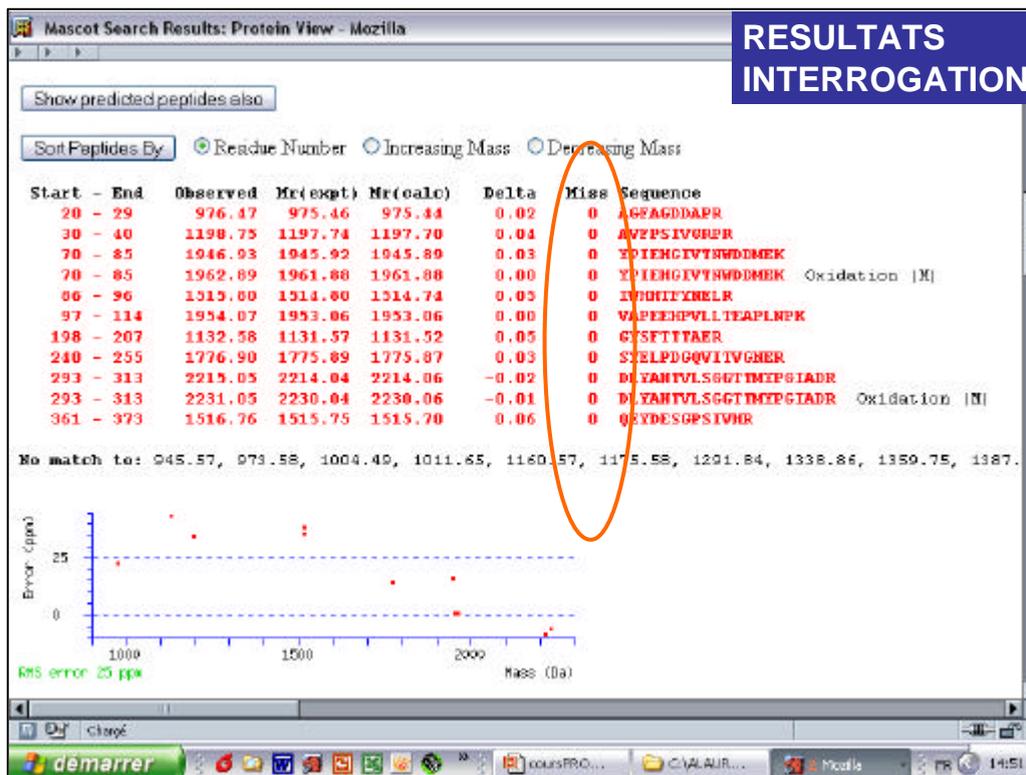
1  NCDDVVAALV VMNGSGMCEA GPAGDDAPRA VFPSIVGRPR HQGVHVGNGQ
31  KDSYVGDPAQ SERGILTLEY PIERGVYVW DDMKIVHHT FXNELEVAEK
101 EMFWLITLAP LMPKAREEM TQMFETFTNT PANTVAIQAV LSLYASGRTT
151  GIVLDSGGCV THTVPIYEGY ALPHAILRLD LAGRPLTDYL NKILTERGYS
201  FVYIAEREIV RPIREKLCYV ALDFEQENAT AASSSSLEKS YELPDGQVIT
251  VUNERFRCPE ALFQPSFLGM ESAGIHETSY NSINKCDIDI RKDLVANTVL
301  SGGYTNYEPI ADRNCKEITA LAPSTHEIKI LAPPERKTSV WIGGSILASL
351  STFQQEUIBK QEYDESGPSI VHRKCF

```

Show predicted peptides also

Sort Peptides By Residue Number Increasing Mass Decreasing Mass

Nous avons 11 peptides sur 48 qui sont correctes, et ces peptides sont assez bien dispersés sur la séquences. Ils couvrent 33 % de la séquence en question ce qui est un bon résultats.



La donnée misscleavage ou mauvaise coupure en français, nous indique qu'il n'y a pas d'erreur de coupure de la trypsine ceci renforce les résultats puisqu'il n'y a pas de faux positifs parmi les masses soumis à la banque qui résulteraient d'une mauvaise coupure de l'enzyme.

Précautions et validité des résultats

- **Nombre des peptides soumis (de 20 à 30)**
- **Pourcentage de recouvrement (doit être au dessus de 10 à 20 %)**
- **Résultats identiques avec les différentes interfaces car elles utilisent des algorithmes différents**

Comme ces interfaces donnent des résultats uniquement sous la forme de probabilité d'identité, il est très important de prendre le maximum de précaution avant de pouvoir affirmer que nous sommes bien en présence de telle ou telle protéine. Il faut donc s'astreindre à quelques règles qui sont : soumettre au minimum 20 à 30 peptides, ne retenir les résultats pour lesquels on obtient un pourcentage de recouvrement supérieur à 10-20%. Enfin, il faut réussir à obtenir des résultats différents avec diverses interfaces.

