

# Protéomique quantitative : électrophorèse bidimensionnelle

DR Delphine Vincent  
Cours Master 2 / UE Protéomique  
Université Bordeaux 2  
IECB 6 octobre 2008

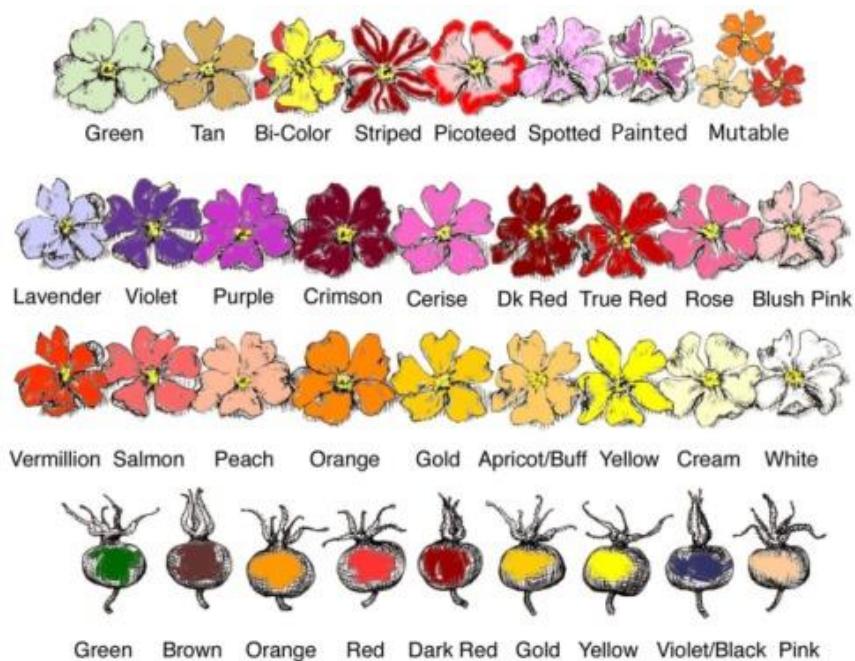
# Introduction

Pourquoi s'intéresser aux protéines et à leurs quantités ?

# Introduction: caractères qualitatifs/quantitatifs

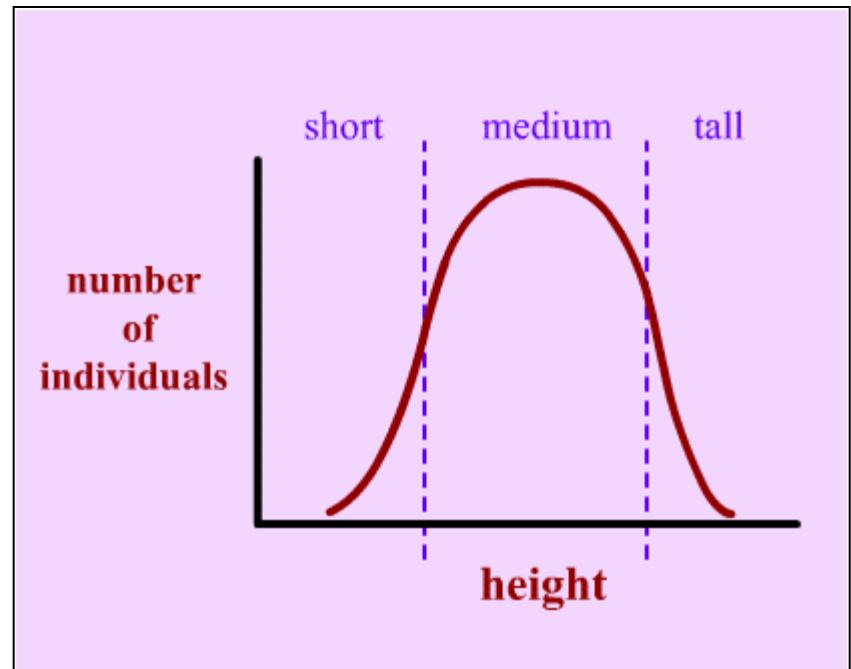
**Caractère qualitatif:**

caractère qui ne peut se mesurer.



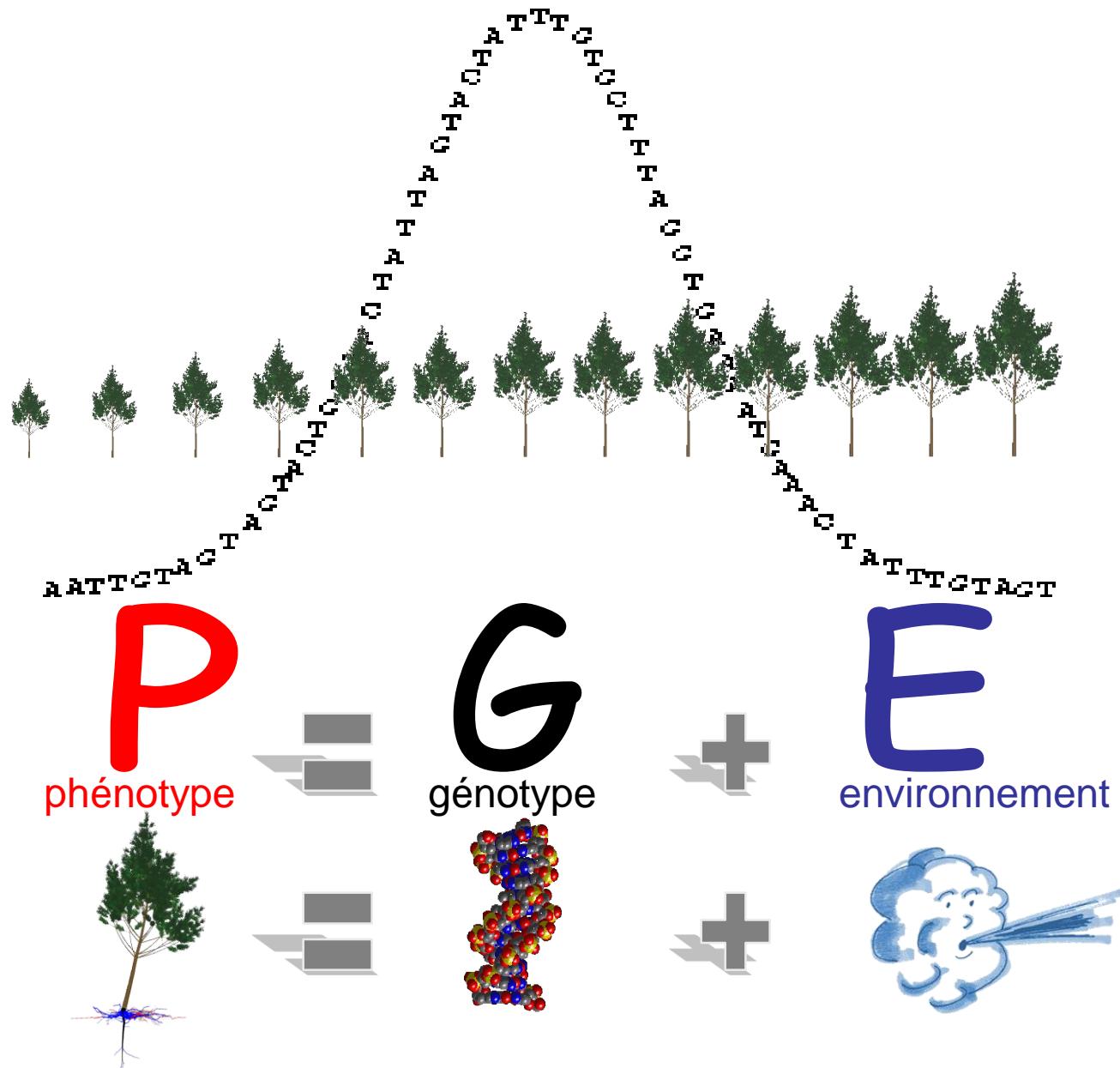
**Caractère quantitatif:**

caractère à variation continue qui peut se mesurer.



# Introduction: caractères qualitatifs/quantitatifs

La plupart des caractères d'importance économique et écologique sont de nature quantitative.



# Introduction: pourquoi s'intéresse-t-on aux changements (variations) ?

Comparer des états différents bien décrits aide à comprendre les phénomènes biologiques associés.

espèces



variétés/clones



tissus



traitement



développement



# Introduction: limite des études post-génomiques

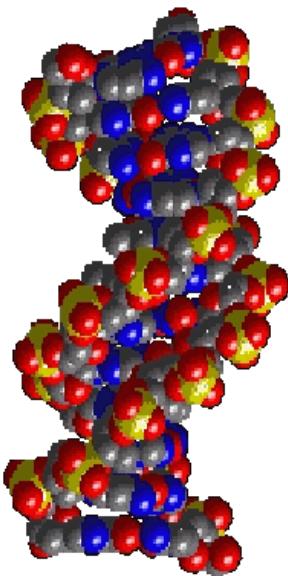
Animals	Plants	Protist	Fungi					
		Aquilegia 2.0 20-09-05		Arabidopsis 13.0 6-16-06		Barley 9.0 09-15-04		Beet 1.0 02-03-05
		Brassica napa 3.0 7-30-07		Chlamydomonas reinhardtii 5.0 09-28-04		Cocoa 1.0 05-20-04		Common Bean 1.0 21-7-05

Connaître la liste des gènes d'un organisme ne donne aucun renseignement sur leur fonction.

Les études post-génomiques (transcriptomique, protéomique, métabolomique) permettent d'identifier les gènes impliqués dans les phénomènes étudiés.

	Onion 1.0 09-04-03		Pepper 2.0 11-09-04		Petunia 1.0 09-29-04		Pine 6.0 06-19-05
	Poplar 3.0 06-19-06		Potato 11.0 06-19-06		Rice 17.0 06-20-06		Rye 3.0 12-22-03
	Sorghum bicolor 8.0 05-11-04		Soybean 12.0 09-20-04		Spruce 2.0 06-21-06		Sugarcane 2.1 10-13-04
	Sunflower 4.0 06-17-06		Tall Fescue 1.0 29-10-05		Tobacco 3.0 06-20-06		Tomato 11.0 06-21-06
	Triphysaria 1.0 08-30-05		Wheat 10.0 01-14-04				

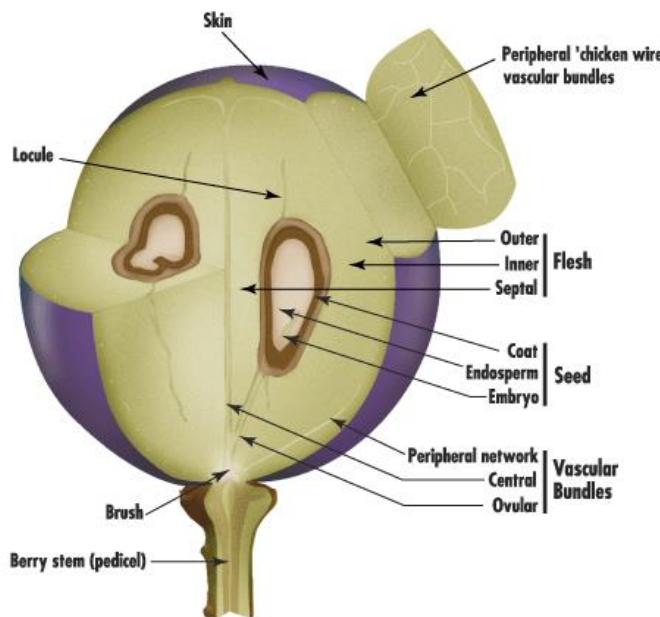
# Introduction: intérêt des études post-génomiques



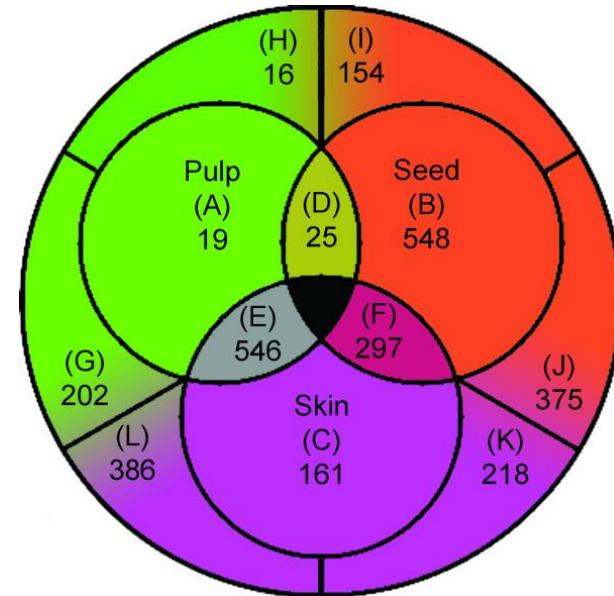
L'ADN n'offre qu'une vue statique du potentiel d'un organisme.  
Il ne permet pas de décrire les processus dynamiques d'un organisme vivant.  
Tous les gènes ne sont pas exprimés (transcrits puis traduits) en même temps par un organisme. Chaque individu, organe, tissu réagit différemment.

Etudier les niveaux de transcrits (ARN) et de protéines en réponse à un phénomène donne accès à la partie exprimée du génome nécessaire aux métabolismes de l'organisme impliqués dans ce phénomène.  
Ainsi les gènes-clé associés au phénomène considéré peuvent être identifiés.

Les études post-génomiques offrent en outre l'avantage considérable d'observer simultanément plusieurs dizaines de milliers (transcriptomique) ou plusieurs centaines voire milliers (protéomique) de produits de gènes.



Ex: expression de gène (ARN) au sein des différents tissus d'une baie de raisin.



# Introduction: quels sont les avantages à travailler avec les protéines ?

Observer les niveaux de transcrits n'informe en rien des régulations post-transcriptionnelles ou post-traductionnelles.

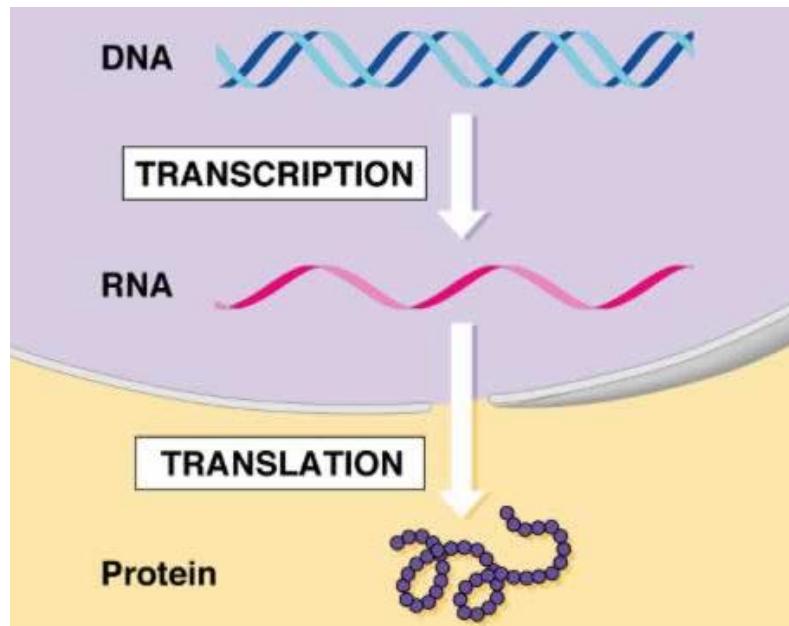
Chez les eucaryotes, 1 gène = 6-8 protéines (1 épissage alternatif + modifications post-traductionnelles)

De plus en plus d'études indiquent que les niveaux d'expression de protéines ne peuvent pas être prédits par les niveaux d'expression de transcrits.

→  $1 \text{ ARNm} \neq 1 \text{ protéine}$

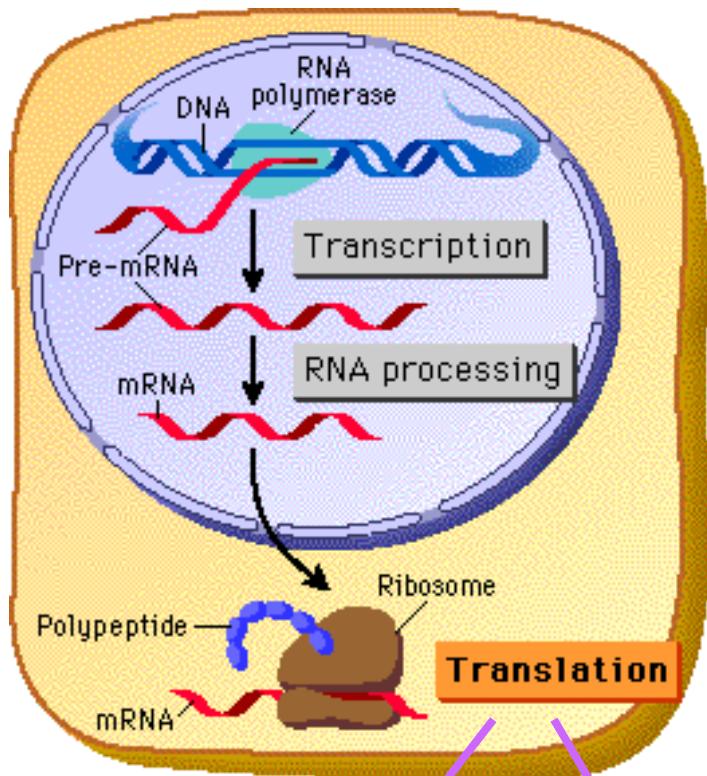
Beaucoup de caractéristiques des protéines (quantité, activité, localisation, interaction) ne peuvent pas être déduites de la séquence des gènes ou bien des niveaux de leur transcrits.

Les protéines représentent le niveau ultime de l'expression des gènes.



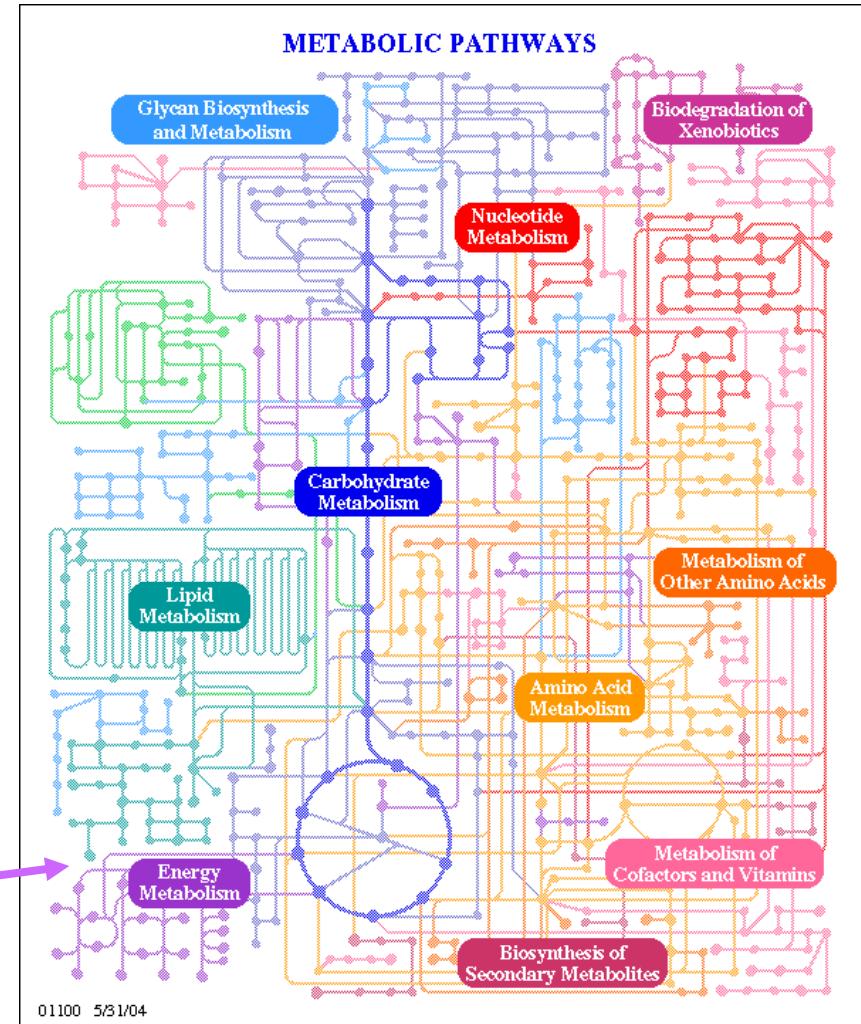
# Introduction: quels sont les avantages à travailler avec les protéines ?

Les protéines participent à tous les métabolismes biologiques et composent la plupart des structures.