Si aucun résultat



▪ **MS/MS**

▪ **Séquençage de novo (Edman)**

Lorsque les données de MALDI-TOF ne donnent aucun résultats on peut utiliser alors la MS/MS qui nous donnera des informations sur les séquences, la recherche dans les bases de données sera alors plus performante.

Enfin pour des protéines qui sont nouvelles et absentes des banques il nécessaire de réaliser un séquençage d'Edman de novo afin d'obtenir la vraie séquence de la protéine.

Les domaines d'application



Médecine-Pharmacie

- **Diagnostic (Cancer prostate...)**
- **Nouvelle cible pharmaceutique**
- **Nouveau médicament**



Cancer, SIDA, AVC, Anémie, Allergie, Antibiotique

Les domaines où la protéomique trouve une application sont de plus en plus nombreux ces dernières années.

On peut donner à titre d'exemple la protéine P53 qui est un marqueur du cancer de la prostate a été trouvée après étude protéomique. Aujourd'hui on utilise un test qui permet de diagnostiquer le cancer de la prostate grâce à cette découverte.

(AVC : ACCIDENT VASCULAIRES CEREBRAUX)

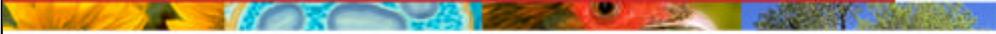
Biologie-Agronomie

- **protection des plantes**
- **produits vétérinaires**
- **Physiologie**
- **Parasitologie**
- **Sécurité alimentaire** (contaminants biologiques, OGM)

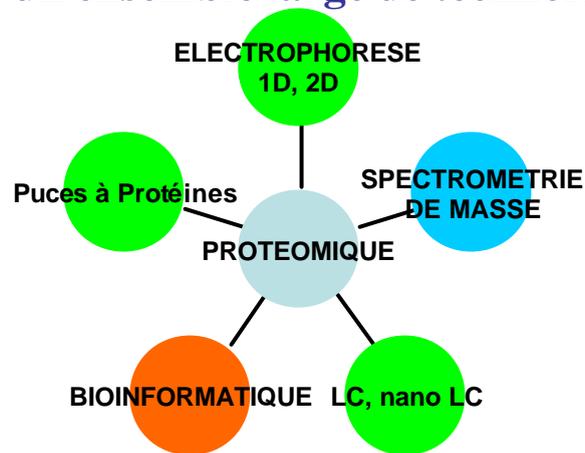
...

En protection des plantes on a pu identifier des protéines impliquées soit dans l'attaque d'agresseurs, soit dans la réponse des plantes. En sécurité alimentaire, certaines protéines servent là aussi de marqueurs qui permettent de certifier la sécurité ou la qualité des aliments.

L'analyse protéomique



L'analyse protéomique repose sur un ensemble large de technologies



L'analyse protéomique est donc une discipline qui fait appel à un grand nombre de techniques, de compétences et de savoir faire.

Des techniques séparatives, des techniques d'identification, de la bio-informatique aussi.

Production des données en quantité importante



- Echantillons biologiques
- Gels 2D ou chromatogrammes
- Données des spots issues de la MS et/ou MS/MS
- Résultats des interrogations des banques de données

La multiplication des études et de leur étendue favorise l'augmentation des données qui doivent être gérées.

L.I.M.S.

(Laboratory Information Management System Technology)



Nouveaux logiciels de gestions des données expérimentales

- Protein informatics
- ProDB
- CBiB (INRA-CNRS)

Aujourd'hui on peut travailler avec des LIMMS, qui sont en fait des logiciels qui vont gérer l'énorme quantité de données générées par les études. Vous pourrez trouver des LIMMS sur internet.

Références

Liens vers source bibliographique sur la protéomique :

- <http://www.swissproteomicsociety.org/digest/2008/issue56.html>

Site généraux sur la protéomique :

- <http://www.expasy.org/>
- <http://www.mnhn.fr/mnhn/bpy/2DMS/teomique.html>
- <http://pubs.acs.org/journals/jprobs/> Journal of proteome research
- <http://www.hupo.org/> Human of proteome organisation

Revue :

- Proteomics
- Electrophoresis
- Journal of proteome research

Remerciements à :

- Lionel Renault INRA de Rennes
- David Biron CNRS Clermont-Ferrand
- Emmanuelle Demey
(Plateau spectrométrie de masse de la génopole de Montpellier)