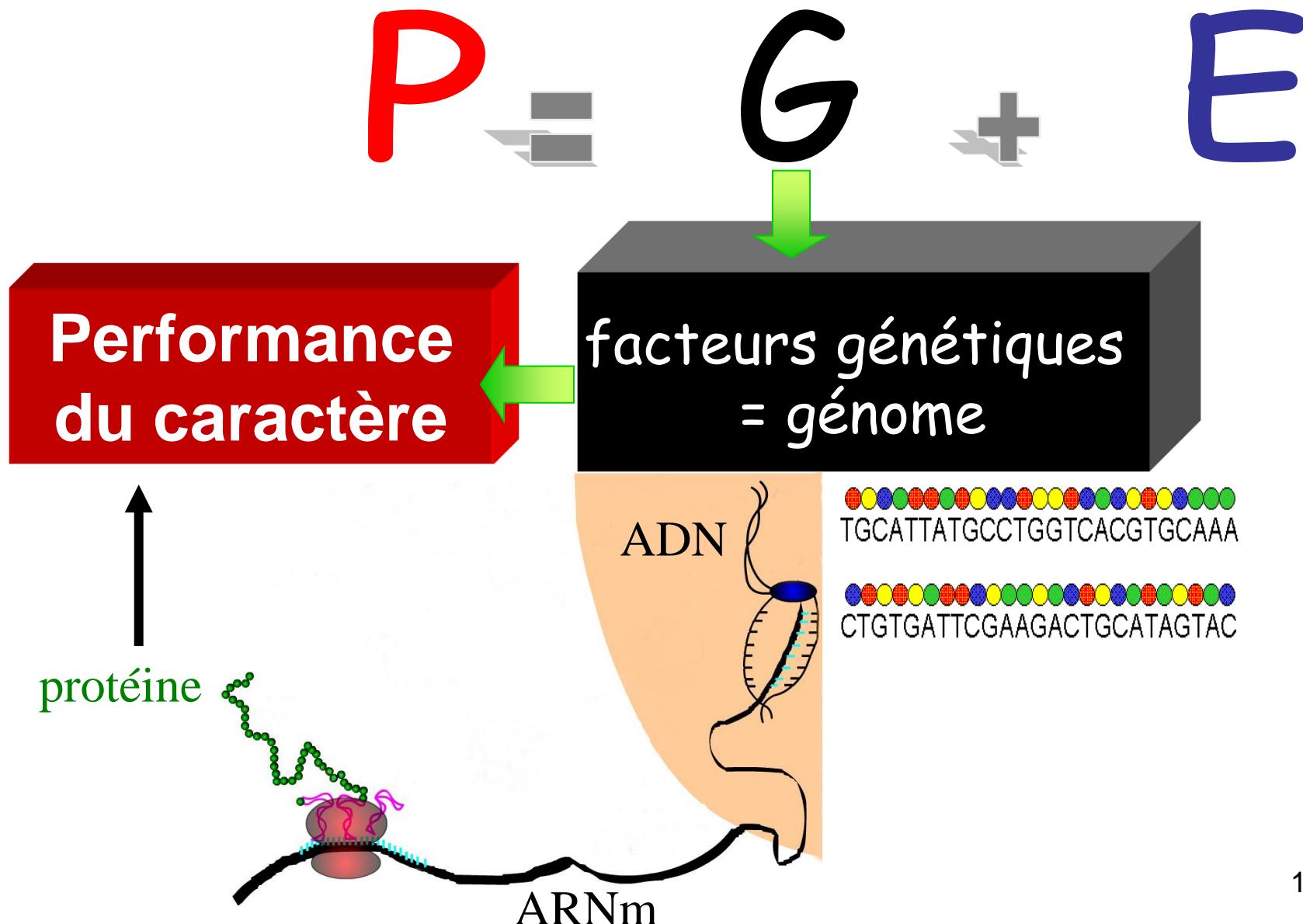
Protéines de structures

Enzymes



# Introduction: quels sont les avantages à travailler avec les protéines ?

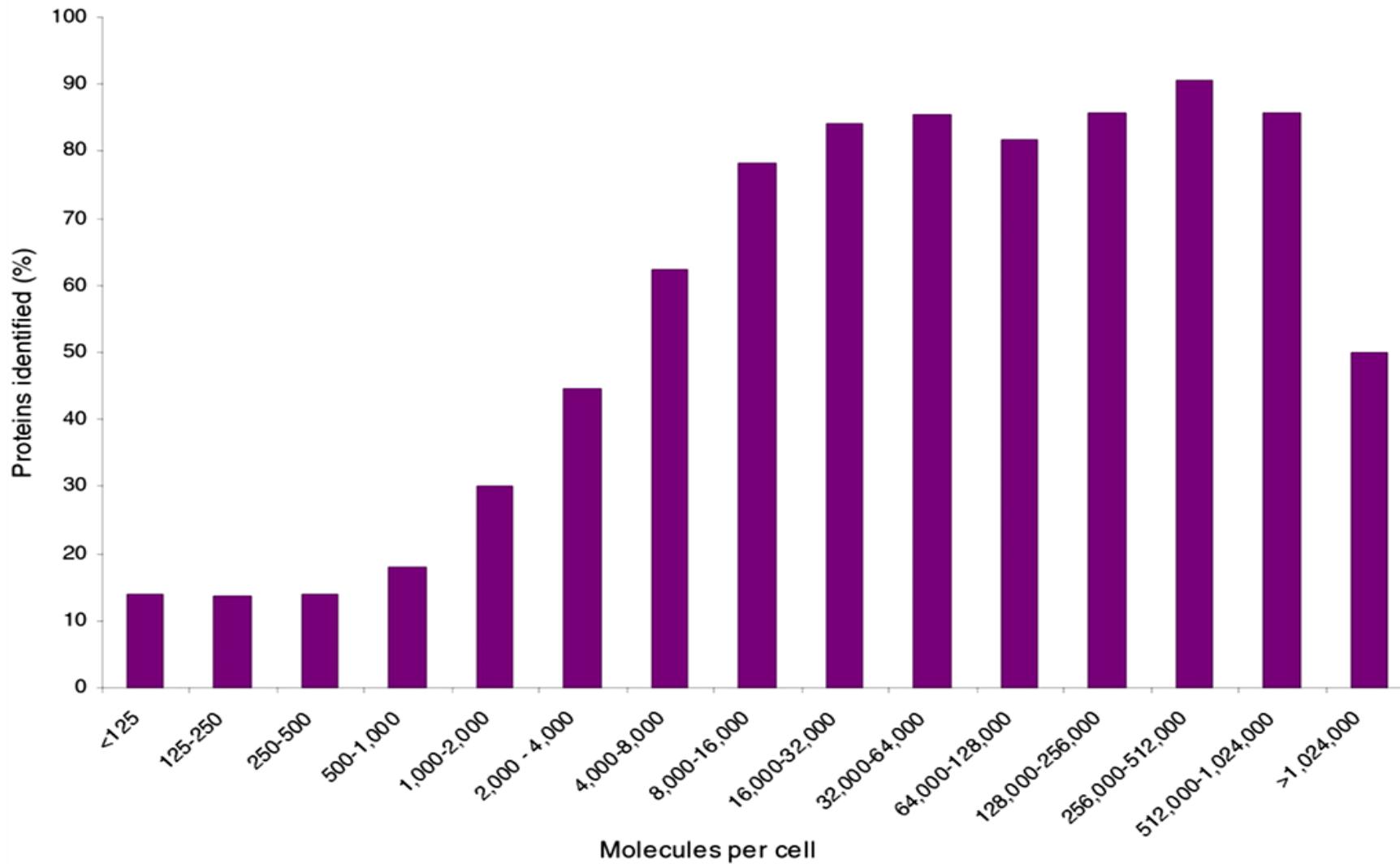
Les protéines étant responsables des processus biochimiques, les étudier permet de mieux appréhender le phénotype observé.



## Introduction: quels sont les défis à relever ?

Protéines: diversité infinie et ordre de grandeur considérable.

Ex: chez la levure (génome séquencé en 1996), il y a 6000 gènes qui potentiellement codent pour des millions de protéines différentes. Cependant, seules quelques milliers de protéines sont traduites à un instant donné. En outre, l'abondance maximale d'un transcrit s'étale de 1 à 100 molécules tandis que l'abondance des protéines s'étale de la dizaine à des millions.



# Introduction: quels sont les défis à relever ?

Diversité infinie (ex. les anticorps).

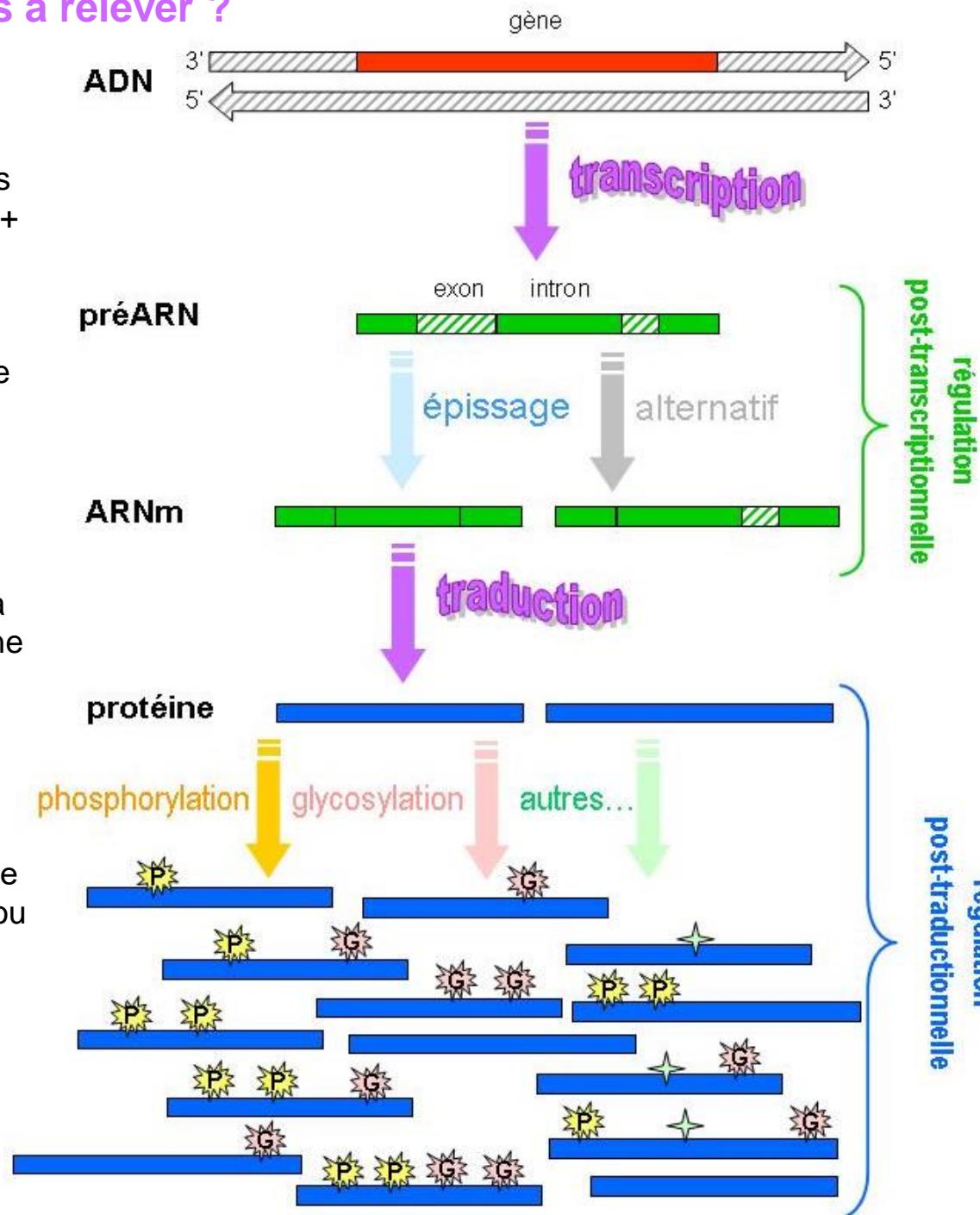
Origine de cette diversité: 20 acides aminés (4 bases azotées chez acides nucléiques) + régulation de l'expression des gènes.

Les outils de protéomique sont de débit moyen (100-1000) contrairement à ceux de transcriptomique (100000).

Les protéines ne peuvent pas être amplifiées comme l'ADN (PCR).

En raison de leur extrême diversité et de la considérable amplitude de leur niveaux, une technique unique ne suffit pas à observer toutes les protéines présentes à un instant donné dans un tissu.

Ex: fractionnement pour observer les protéines d'un organite ou purification d'une catégorie de protéine (phosphoprotéines) ou étude des interactions protéines-protéines.

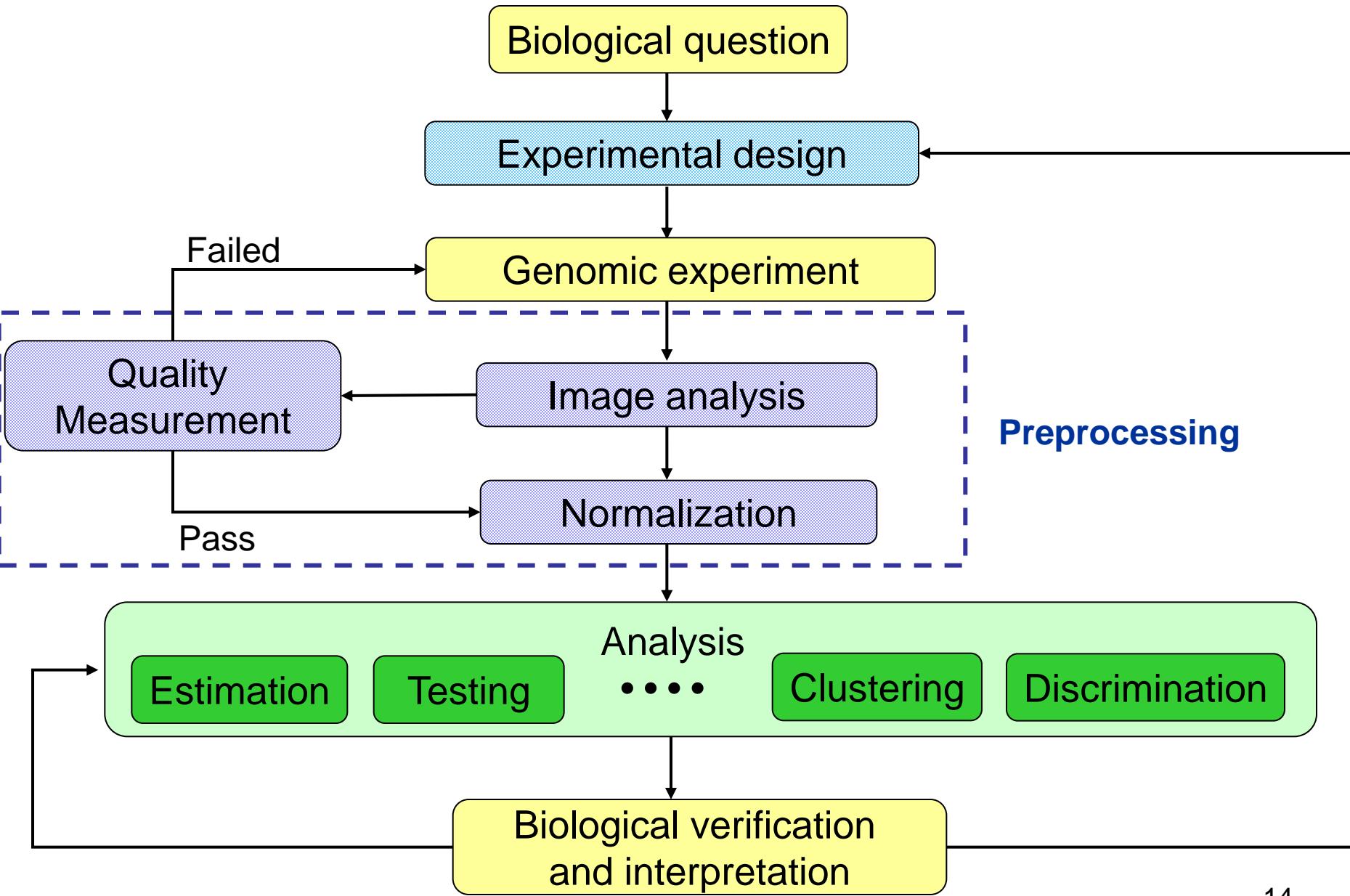


# Introduction: quelles sont les solutions pour observer les variations des quantités de protéines?

## Quelques techniques de protéomique:

	Quantitatif ? (variations)	Qualitatif ? (absence/présence)	Protéines connues ?
gel	SDS-PAGE	non	oui
	2-DE	oui	oui
	BN-PAGE	non	oui
	CE	oui	oui
MS	Protein chip	non	oui
	ICAT	oui	oui
	iTRAQ	oui	oui
	LC-MS	non	oui
	MudPIT	non	oui
	TAP	non	non

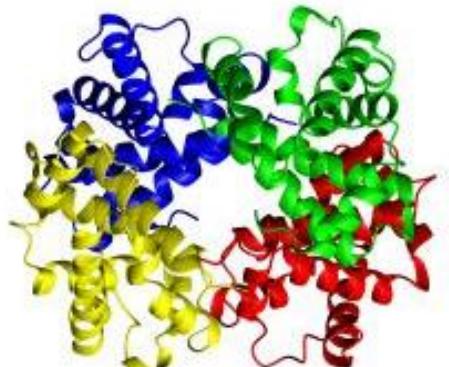
# Introduction: déroulement d'une étude post-génomique



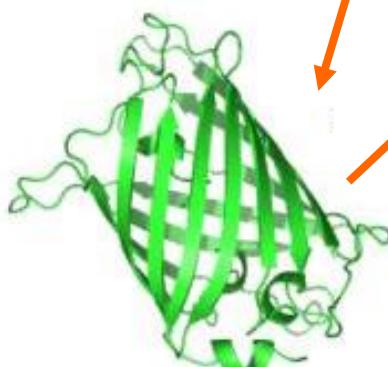
# Introduction: structure des protéines

Structure secondaire

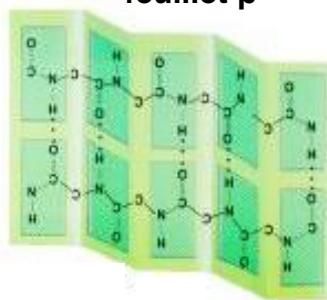
Structure quaternaire



Structure tertiaire



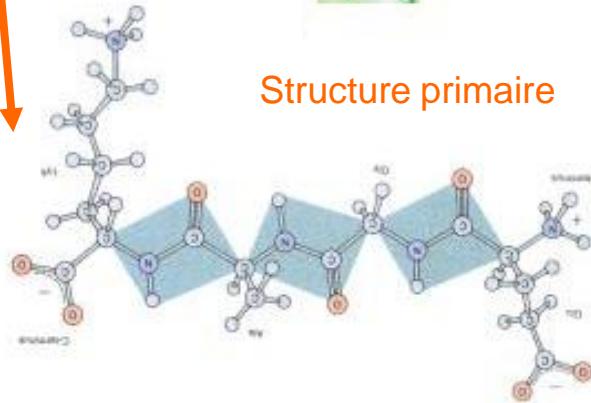
feuillet  $\beta$



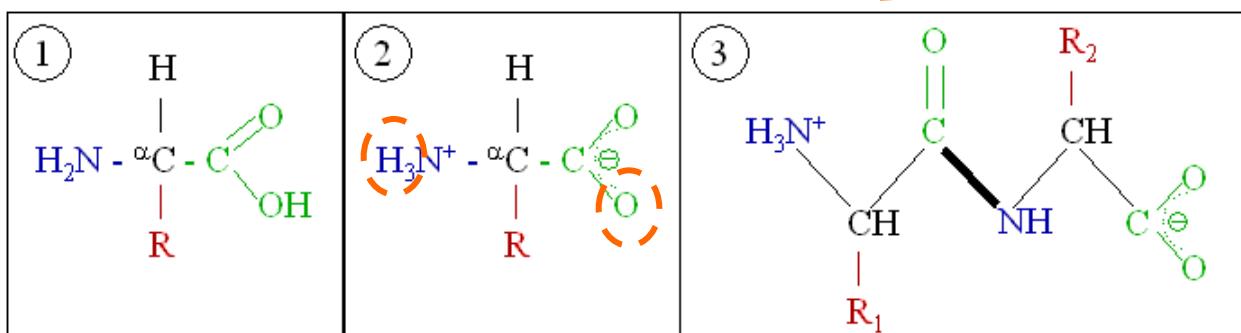
hélice  $\alpha$



Structure primaire

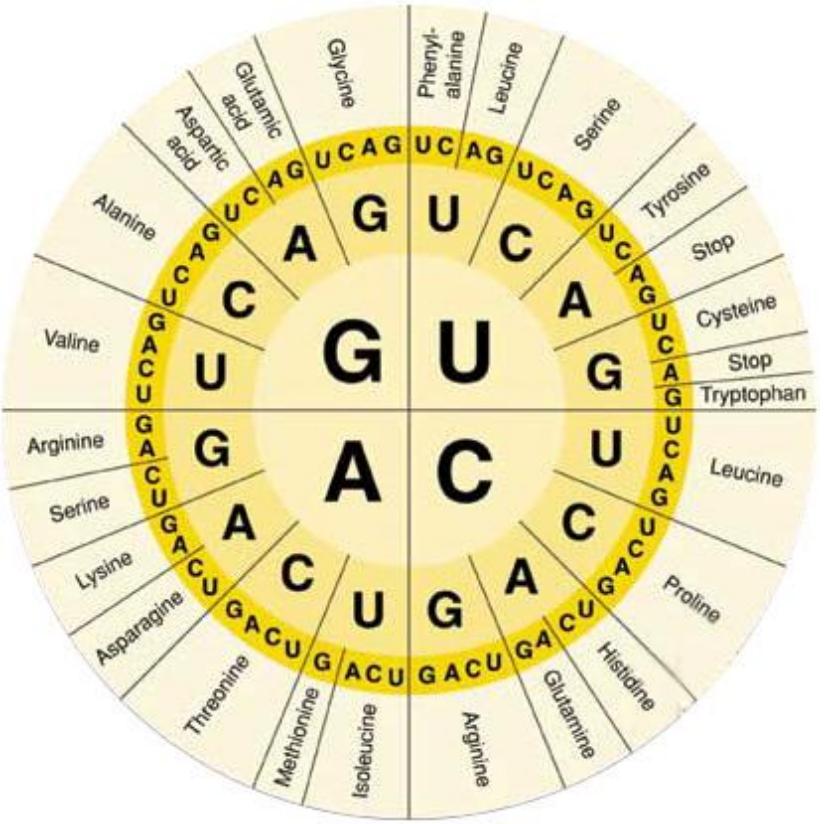


Elongation peptidique



# Introduction: acides aminés, unités de base des protéines

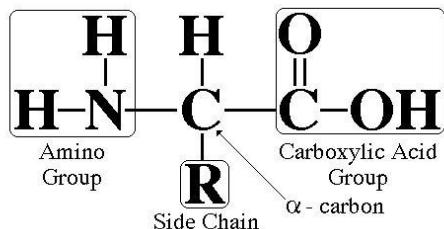
>P13645|K1C10\_HUMAN Keratin, type I cytoskeletal 10  
MSVRYSSSKHYSSRSGGGGGGGGCGGGGVSSLRISSSKGSLGGGFSS  
GGFSGGSFSRGSSGGCGFGSSGGYGGLGGFGGGSFRGSYGSSSFGGS  
YGGSGGGSFGGGSFGGGSGGGGGFGGGGGFGGGGGFGGDGGLLSG  
NEKVTMQLNDRLASYLDKVRALESNEYLEGKIKEWYEKGNSHQGEPR  
DYSKYYKTIDDLKNQILNLTTDNANILLQIDNARLAADDFLRKYENEVALRQSV  
EADINGLRRVLDETLTKADLEMQIESLTELAYLKKNHEEEMKDLRNVSTG  
DVNVEMNAAPGVDLTQLNNMRSQYEQLAEQNPKDAEAWFNEKSKELTTE  
IDNNIEQISSYKSEITLRRNVQALEIELQLSQLALKQSLEASLAETEGRYCVQL  
SQIQAQISALEEQLQQIRAEATECQNTEYQQLLDIKIRLENEIQTYRSLLEGEGS  
SGGGGRGGGSFGGGYGGGSSGGGGSSGGYGGGGHGGSSGGYGGGGSSGGGG  
GGGSSGGGYGGGSSSSGGHGGGSSSSGGHGGSSSGGYGGGGSSGGGGGG  
YGGGSSGGGSSSSGGYGGGSSSSGGHKSSSSGSGVGESSSKGPRY



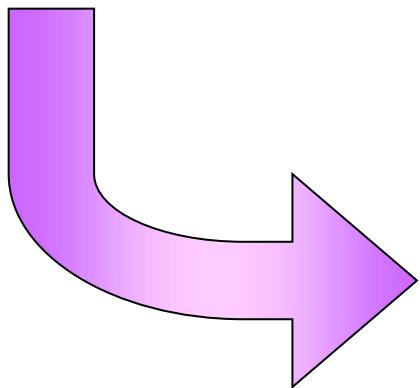
amino acid			mass	surface <sup>b</sup>	volume <sup>c</sup>	pK <sub>a</sub> <sup>d</sup>	pI <sup>e</sup>	solubility <sup>e</sup>	density <sup>e</sup>
Alanine	ALA	A	71.09	115	88.6	-	6.107	16.65	1.401
Arginine	ARG	R	156.19	225	173.4	~12	10.76	15	1.1
Aspartic Acid	ASP	D	115.09	150	111.1	4.5	2.98	0.778	1.66
Asparagine	ASN	N	114.11	160	114.1	-	-	3.53	1.54
Cysteine	CYS	C	103.15	135	108.5	9.1-9.5	5.02	very high	-
Glutamic Acid	GLU	E	129.12	190	138.4	4.6	3.08	0.864	1.460
Glutamine	GLN	Q	128.14	180	143.8	-	-	2.5	-
Glycine	GLY	G	57.05	75	60.1	-	6.064	24.99	1.607
Histidine	HIS	H	137.14	195	153.2	6.2	7.64	4.19	-
Isoleucine	ILE	I	113.16	175	166.7	-	6.038	4.117	-
Leucine	LEU	L	113.16	170	166.7	-	6.036	2.426	1.191
Lysine	LYS	K	128.17	200	168.6	10.4	9.47	very high	-
Methionine	MET	M	131.19	185	162.9	-	5.74	3.381	1.340
Phenylalanine	PHE	F	147.18	210	189.9	-	5.91	2.965	-
Proline	PRO	P	97.12	145	112.7	-	6.3	162.3	-
Serine	SER	S	87.08	115	89.0	-	5.68	5.023	1.537
Threonine	THR	T	101.11	140	116.1	-	-	very high	-
Tryptophan	TRP	W	186.12	255	227.8	-	5.88	1.136	-
Tyrosine	TYR	Y	163.18	230	193.6	9.7	5.63	0.0453	1.456
Valine	VAL	V	99.14	155	140.0	-	6.002	8.85	1.230

# Introduction: acides aminés, unités de base des protéines

## Amino Acid Structure



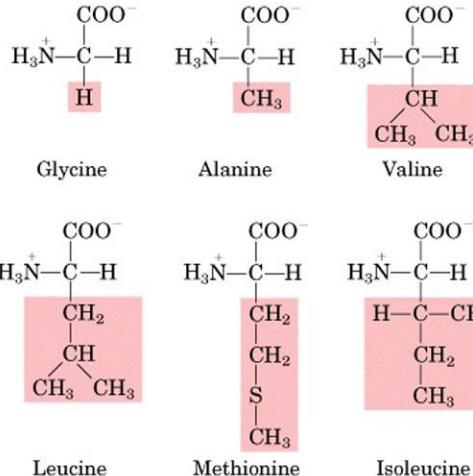
R: groupe variable



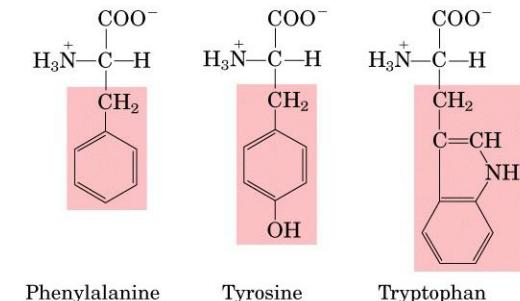
Charge d'une protéine = somme des charges des AA qui la composent.  
3 AA chargés  $\oplus$  (Lys, Arg, His)  
2 AA chargés  $\ominus$  (Asp, Glu)

## Twenty standard Amino Acids

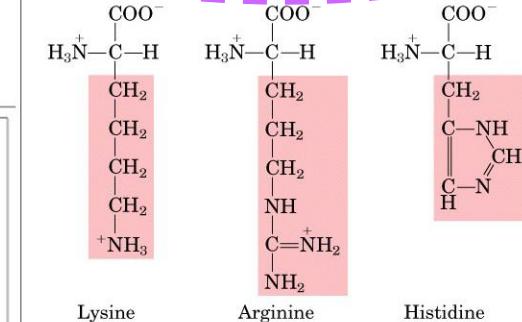
### Nonpolar, aliphatic R groups



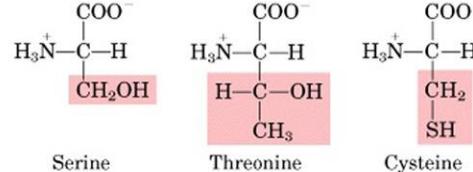
### Aromatic R groups



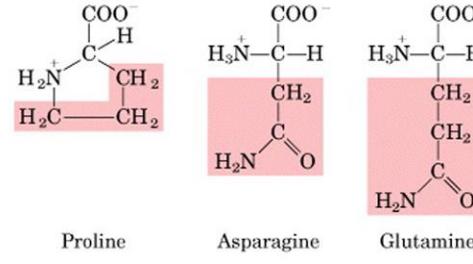
### Positively charged R groups



### Polar, uncharged R groups



### Negatively charged R groups



Aspartate

Glutamate

**2-DE**

**two-dimensional electrophoresis**

# Historique

# Historique: électrophorèse des protéines, technique qui remonte aux années 1930

Biochemical J. 1937 31(2): 313-7

## XLV. ELECTROPHORESIS OF SERUM GLOBULIN. I

By ARNE TISELIUS

*From the Institute of Physical Chemistry, University of Upsala*

(Received 2 December 1936)

Biochemical J. 1937 31(9): 1464-77

## CLXXXII. ELECTROPHORESIS OF SERUM GLOBULIN

### II. ELECTROPHORETIC ANALYSIS OF NORMAL AND IMMUNE SERA

By ARNE TISELIUS

*From the Institute of Physical Chemistry, University of Upsala*

(Received 1 July 1937)

Electrophorèse:  
Séparation en tube  
Selon un gradient de potentiel électrique  
Suivi par photographie UV

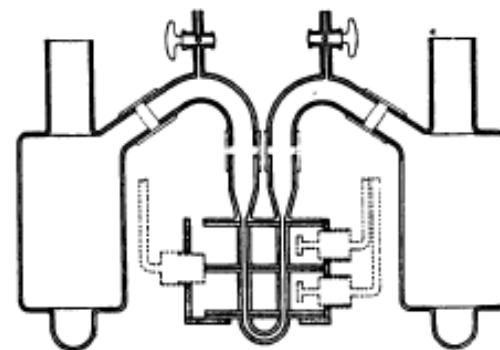


Fig. 1. Apparatus for electrophoretic analysis.



Fig. 2.

ovalbumine purifiée

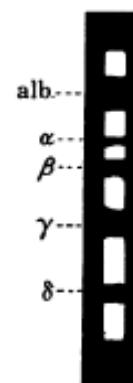


Fig. 3a.

sérum



Fig. 3b.

sérum albumine

# Historique: utilisation de polyacrylamide, invention de la SDS-PAGE en 1970

Nature, Vol. 227, No. 5259, pp. 680–685, August 15, 1970.

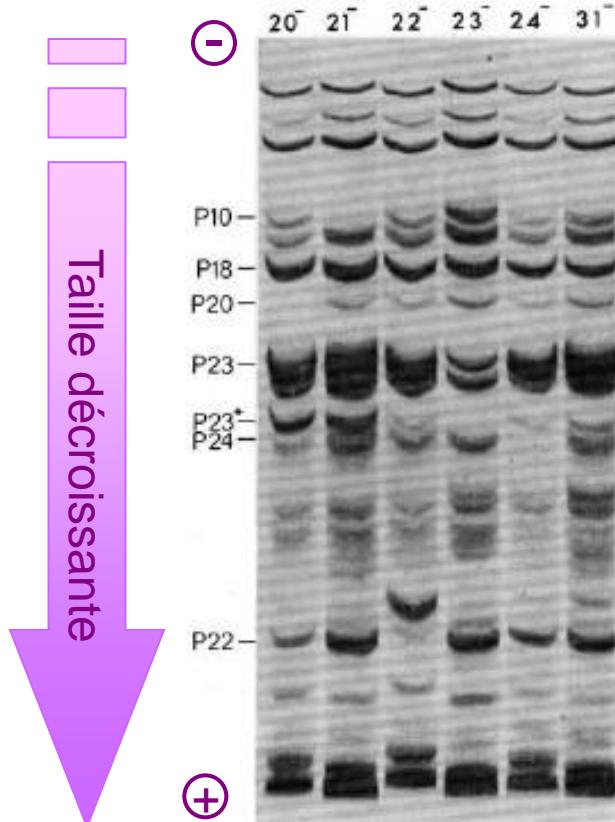
## Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4

by

U. K. LAEMMLI

MRC Laboratory of Molecular Biology,  
Hills Road, Cambridge

Using an improved method of gel electrophoresis, many hitherto unknown proteins have been found in bacteriophage T4 and some of these have been identified with specific gene products. Four major components of the head are cleaved during the process of assembly, apparently after the precursor proteins have assembled into some large intermediate structure.



### Tampon de Laemmli:

Tris-HCl	62.5 mM, pH 6.8
Glycine	10 %
SDS	3 %
2-ME	5 %

Ebouillanter l'échantillon en présence de sodium dodecyl sulphate (SDS, cation) qui dénature les structures tertiaire (autres que résultant d'un pont S-S) et secondaire des protéines.

Le 2-mercaptopropanoïne (2-ME) rompt les ponts disulfures, dénaturant davantage les protéines. Les sous-unités de protéines rendues négatives par le SDS migrent toutes vers l'anode (+) et sont séparées en fonction de leur taille (déplacement d'autant plus rapide que la molécule est petite). La polyacrylamide (10%) permet une excellente résolution des bandes.

# Historique: 2-DE, une technique inventée en 1975

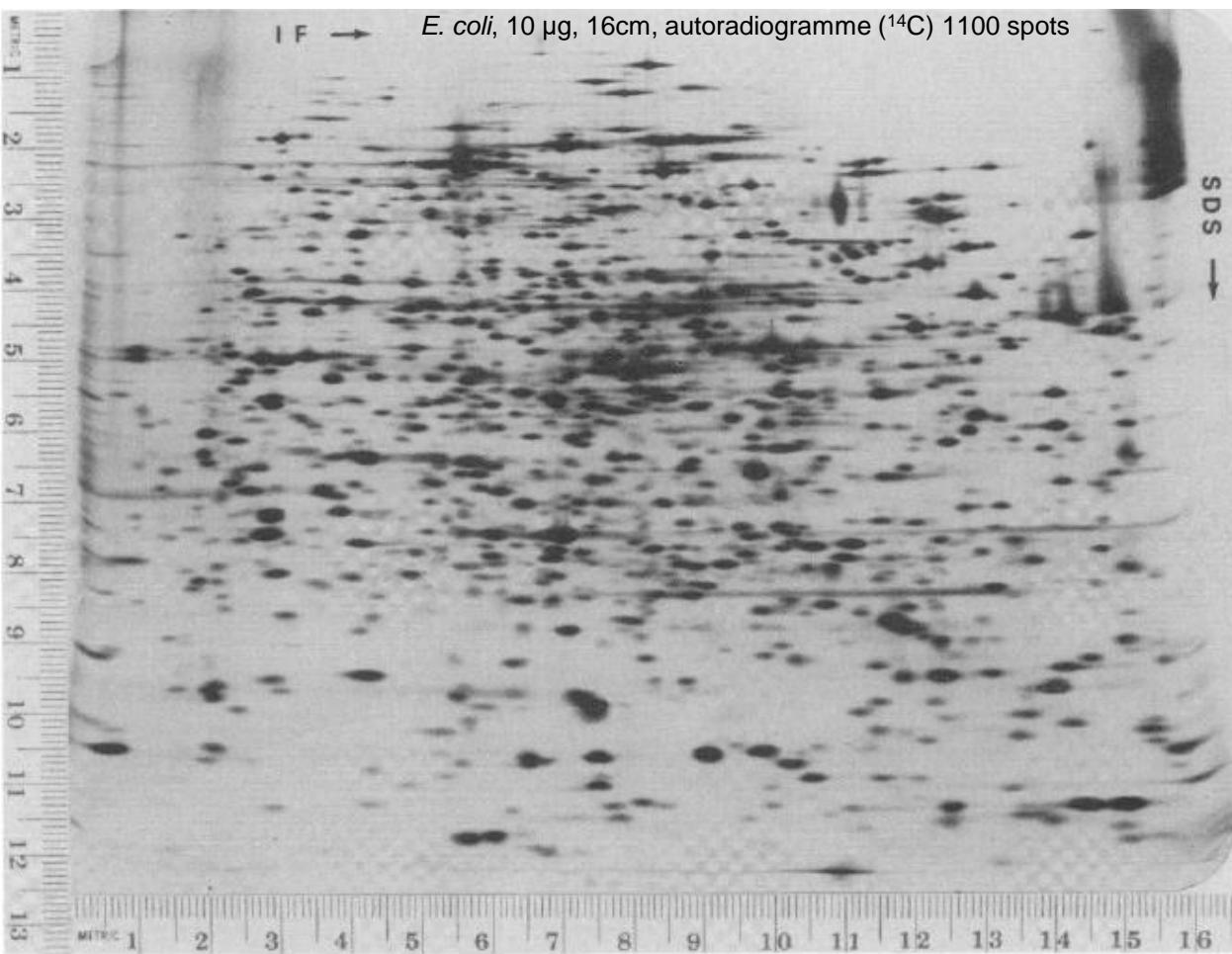
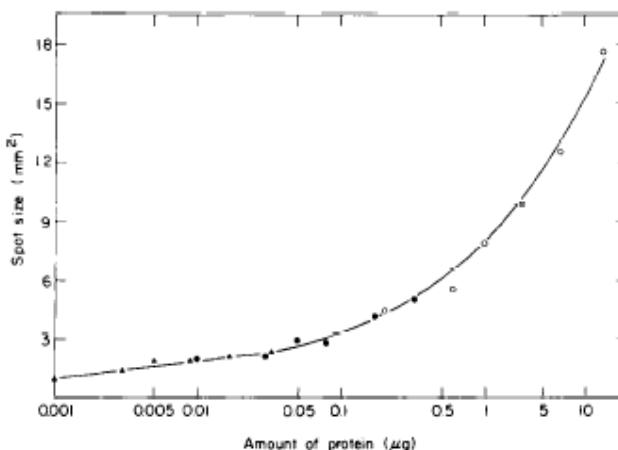
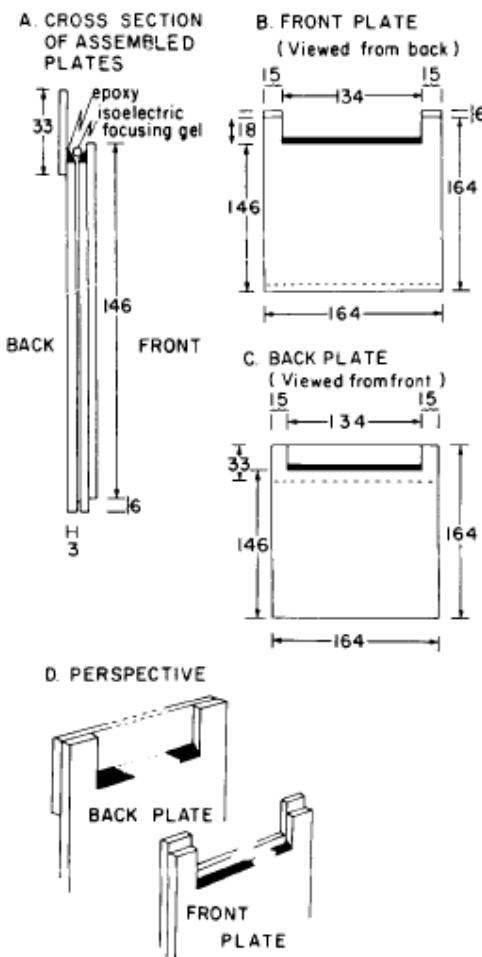
THE JOURNAL OF BIOLOGICAL CHEMISTRY  
Vol. 250, No. 10, Issue of May 25, pp. 4007-4021, 1975  
Printed in U.S.A.

## High Resolution Two-Dimensional Electrophoresis of Proteins\*

(Received for publication, September 5, 1974)

PATRICK H. O'FARRELL†

From the Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder,  
Colorado 80302



# Historique: IPG strip, amélioration de la reproductibilité, 1982

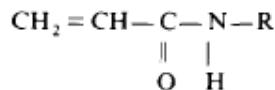
Journal of Biochemical and Biophysical Methods, 6 (1982) 317-339  
Elsevier Biomedical Press

317

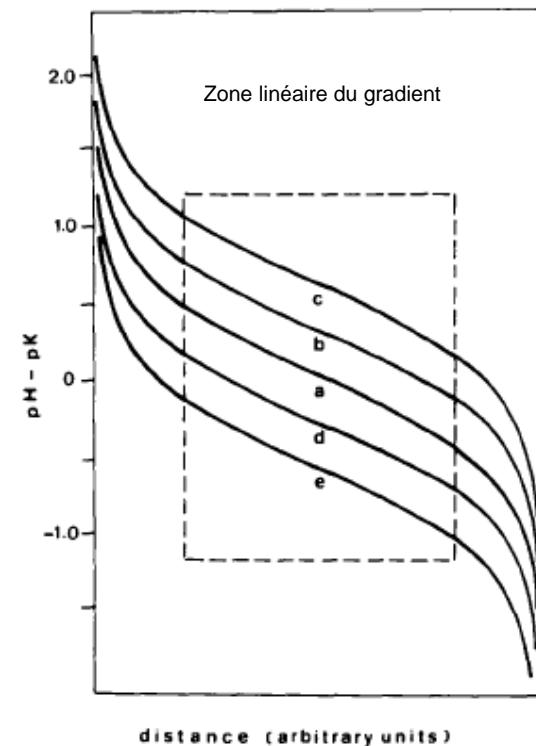
## Isoelectric focusing in immobilized pH gradients: Principle, methodology and some applications \*

Bengt Bjellqvist <sup>1</sup>, Kristina Ek <sup>1</sup>, Pier Giorgio Righetti <sup>2</sup>, Elisabetta  
Gianazza <sup>2</sup>, Angelika Görg <sup>3</sup>, Reiner Westermeier <sup>3</sup> and Wilhelm Postel <sup>3</sup>

<sup>1</sup> LKB Produkter AB, Box 305, S-161 26 Bromma, Sweden, <sup>2</sup> Department of Biochemistry, University of  
Milan, Via Celoria 2, Milan 20133, Italy, and <sup>3</sup> Technical University of Munich, D-8050  
Freising-Weihenstephan, F.R.G.



Immobilines: dérivés d'acrylamide pouvant être incorporés à des positions fixes le long du strip.  
R = contient un groupe carboxyl ou bien une amine tertiaire



	H <sub>2</sub> O		Polyacrylamide gel <sup>b</sup> T = 5%, C = 3%		Polyacrylamide gel <sup>b</sup> T = 5%, C = 3% glycerol 25% (w/v)		Physical state at room temperature	
	10°C	25°C	10°C	25°C	10°C	25°C		
<i>Acids with carboxyl as buffering group</i>								
Immobiline pK 3.6	3.57	3.58	-	-	3.68 ± 0.02	3.75 ± 0.02	solid	
Immobiline pK 4.4	4.39	4.39	4.30 ± 0.02	4.36 ± 0.02	4.40 ± 0.03	4.47 ± 0.03	solid	
Immobiline pK 4.6	4.60	4.61	4.51 ± 0.02	4.61 ± 0.02	4.61 ± 0.02	4.71 ± 0.03	solid	
<i>Bases with tertiary amine as buffering group</i>								
Immobiline pK 6.2	6.41	6.23	6.21 ± 0.05	6.15 ± 0.03	6.32 ± 0.08	6.24 ± 0.07	solid	
Immobiline pK 7.0	7.12	6.97	7.06 ± 0.07	6.96 ± 0.05	7.08 ± 0.07	6.95 ± 0.06	solid	
Immobiline pK 8.5	8.96	8.53	8.50 ± 0.06	8.38 ± 0.06	8.66 ± 0.09	8.45 ± 0.07	liquid	
Immobiline pK 9.3	9.64	9.28	9.59 ± 0.08	9.31 ± 0.07	9.57 ± 0.06	9.30 ± 0.05	liquid	

# Historique: Difference in Gel Electrophoresis (DiGE), 1997

1: [Electrophoresis](#). 1997 Oct;18(11):2071-7.

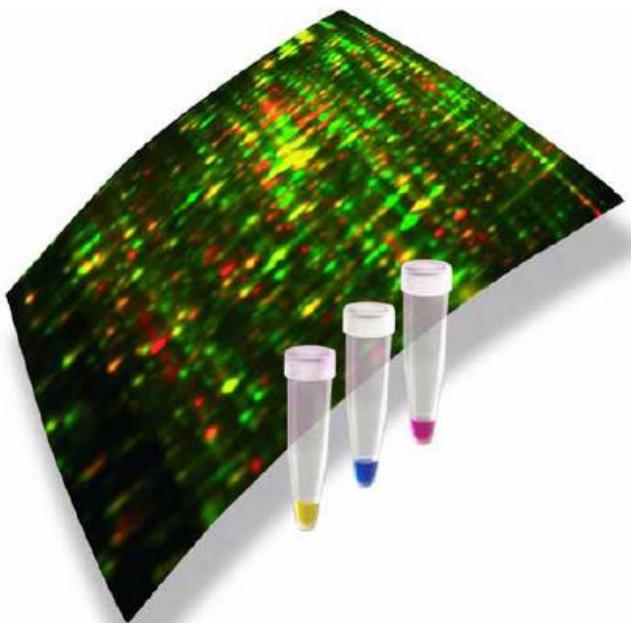
Difference gel electrophoresis: a single gel method for detecting changes in protein extracts.

Unlü M, Morgan ME, Minden JS.

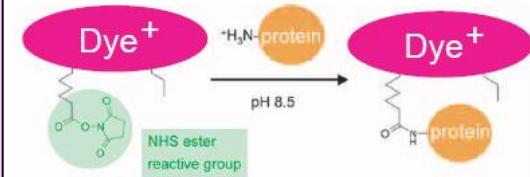
Department of Chemistry, Center for Light Microscope Imaging and Biotechnology, Carnegie Mellon University, Pittsburgh, PA, USA.

We describe a modification of two-dimensional (2-D) polyacrylamide gel electrophoresis that requires only a single gel to reproducibly detect differences between two protein samples. This was accomplished by fluorescently tagging the two samples with two different dyes, running them on the same 2-D gel, post-run fluorescence imaging of the gel into two images, and superimposing the images. The amine reactive dyes were designed to insure that proteins common to both samples have the same relative mobility regardless of the dye used to tag them. Thus, this technique, called difference gel electrophoresis (DIGE), circumvents the need to compare several 2-D gels. DIGE is reproducible, sensitive, and can detect an exogenous difference between two *Drosophila* embryo extracts at nanogram levels. Moreover, an inducible protein from *E. coli* was detected after 15 min of induction and identified using DIGE preparatively.

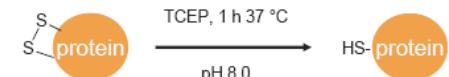
PMID: 9420172 [PubMed - indexed for MEDLINE]



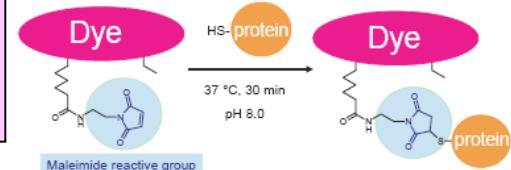
Minimal:  
Quantité non limitante d'échantillon  
50 µg de protéines  
3 marqueurs  
Marquage des lysines  
2% du spot marqué



Reduction step



Coupling step



## Historique: émergence de la protéomique, notion d'analyse globale

**Proteome:** entire complement of proteins expressed by a genome, cell, tissue or organism (Wilkins, 1994; Wasinger *et al.*, 1995). All the expressed proteins at a given time point under defined conditions.

**Proteomics:** large-scale study of proteins, particularly their structure and functions.

The screenshot shows the PubMed search interface. The top navigation bar includes links for All Databases, PubMed, Nucleotide, Protein, and Genome. The main search bar contains the text "PubMed" with dropdown options "for" and "Search". Below the search bar are buttons for Limits, Preview/Index, History, Clipboard, and Details. A "Display" dropdown is set to "AbstractPlus", with options for "Show 20" and "Sort By". The search results are displayed in a table with columns for document type, title, and authors. The first result is highlighted with a blue border. The title of the first result is "Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium." The authors listed are Wasinger VC, Cordwell SJ, Cerpa-Poljak A, Yan JX, Gooley AA, Wilkins MR, Duncan MW, Harris R, Williams KL, Humphrey-Smith I. The text below the title indicates the work was done at the Department of Microbiology, University of Sydney, Australia. The main text discusses the protein map of Mycoplasma genitalium, mentioning the use of two-dimensional gel electrophoresis, peptide-mass fingerprinting, and Edman degradation to identify proteins. It notes the identification of homologous genes in other organisms and the use of hierarchical analysis for mass screening. The text concludes with a note about the automation and robotics involved in initial screening.

NCBI PubMed A service of the National Library of Medicine and the National Institutes of Health www.pubmed.gov

All Databases PubMed Nucleotide Protein Genome

Search PubMed for | Go Clear

Limits Preview/Index History Clipboard Details

Display AbstractPlus Show 20 Sort By Send to

All: 1 Review: 0

1: [Electrophoresis](#). 1995 Jul;16(7):1090-4.

Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*.

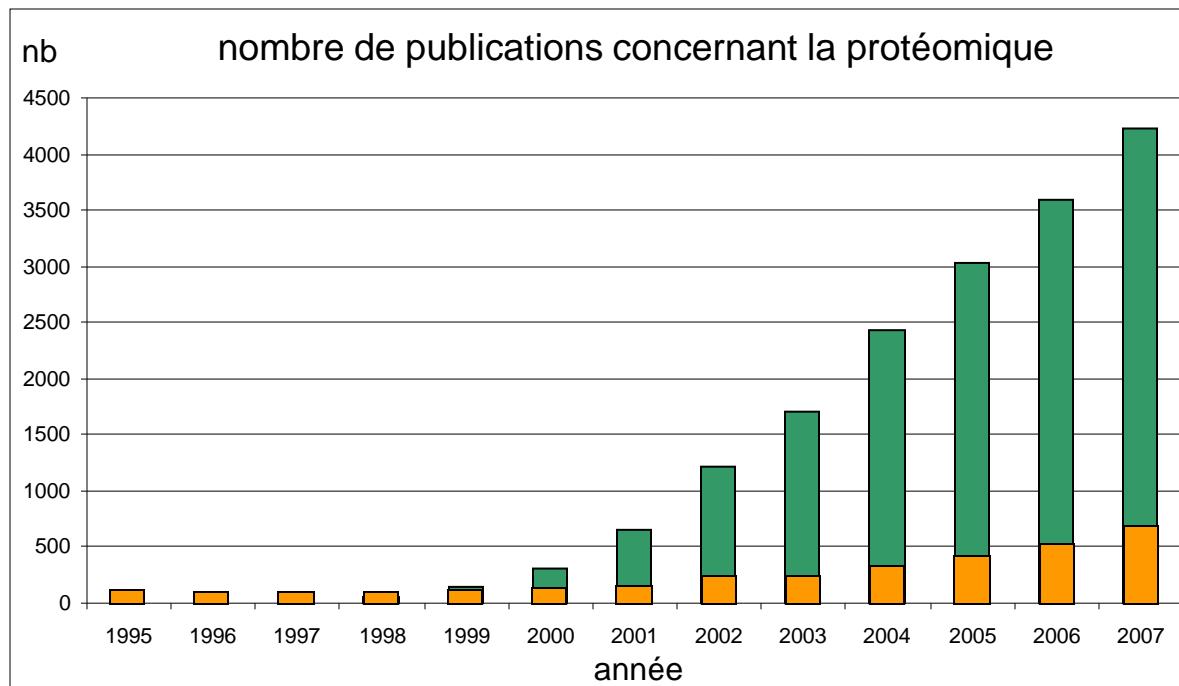
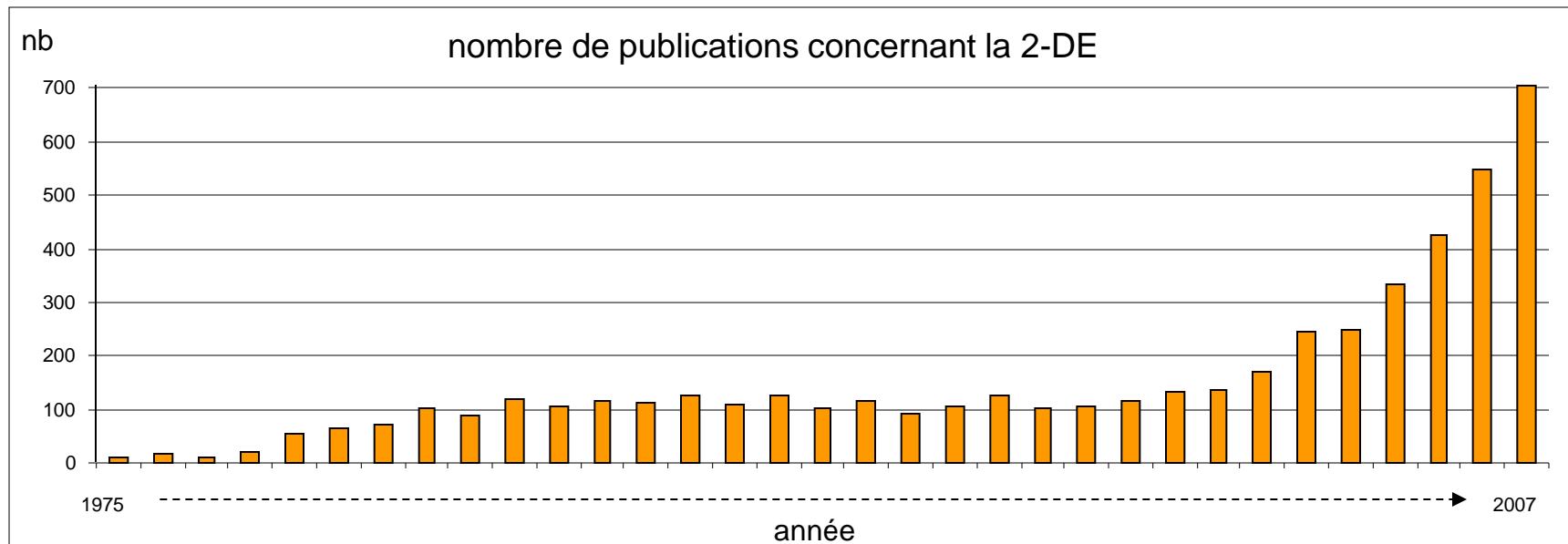
[Wasinger VC](#), [Cordwell SJ](#), [Cerpa-Poljak A](#), [Yan JX](#), [Gooley AA](#), [Wilkins MR](#), [Duncan MW](#), [Harris R](#), [Williams KL](#), [Humphrey-Smith I](#).

Department of Microbiology, University of Sydney, Australia.

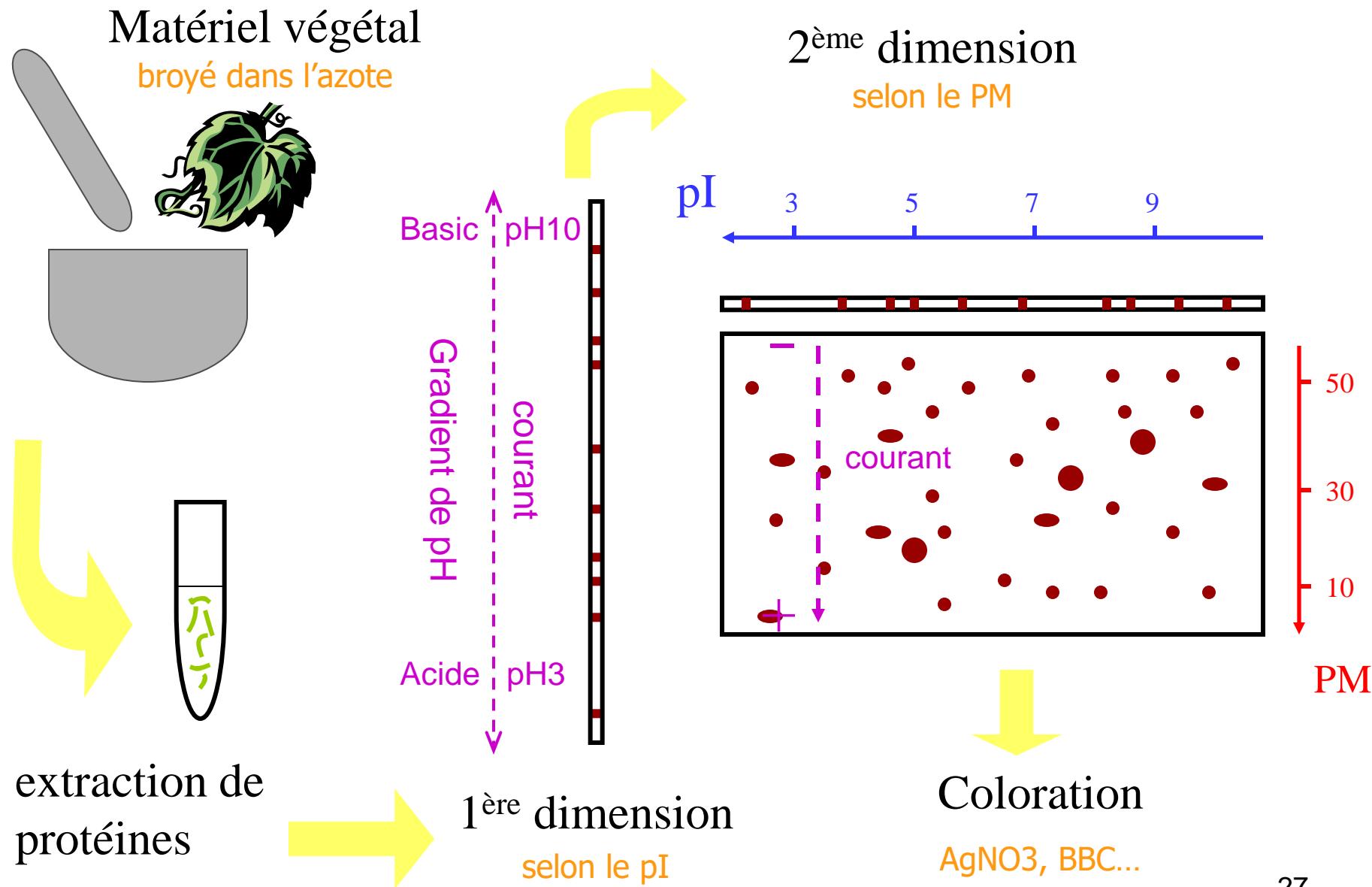
A protein map of the smallest known self-replicating organism, *Mycoplasma genitalium* (Class: Mollicutes), revealed a high proportion of acidic proteins. Amino acid composition was used to putatively identify, or provide unique parameters, for 50 gene products separated by two-dimensional gel electrophoresis. A further 19 proteins were subjected to peptide-mass fingerprinting using matrix-assisted laser desorption ionisation-time of flight (MALDI-TOF) mass spectrometry and 4 were subjected to N-terminal Edman degradation. The majority of *M. genitalium* proteins remain uncharacterised. However, the combined approach of amino acid analysis and peptide-mass fingerprinting allowed gene products to be linked to homologous genes in a variety of organisms. This has allowed proteins to be identified prior to detection of their respective genes via the *M. genitalium* sequencing initiative. The principle of 'hierarchical' analysis for the mass screening of proteins and the analysis of microbial genomes via their protein complement or 'proteome' is detailed. Here, characterisation of gene products depends upon the quickest and most economical technologies being employed initially, so as to determine if a large number of proteins are already present in both homologous and heterologous species databases. Initial screening, which lends itself to automation and robotics, can then be followed by more time and cost intensive procedures, when necessary.

PMID: 7498152 [PubMed - indexed for MEDLINE]

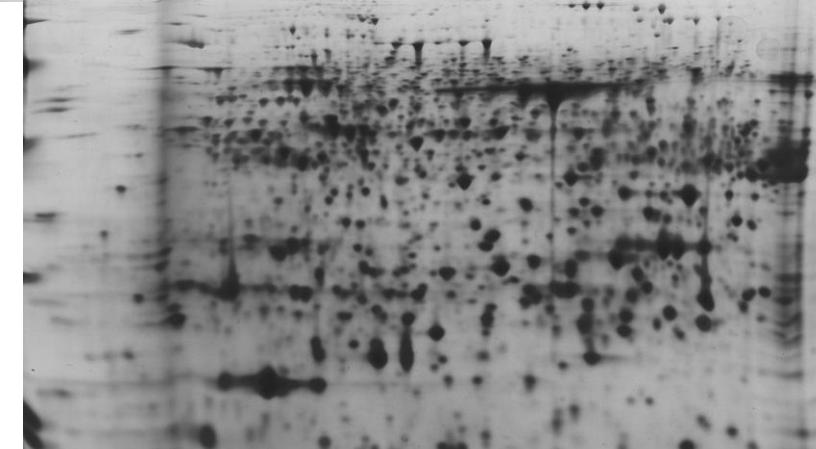
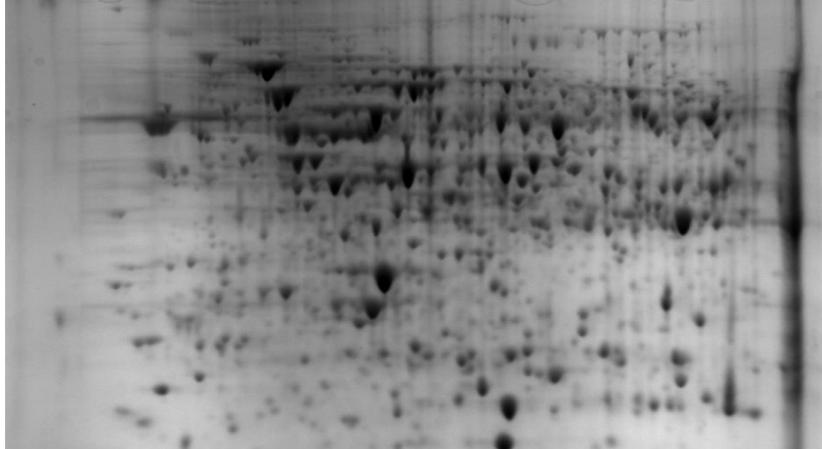
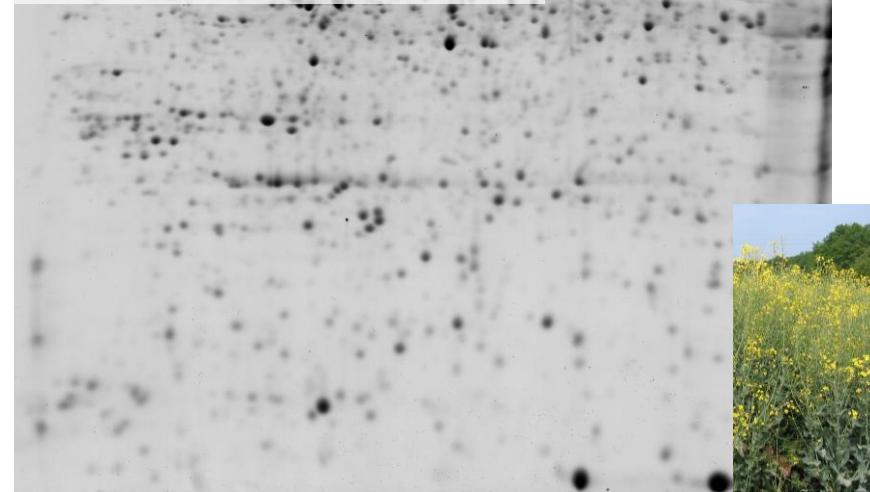
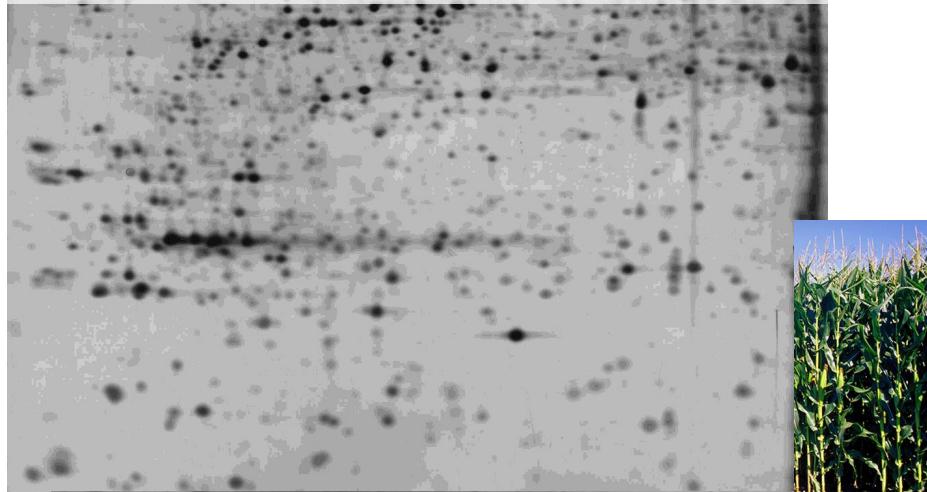
## **Historique: protéomique et 2-DE**



# Historique: principe de l'électrophorèse bidimensionnelle (2-DE)

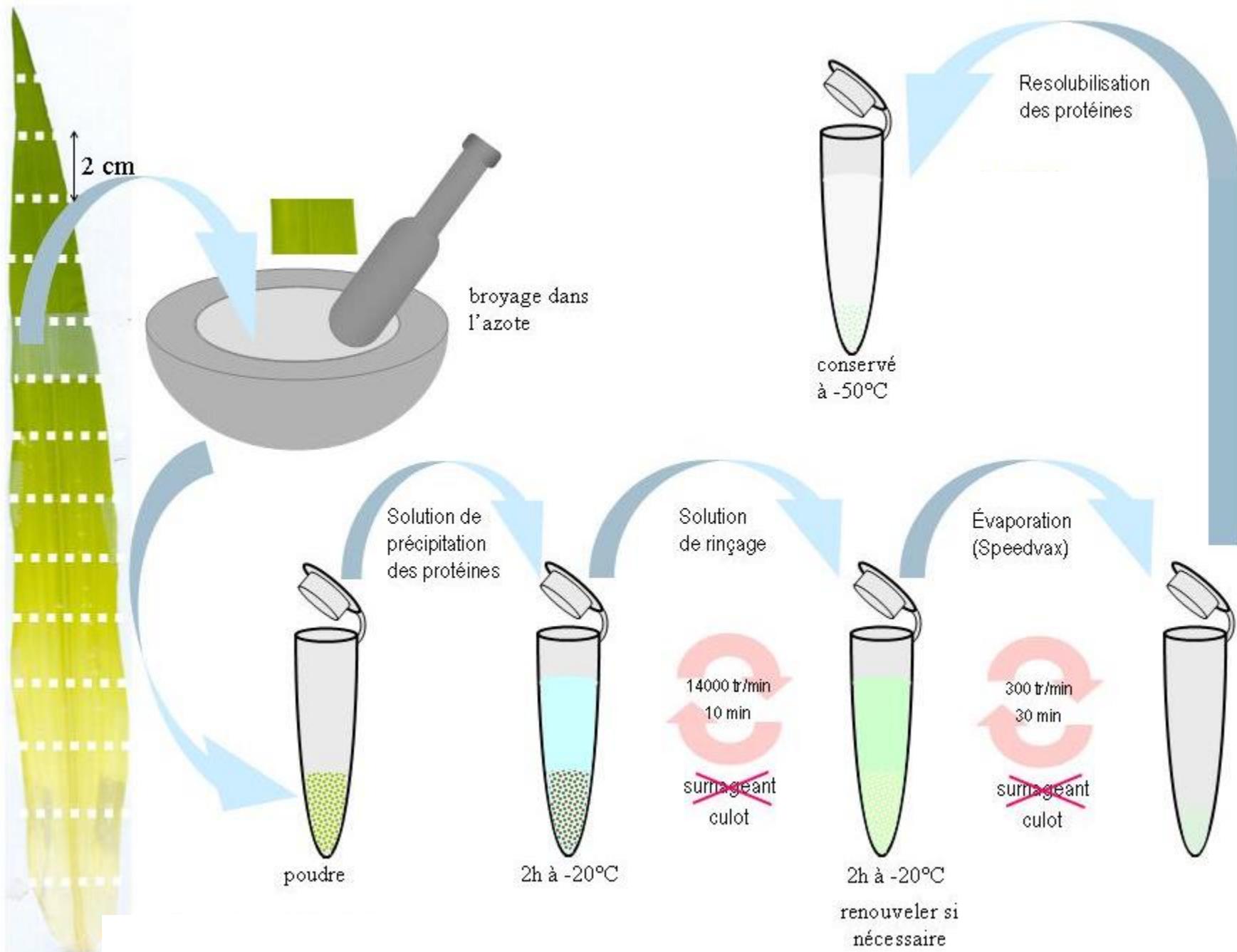


**Variation des profils en fonctions des espèces/tissus.  
même protocole d'extraction, mêmes conditions d'électrophorèse  
feuilles de différentes espèces**



# Extraction des protéines

# Extraction des protéines: principe



# Extraction des protéines: objectif

## Objectif:

Isoler le maximum de protéines en limitant les contaminations par les éléments non protéiques.

Comme l'on veut isoler l'ensemble des protéines et qu'elles sont très diverses, on ne peut qu'optimiser les conditions d'extraction afin de satisfaire au plus grand nombre de protéines, dans un temps raisonnable.

Il n'est pas question de purification de protéines.

Il n'est pas possible d'être exhaustif.

L'enjeu est donc d'extraire la majorité des protéines qui peuvent être résolues par 2-DE (ce qui suppose de connaître les limites de la technique analytique en aval).

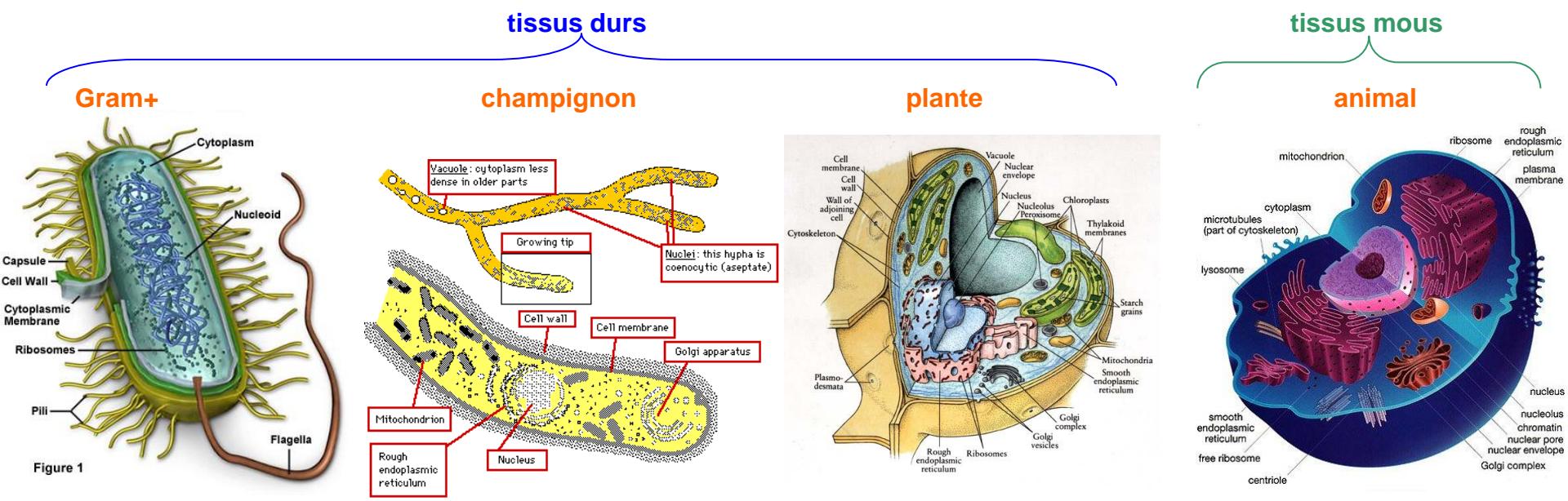
## Procédure:

- 1. pulvérisation des cellules
- 2. précipitation des protéines
- 3. solubilisation des protéines
- 4. protection contre les activités protease/oxydase
- 5. élimination des éléments non protéiques
  - acides nucléiques
  - lipides
  - sels, molécules ioniques/chargées
  - matériel insoluble (paroi cellulaire)

# Extraction des protéines: pulvérisation des cellules pour accéder aux protéines

La méthode employée va dépendre du tissu étudié:

présence d'une paroi cellulaire/enveloppe (bactérie Gram +, champignon, plantes) ou non (bactérie Gram -, animaux).



Lyse osmotique (sels) ou cycle de congélation/dégel pour protéines cytoplasmiques

Lyse chimique via des détergents (SDS) ou des agents chaotropes (urée) pour culture cellulaire

Homogénéisation en présence d'un tampon

Pulvérisation mécanique par pression (French press)

Sonication pour cellules avec parois

Lyse enzymatique (enzymes de dégradation des parois)

Pulvérisation mécanique par broyage (billes, sable ou azote liquide) pour tissus solides

tissus mous

tissus durs

# Extraction des protéines: précipitation pour concentrer les protéines et éliminer les éléments non protéiques

Quatre principes pour précipiter les protéines:

## 1- avec des sels (substance formée d'un acide et d'une base)

Si une faible concentration en sels favorise la solubilisation des protéines, une concentration trop élevée les précipite.

Efficacité variable des anions constituants les sels:

citrate > phosphate > sulphate > acetate

chloride > nitrate > thiocyanate

Sel couramment employé: ammonium acétate ou ammonium sulfate

## 2- variation de pH

A bas pH, une protéine présente une charge nette positive (gain de proton des groupes amines), à haut pH sa charge nette devient négative (perte de proton des groupes carboxyl); à son point isoélectrique (pl), une protéine a une charge nulle. N'interagissant plus avec son milieu, elle précipite.

L'acide trichloroacétique (TCA), très efficace, est souvent utilisé, cependant la résolubilisation qui s'ensuit est difficile.

## 3- avec des solvants

L'ajout d'un solvant réduit la constante diélectrique du milieu, conduisant à la précipitation des protéines. Cependant, les solvants présentent une affinité pour les parties hydrophobes des protéines, ce qui les dénature en même temps qu'elles précipitent. En outre, cela ne précipite pas toutes les protéines.

Solvants communément utilisés: acétone, éthanol, méthanol, chloroforme/phénol.

## 4- élimination de l'eau

Certaines molécules (dextran, PEG) limitent la quantité d'eau nécessaire à la solubilisation des protéines; les protéines s'agglomèrent et précipitent.

NB: une combinaison est possible et accentue le phénomène. Ex. TCA/acétone



Une fois précipités dans le culot, le surnageant est jeté. La précipitation permet ainsi d'éliminer des composés qui restent en solution dans le surnageant. Toutefois, cette étape conduit à des pertes de protéines (celles qui ne précipitent pas).

# Extraction des protéines: solubilisation des protéines et elles seulement !

La plupart des protéines doivent être dénaturées pour être résolubilisées.

Ce que doit contenir un tampon de résolubilisation de protéines:

## 1- agents chaotropes

Ils rompent les liaisons hydrogènes et les interactions hydrophobes à la fois entre et au sein des protéines. A forte concentration, ils détruisent la structure secondaire des protéines, les rendant à nouveau solubles.

Composés les plus utilisés (concentration à saturation): urée 7M en combinaison avec thiourée 2M (éviter  $t > 35^\circ\text{C}$ ). Neutre donc compatible avec IEF.

## 2- détergents

Ils rompent également les interactions hydrophobes à la fois entre et au sein des protéines. Ils améliorent la solubilisation des protéines peu solubles. Ils sont multiples, anioniques, cationiques, neutres ou zwitterioniques (amphotères).

Composés les plus utilisés (1-4%): triton X-100, CHAPS, ASB-14, SB3-10. SDS (cation) doit être éliminé avant IEF.

## 3- agents réducteurs

Ils clivent les ponts disulfures au sein des protéines et entre les sous-unités (structures tertiaire et secondaire perdues). Deux grands types: composés sulfhydryls (DTT, DTE, 2-ME) et phosphines (TBP, TCEP). La réduction des liaisons S-S peut être stabilisée par alkylation (iodoacétamide), limitant ainsi les agrégations et favorisant la solubilisation.

## 4- tampons, sels et ampholytes

Le pH et la force ionique influencent la solubilité des protéines. Nombre de protéines deviennent solubles à haut pH d'où l'addition de Tris base. Cependant, le pH optimal de solubilisation varie d'une protéine à l'autre d'où l'intérêt d'utiliser un tampon. L'ajout de sel (150 mM NaCl, incompatible avec IEF) solubilise les protéines par interactions ioniques. Les ampholytes augmentent le pouvoir tampon et la force ionique de la solution. La sonication (chauffe!) peut favoriser la résolubilisation.



Une fois en solution, les protéines peuvent être séparées par 2-DE. Encore une fois, toutes les protéines ne sont pas résolubilisées. Cette étape conduit à d'autres pertes qui s'ajoutent à celles survenues lors de la précipitation. Avantages: les composés non protéiques non solubles sont également éliminés.

## Extraction des protéines: limiter les dégradations des protéines solubilisées

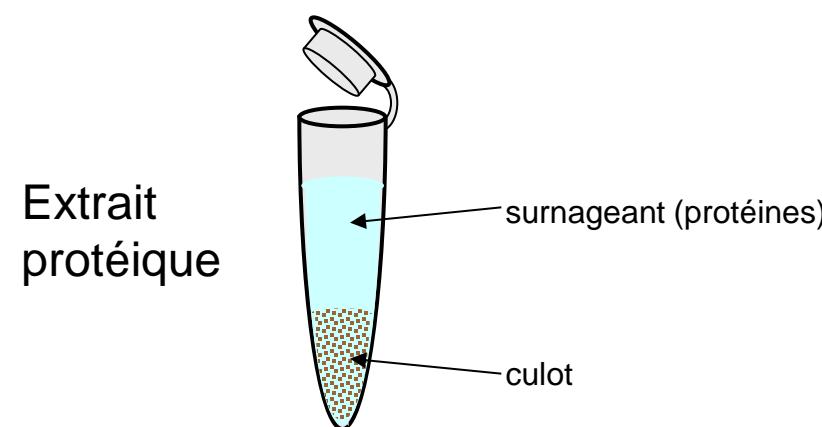
Jusqu'à la resolubilisation des protéines, l'extraction se déroule à basses températures ( $< 0^{\circ}\text{C}$ ) ce qui inactive toutes les enzymes. Cependant, la resolubilisation et les étapes de séparation subséquentes sont effectuées à température ambiante (sous peine de cristallisation des tampons). C'est pourquoi il est important d'établir et maintenir des conditions dénaturantes.

En solution, les protéines peuvent être modifiées par oxydation via l'activité polyphénol oxidase (PPO) ou dégradées par l'activité des protéases. Il est possible d'ajouter des composés contrecarrant ces activités au tampon de resolubilisation des protéines.

La PPO peut être inactivée par addition de polyvinyl polypyrrolidone (PVPP).

Les protéases/protéinases peuvent être inhibées par divers agents:

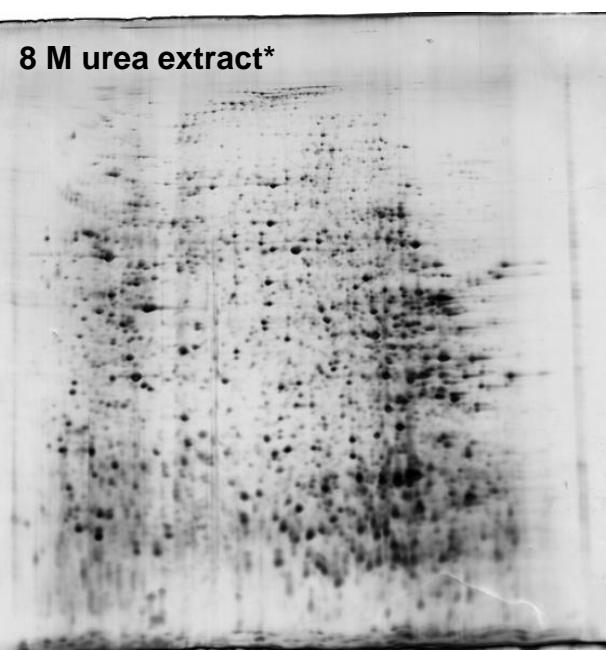
- PMSF (phenylmethyl sulfonide fluoride), inactive les protéases à sérine et cystéine; inactivé par DTT et 2-ME
- EDTA (ethylenediaminetetraacetic acid ), inhibe les metalloprotéases
- pH élevé qui inhibe la plupart des protéases
- cocktail, regroupe une variété d'inhibiteurs de protéases (large spectre)



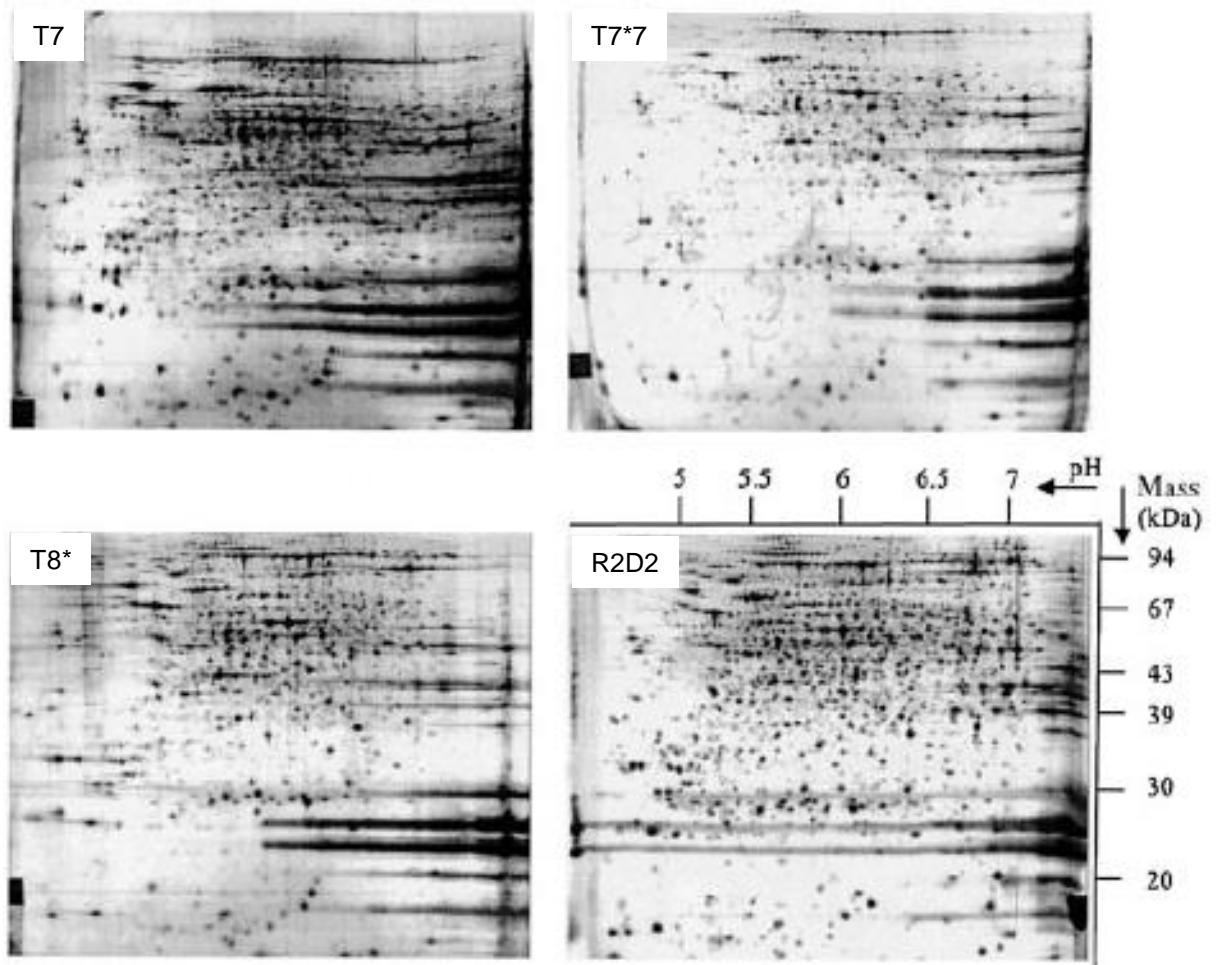
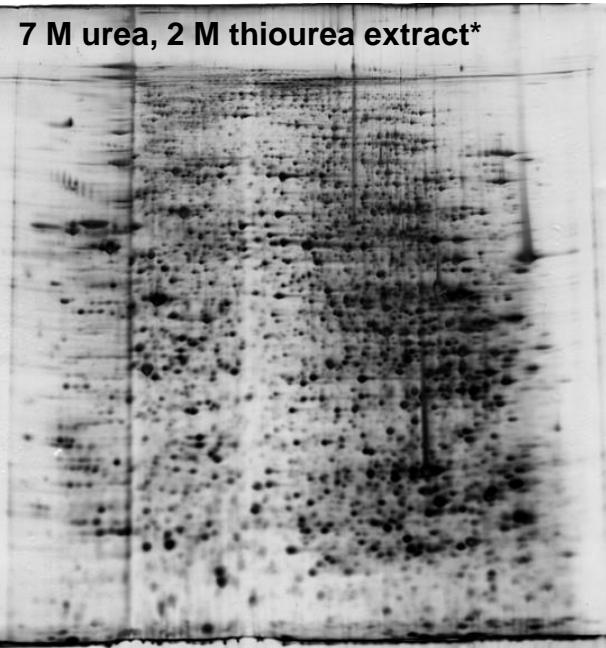
# Extraction des protéines: influence des composés chimiques

Urea vs. Urea/thiourea

8 M urea extract\*



7 M urea, 2 M thiourea extract\*

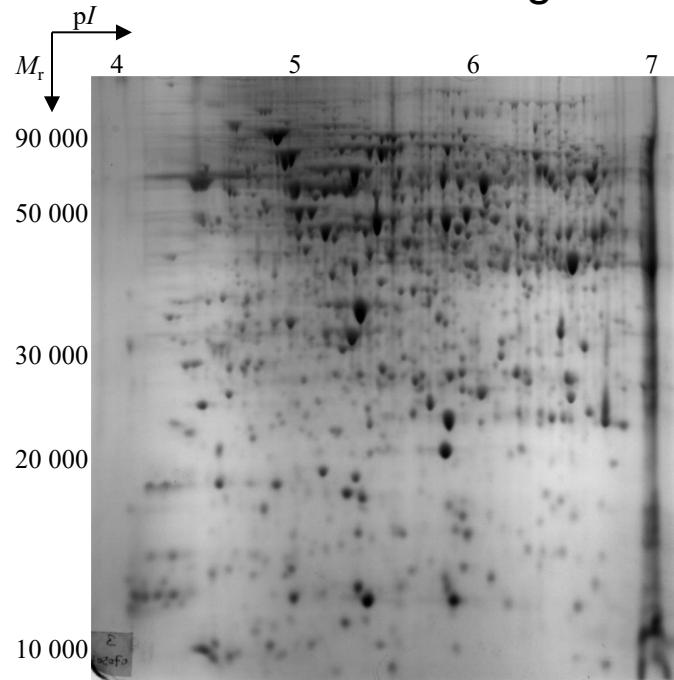


	T2 buffer	T7 buffer	T7*7 buffer	T8* buffer	T10 buffer	R2D2 buffer
Chaotropes	Urea 8 M	Urea 8 M	Urea 8 M	Urea 7 M Thiourea 2 M	Urea 7 M Thiourea 2 M	Urea 5 M Thiourea 2 M
Detergents	CHAPS 2%	CHAPS 2%	CHAPS 4%	CHAPS 4%	CHAPS 2%	CHAPS 2% SB3-10 2%
Reducing agents	DTT 20 mM	DTT 20 mM	DTT 25 mM	DTT 25 mM	TCEP 5 mM	DTT 20 mM TCEP 5 mM
Carrier ampholytes	absent	4–6.5 : 0.5%	4–6.5 : 0.5%	4–6.5 : 1%	4–6.5 : 0.5%	4–6.5 : 0.5% 3–10 : 0.25%

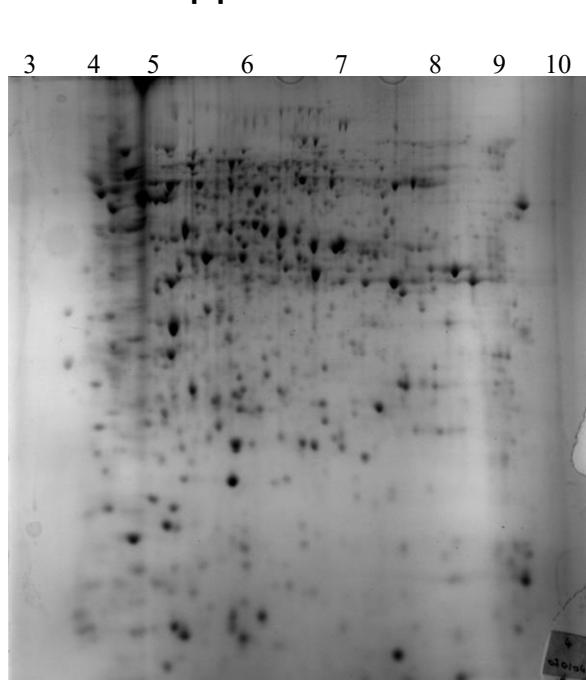
# Extraction des protéines: méthode à optimiser pour chaque échantillon



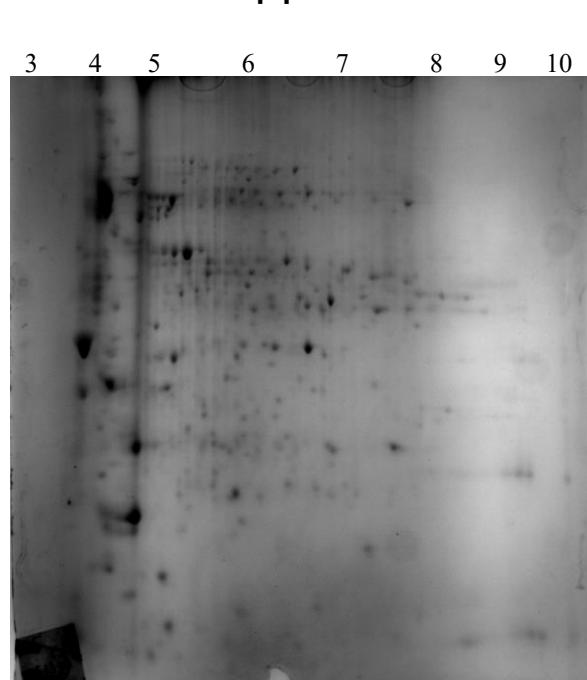
Extrémité de la tige



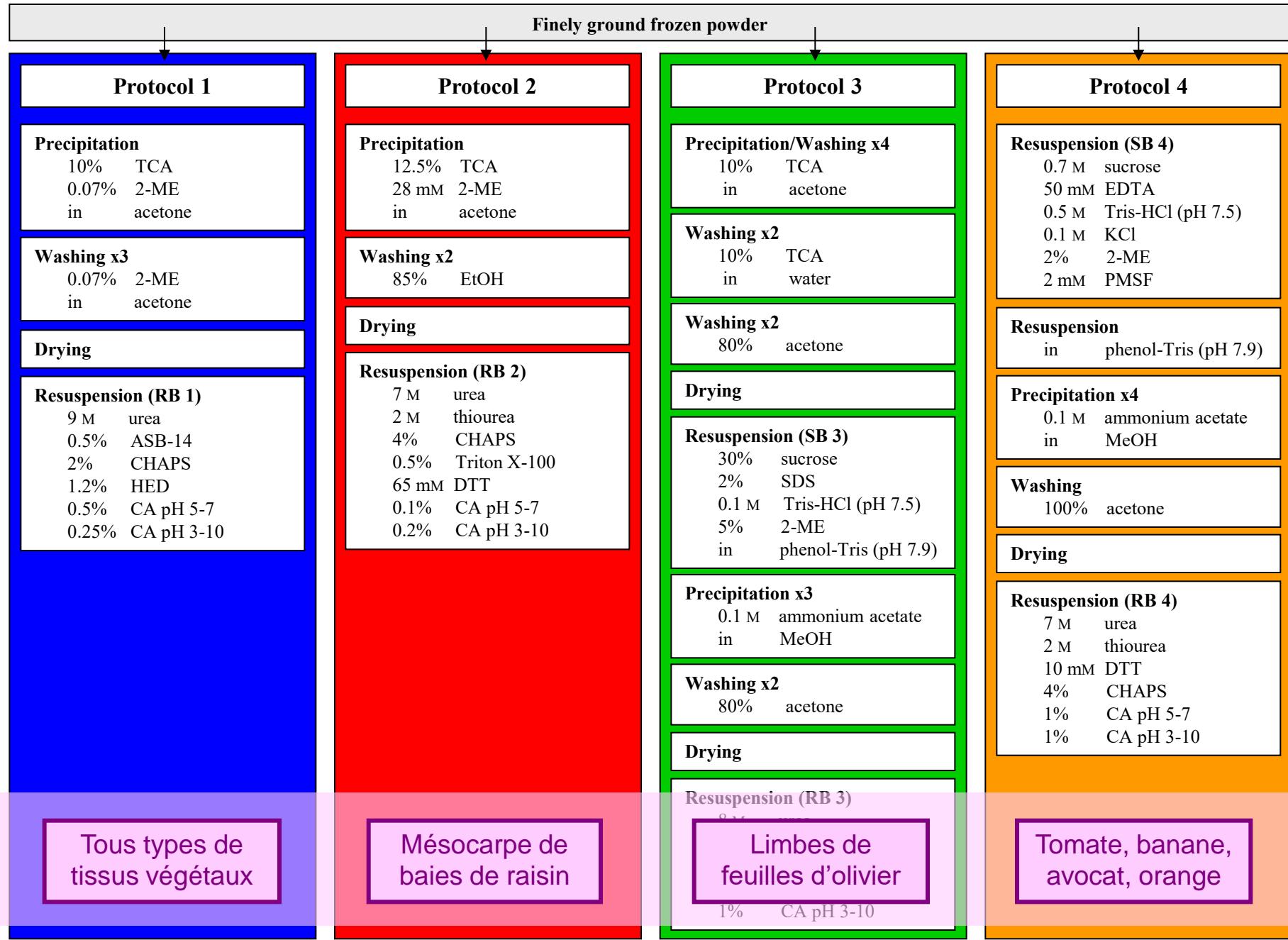
Grappe immature



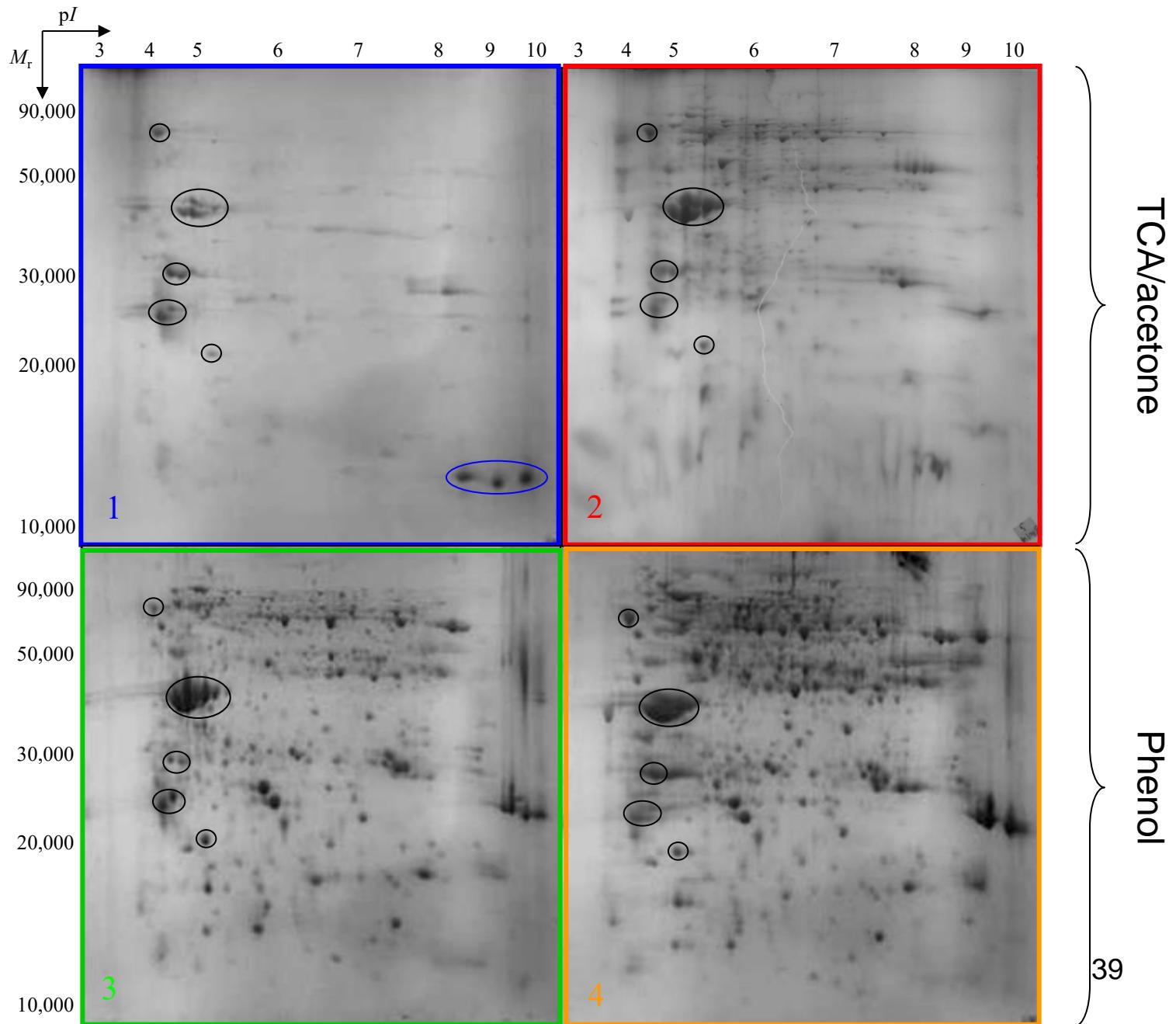
Grappe mûre



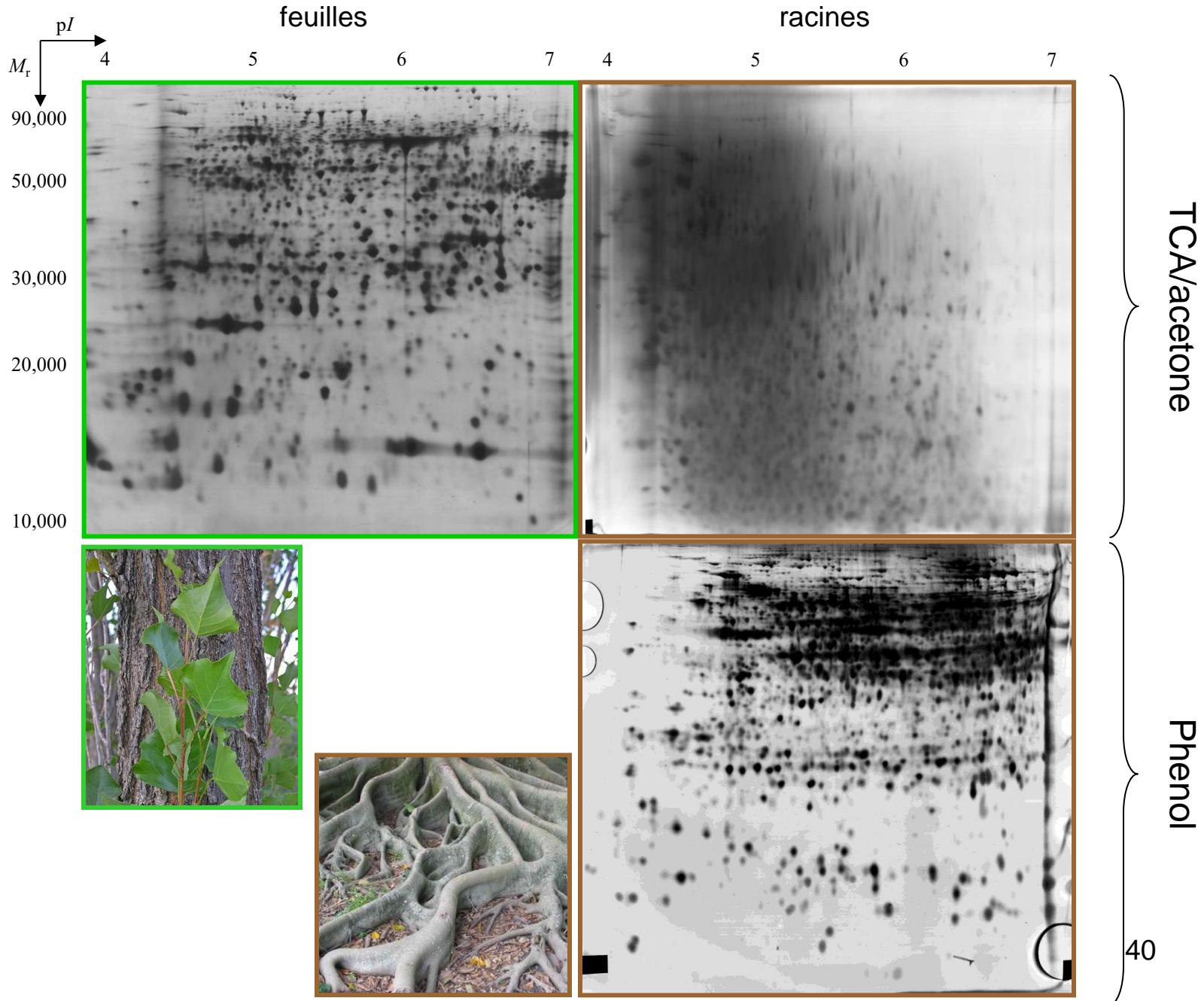
# Extraction des protéines: méthode à optimiser pour chaque échantillon



# Extraction des protéines: méthode à optimiser pour chaque échantillon



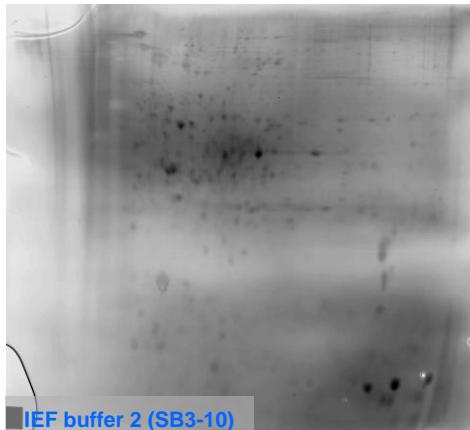
# Extraction des protéines: méthode à optimiser pour chaque échantillon



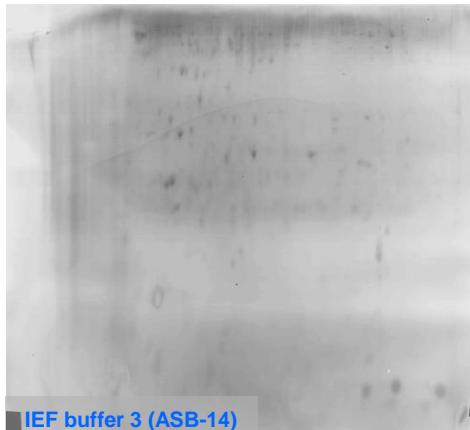
# Extraction des protéines: méthode à optimiser pour chaque échantillon

7M urea  
2M thiourea  
1% DTT  
2% CHAPS  
1% CA's  
0.5% pim

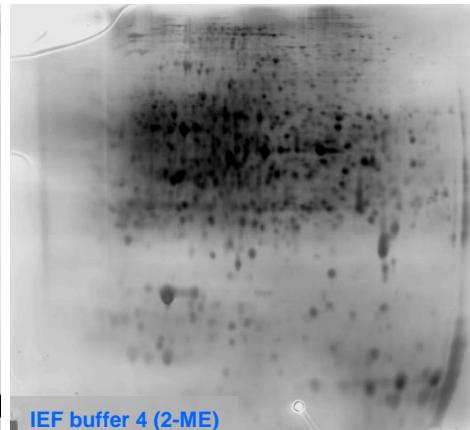
IEF buffer 1



IEF buffer 2 (SB3-10)

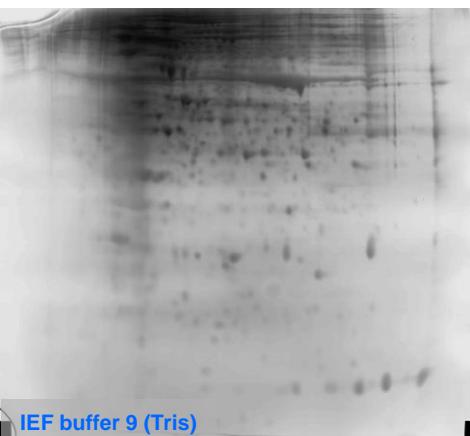


IEF buffer 3 (ASB-14)

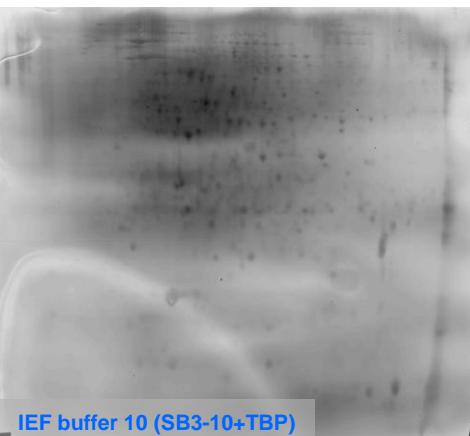


IEF buffer 4 (2-ME)

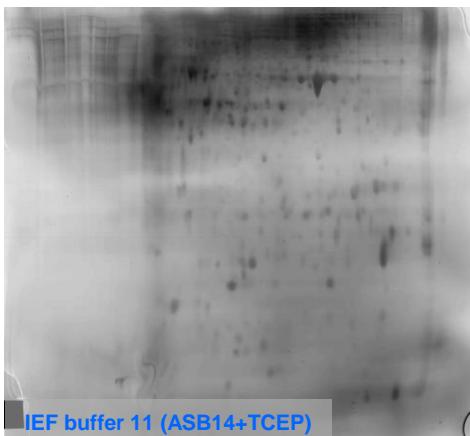
IEF buffer 5 (TCEP)



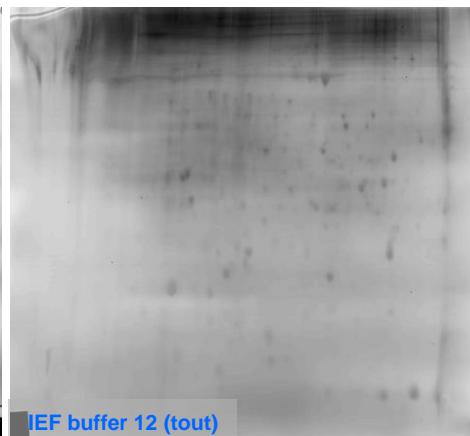
IEF buffer 6 (TBP)



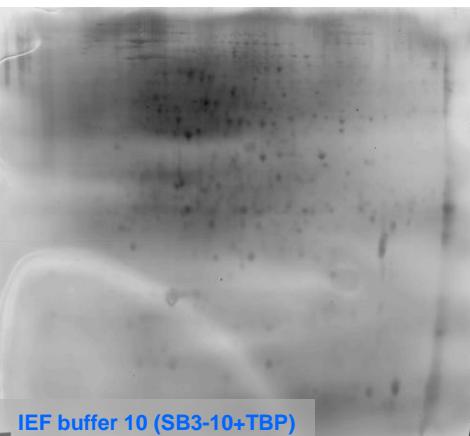
IEF buffer 7 (DNase)



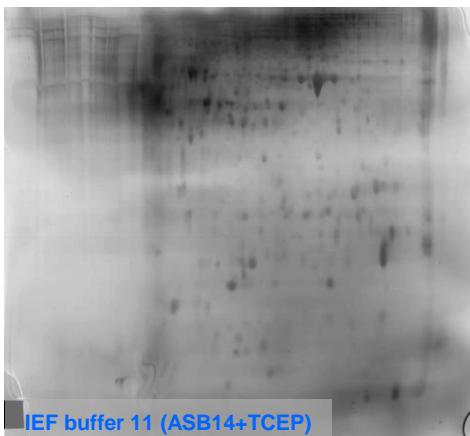
IEF buffer 8 (PVP)



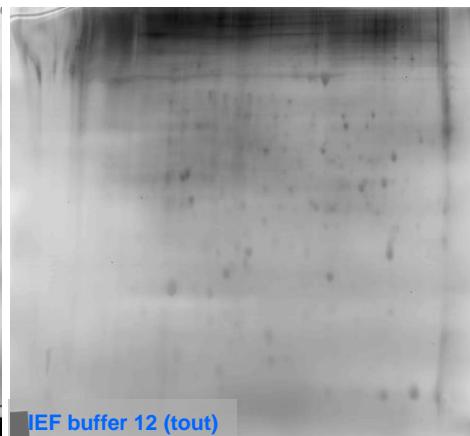
IEF buffer 9 (Tris)



IEF buffer 10 (SB3-10+TBP)

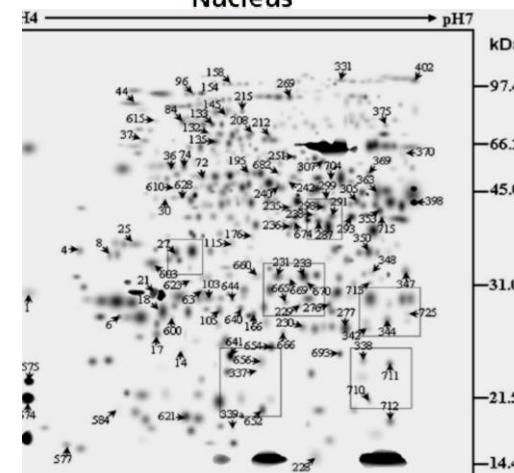
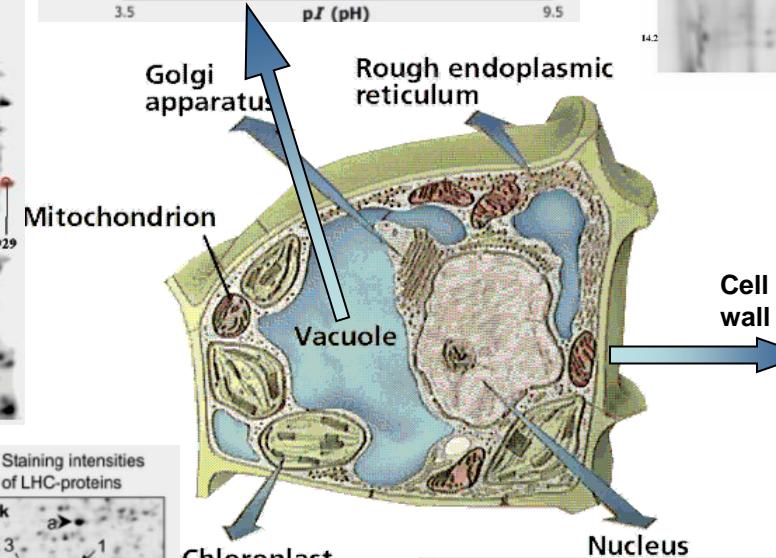
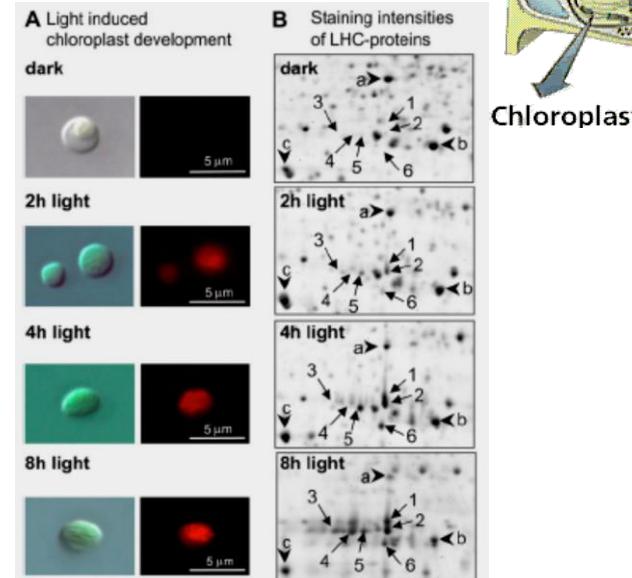
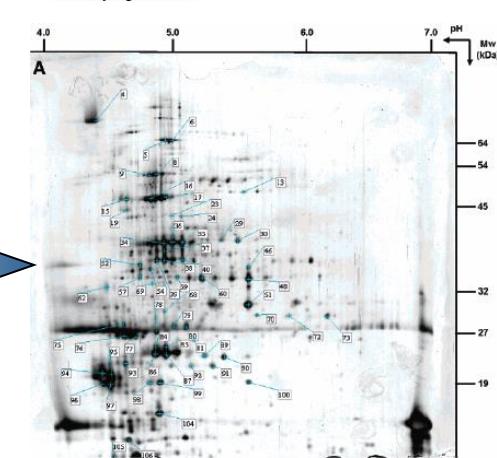
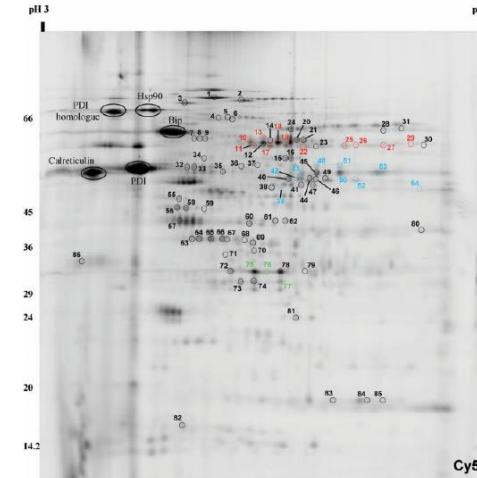
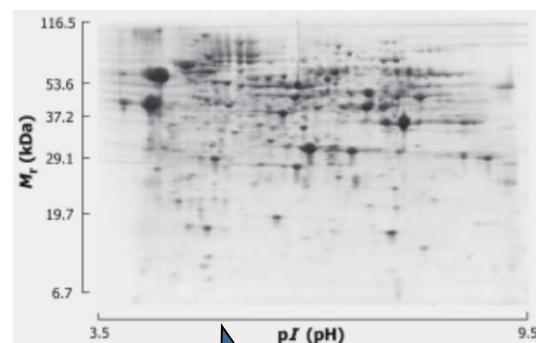
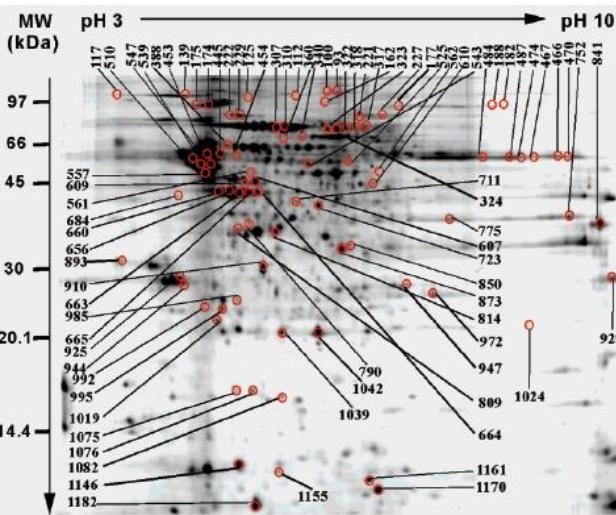


IEF buffer 11 (ASB14+TCEP)



IEF buffer 12 (tout)

# Extraction des protéines: réduire la complexité du mélange de protéines par fractionnement cellulaire

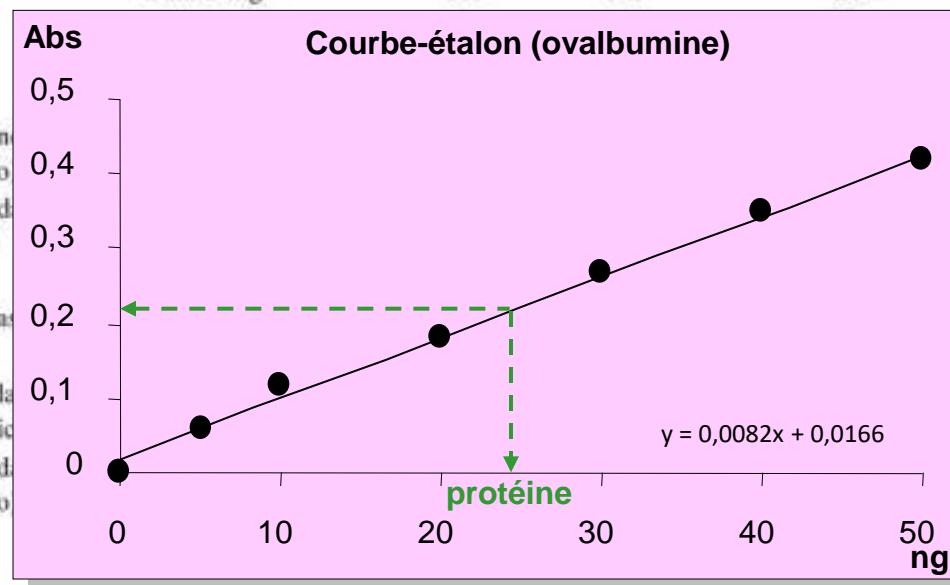


# Dosage des protéines

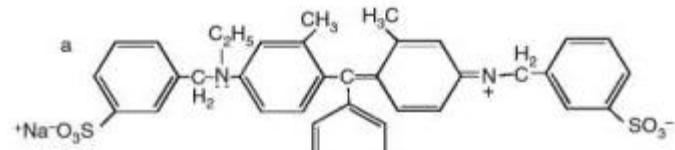
# Dosage des protéines: estimation de leur concentration par rapport à une gamme-étalon

## Advantages and Disadvantages of Some Common Methods for Protein Determination

Method	Range of Sensitivity	Minimum Vol./Amount	Destructive	Time of Assay	Protein-to-Protein Variation	Comments
A <sub>280</sub>	0.2–2 mg/ml	100 µl/20 µg	No	Instant	Large	Interference by UV absorbing materials
A <sub>260</sub>	10–50 µg/ml	100 µl/1 µg	No	Instant	Low	Interference by UV absorbing materials
Biuret	1–5 mg/ml	2 ml/2 mg	Yes	1 h	Low	Rapid color; interference by caustic reagents, NH <sub>4</sub> <sup>+</sup>
Lowry	0.01–1 mg/ml					Many interfering substances; slow color development
BCA	10–1200 µg/ml (standard), 0.5–10 µg/ml (micro)					Warming required, interference by reducing agents
Bradford	25–200 µg/ml (standard), 5–40 µg/ml (micro)					Interference by detergents and strong bases
Colloidal gold	20–200 ng					Assay pH should be acidic (around pH 3)
Ninhydrin	1–10 ng (Modified assay)					Protein must be hydrolyzed
Fluorescamine	0.5–50 µg/ml (standard), 50 ng–25 µg/ml (micro)					Not compatible with buffers containing primary amines
o-phthalaldehyde	10–500 µg/ml (standard), 0.05–2 µg/ml (micro)					Not compatible with buffers containing primary amines
CBQCA	0.12–1.8 µg/ml					Buffers containing primary amines and reducing agents are not compatible
Nano-Orange <sup>®</sup>	0.01–10 µg/ml	2 ml/20 ng	Yes	10 min at 95°C	Low	Not compatible with detergents

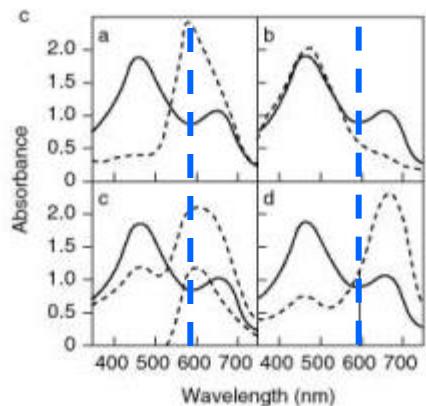
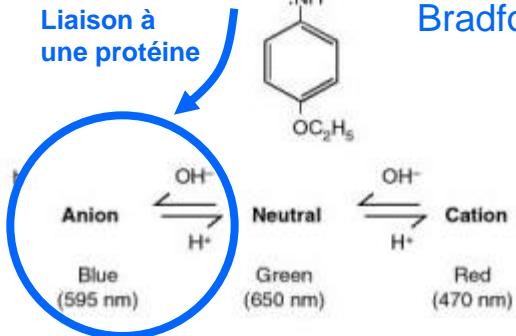


# Dosage des protéines: sensibilité des méthodes

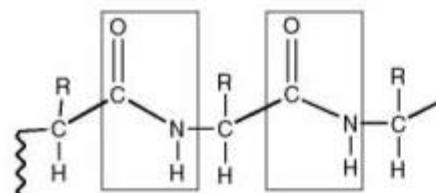


Liaison à une protéine

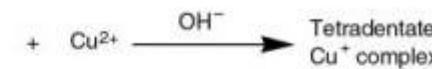
Bradford assay



spectrophotomètre



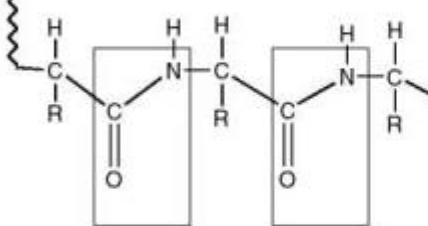
Peptide bond



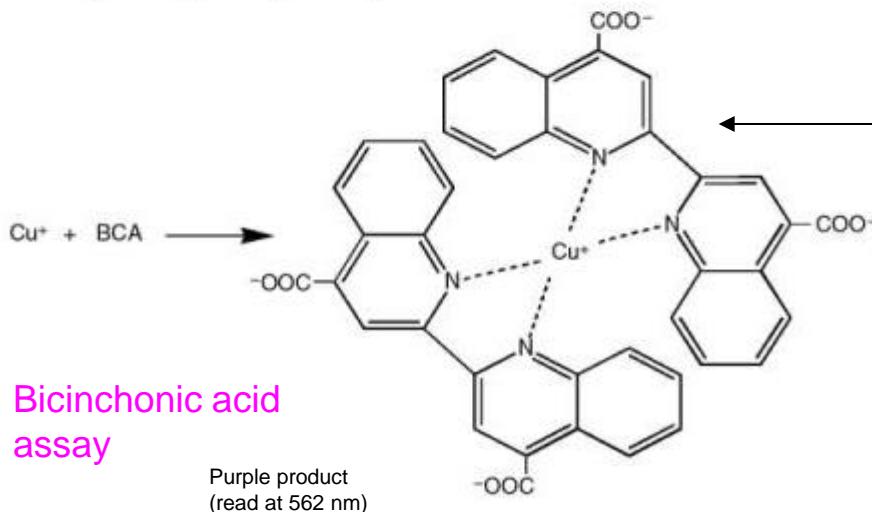
Lowry assay

Biuret assay

Blue product (read at 550 nm)



Blue product (read at 750 nm)



Bicinchoninic acid assay

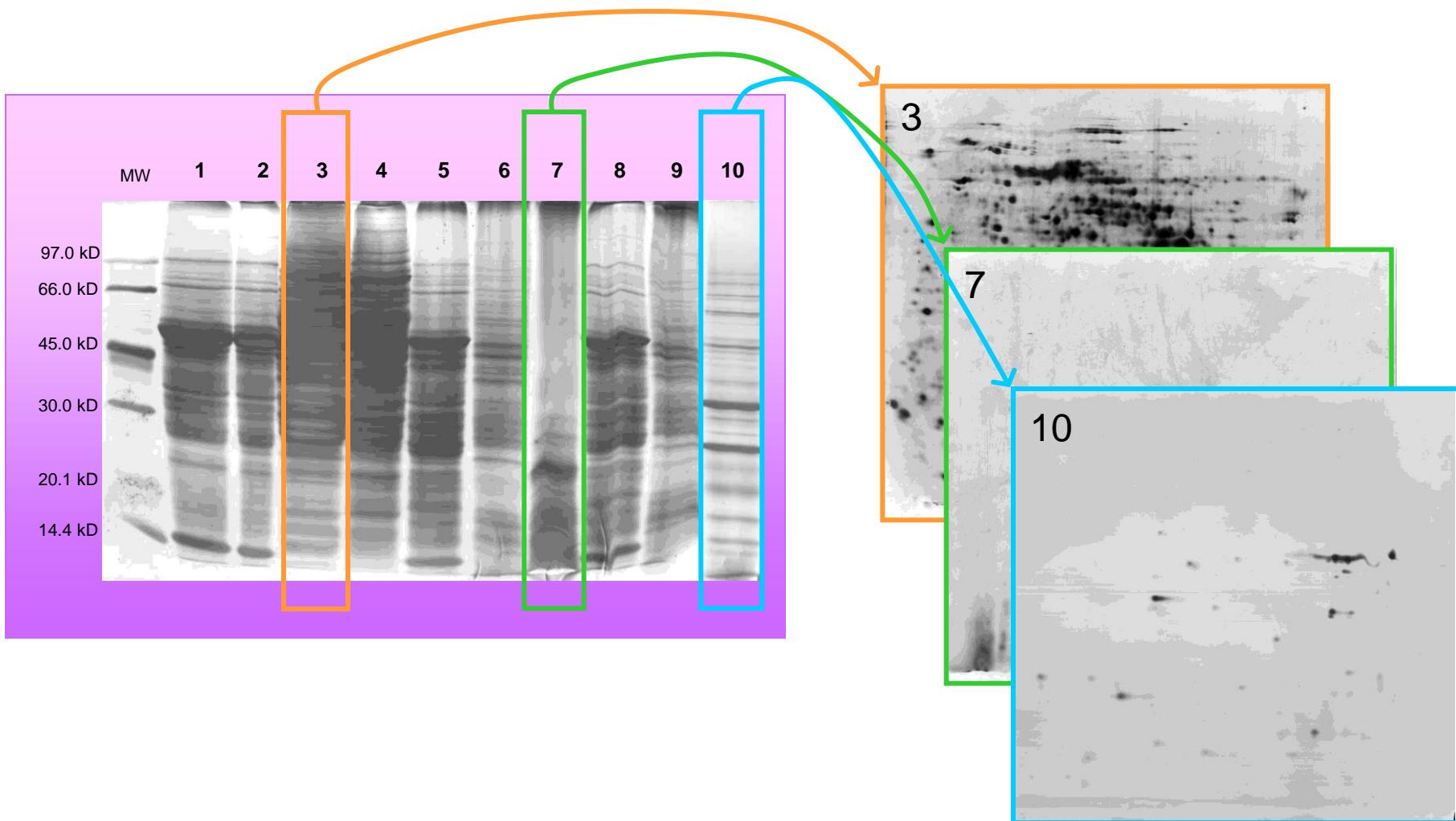
# Dosage des protéines: vérification de la qualité des extraits protéiques

SDS-PAGE système mini-gel: pas de gaspillage de matériel.

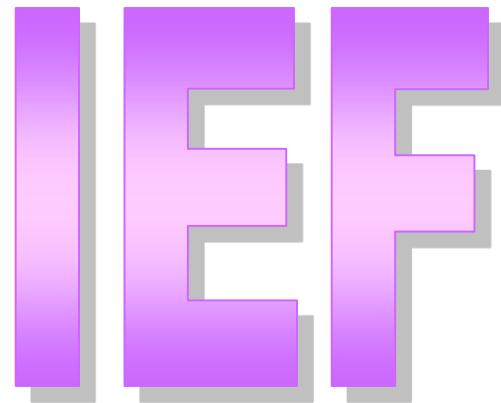
Si l'extrait est compatible avec la 2-DE, les profils protéiques 1-D et 2-D se correspondent.

Permet d'éliminer les échantillons récalcitrants à la méthode d'extraction choisie.

Ne garantit pas le succès de la séparation des protéines par 2-DE.



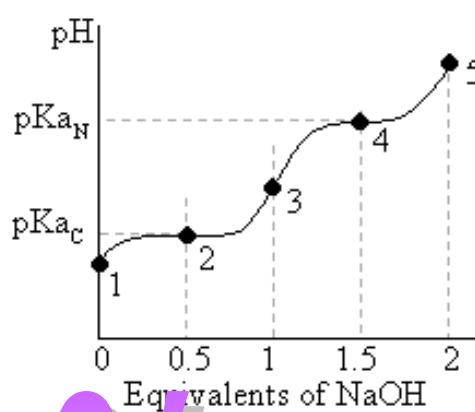
**1<sup>ère</sup> dimension:**



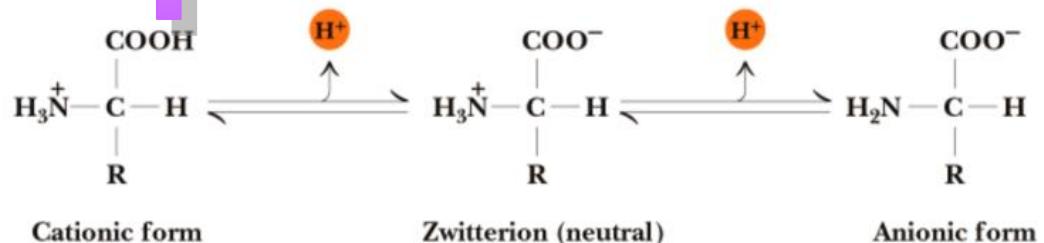
**(isoélectrofocalisation)**

# IEF: principe

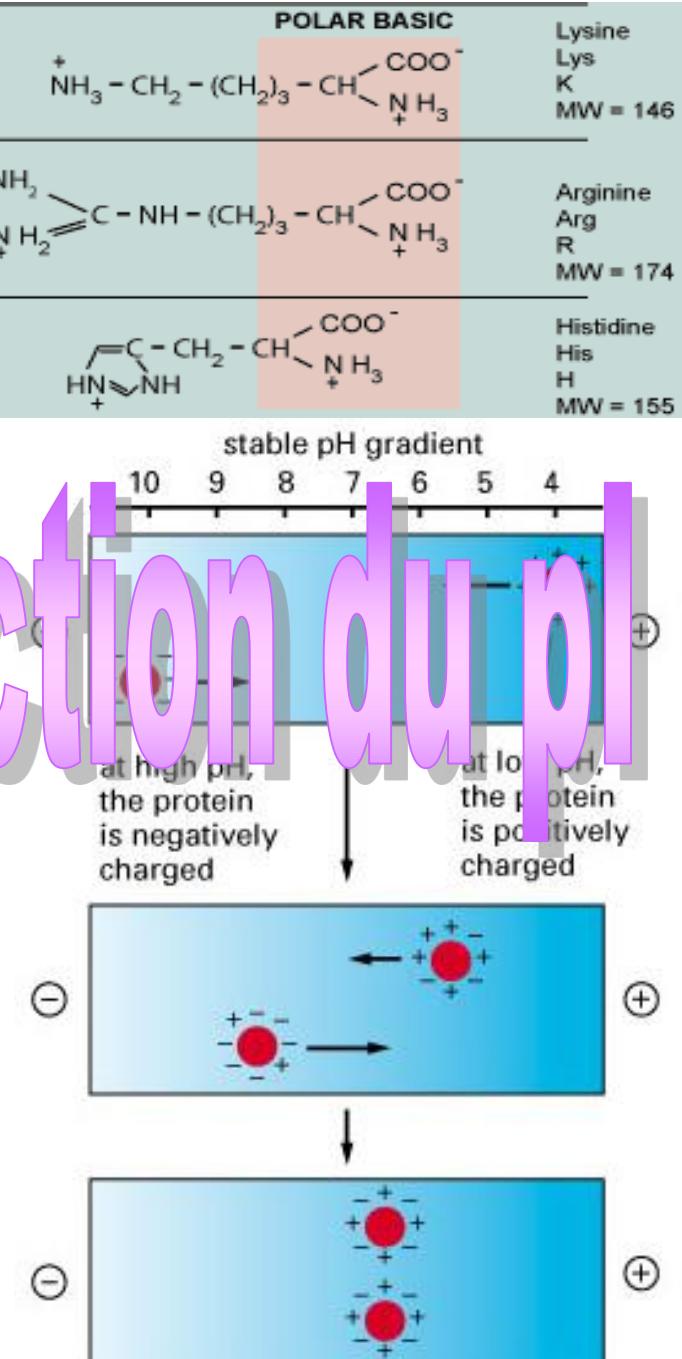
## AA contribuant à la charge des protéines



		POLAR BASIC	Lysine Lys K MW = 146
Aspartic acid	Asp		
Glutamine acid	Glu		
			Histidine His H MW = 155



Le **point isoélectrique (pl)** d'une molécule telle qu'un acide aminé, peptide ou protéine correspond au pH pour lequel sa charge nette s'annule. Par exemple, un acide aminé avec une chaîne latérale indissociable présente une charge nulle lorsque le pH est au milieu des 2 valeurs de pKa [pH = 1/2(pKaC + pKaN)].

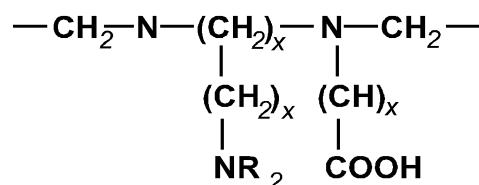


The protein shown here has an isoelectric pH of 6.5.

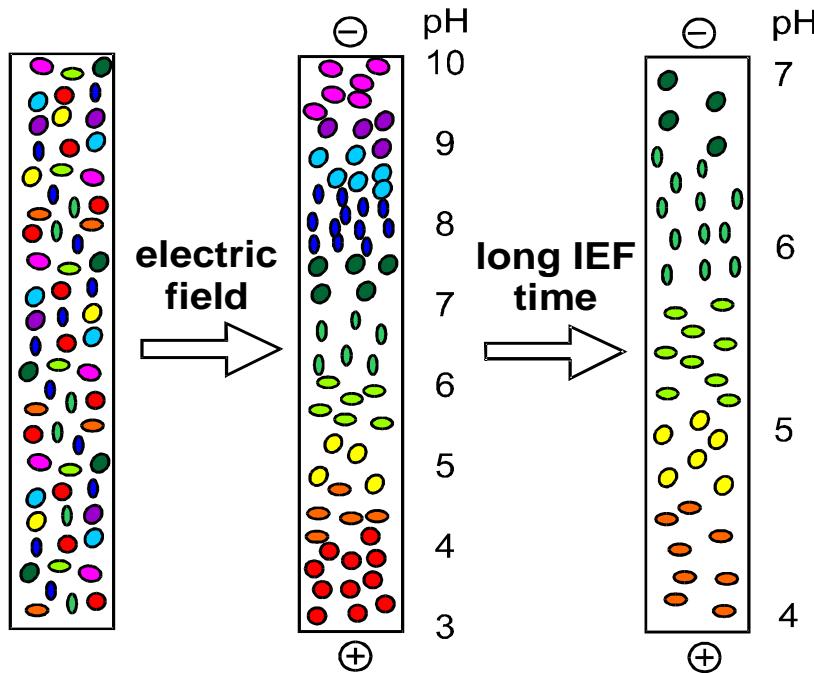
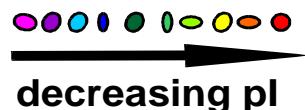
# IEF: Carrier Ampholytes (CA)

## Pharmalytes Ampholines

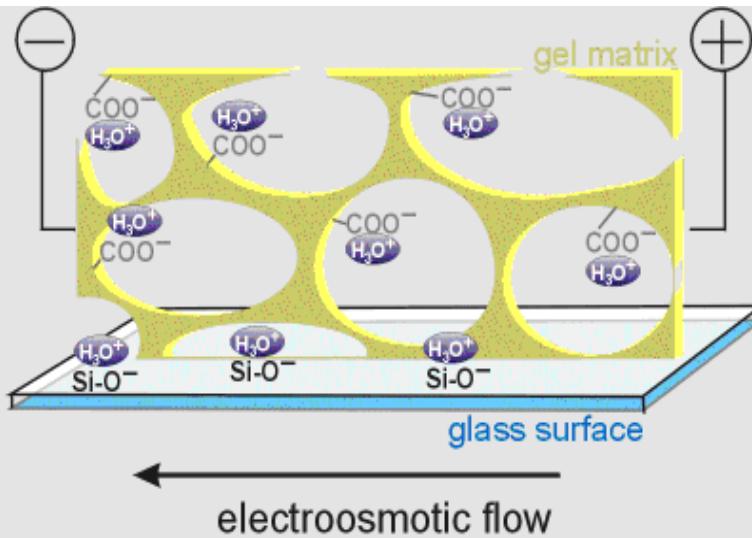
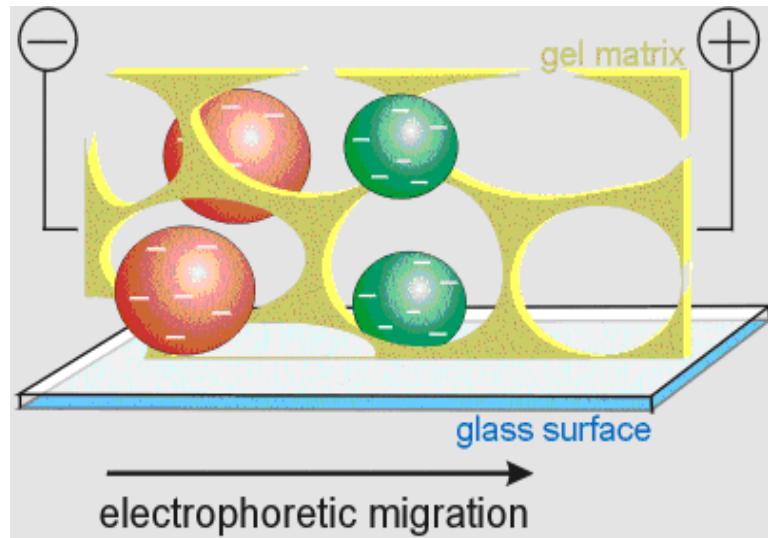
CA-IEF:  
déplacement  
unidirectionnel  
 $(-) \rightarrow (+)$   
NEPHGE  $(-) \leftarrow (+)$



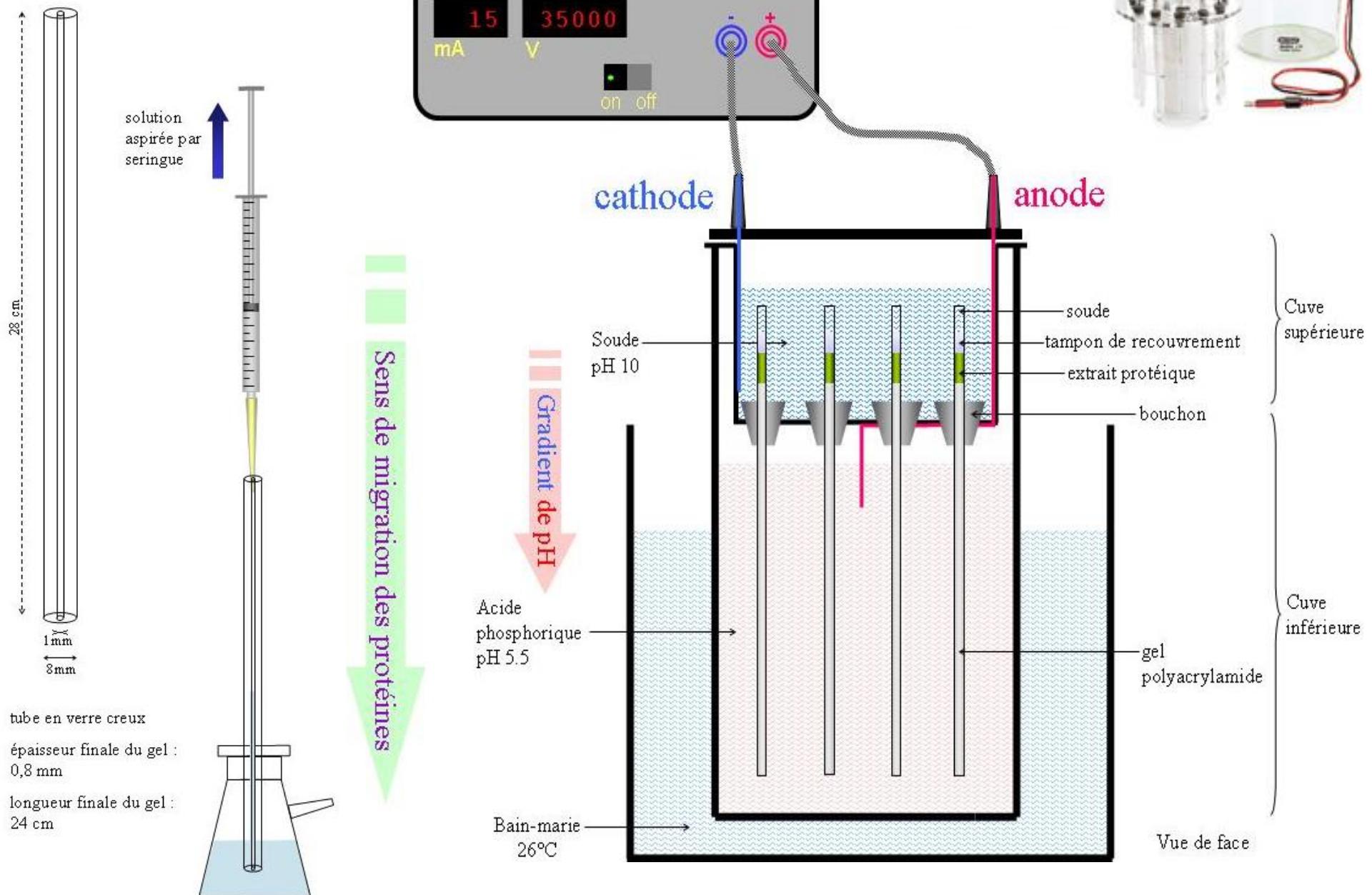
where R = H  
or  $-(\text{CH}_2)_x-\text{COOH}$ ,  
 $x = 2$  or 3



Electroendosmose: déplacement de l'anode (+) vers la cathode (-) d'un liquide polaire au sein d'une structure poreuse chargée sous l'influence d'un champ électrique.



# CA-IEF



## Problèmes associés aux CA:

- Mauvaise reproductibilité d'un lot à l'autre
- Gradient instable
- Dérive du gradient limite la reproductibilité
- Dérive du gradient entraîne la perte des protéines les plus acides et les plus basiques

## Problèmes associés à CA-IEF:

- Fragilité du matériel difficile à manipuler et entretenir
- Difficulté de traçage des échantillons
- Coulage des gels sujet à variation
- Capacité de dépôt d'extrait protéique limité (40 µL)
- Migration dangereuse à haut voltage (2000V) avec tampons liquides (arcs électriques)
- Expulsion des gels très délicate (par pression au moyen d'une seringue)

- mouvement des ions  $\text{H}_3\text{O}^+$  et  $\text{OH}^-$
- accumulation  $\text{H}_3\text{O}^+$  à la cathode
- acidifie le côté cathode (pH10 biaisé)
- idem côté anode

anode « + »

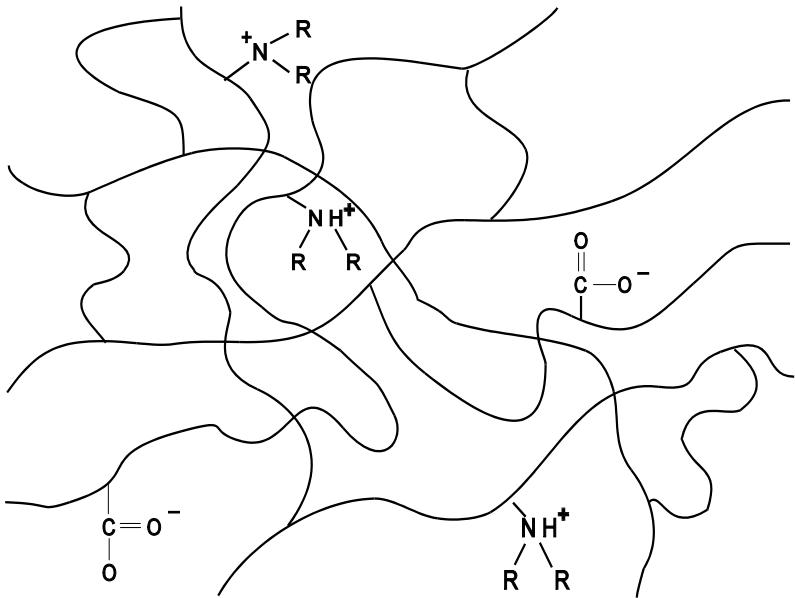
pH 3  
pH 4

cathode « - »

$\text{H}_3\text{O}^+$   
pH 10  
pH 8

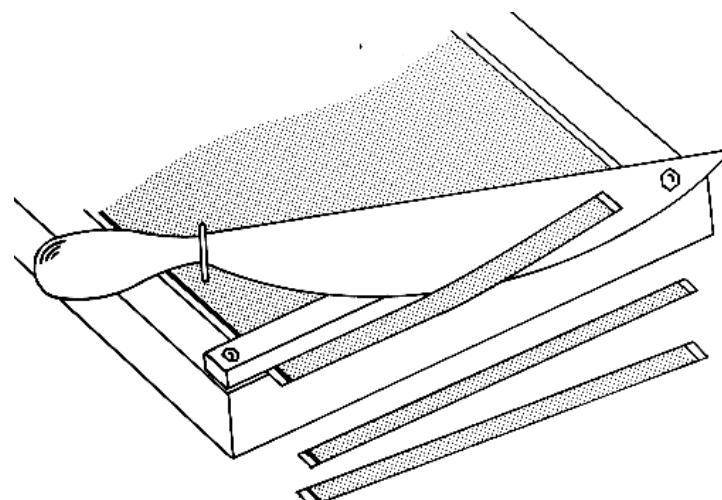
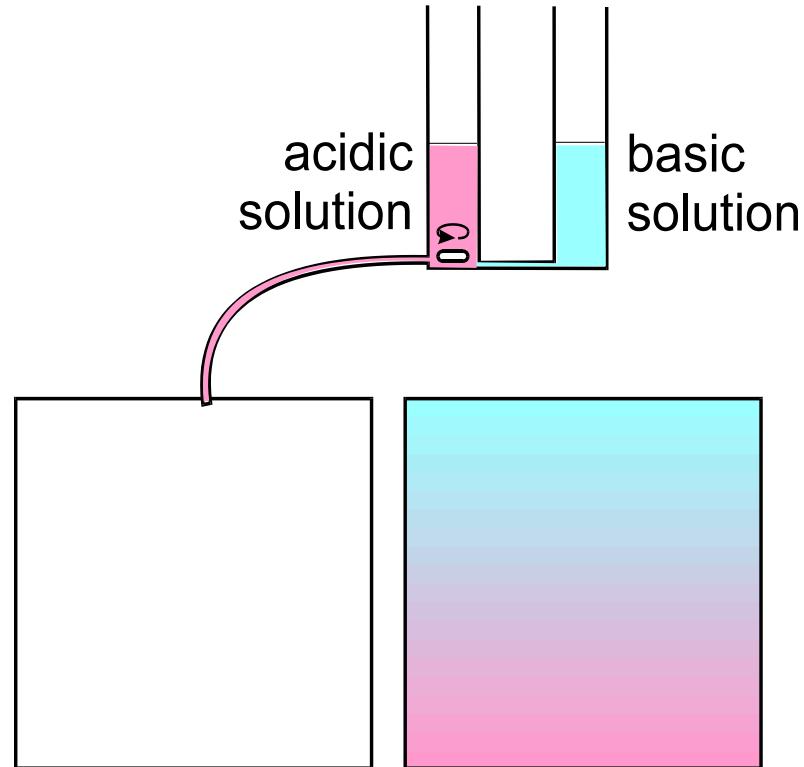
# IEF: immobilized pH gradient (IPG) strip

IPG-IEF: déplacement bidirectionnel  
 $(-) \leftrightarrow (+)$



Acrylamido buffers: **Immobiline®**  
 $\text{CH}_2=\text{CH}-\text{CO}-\text{NH}-\text{R}$ ,  
R contains a carboxylic  
or a tertiary amino group

**Immobiline Gels**  
(0.5 mm gel layers on film supports)

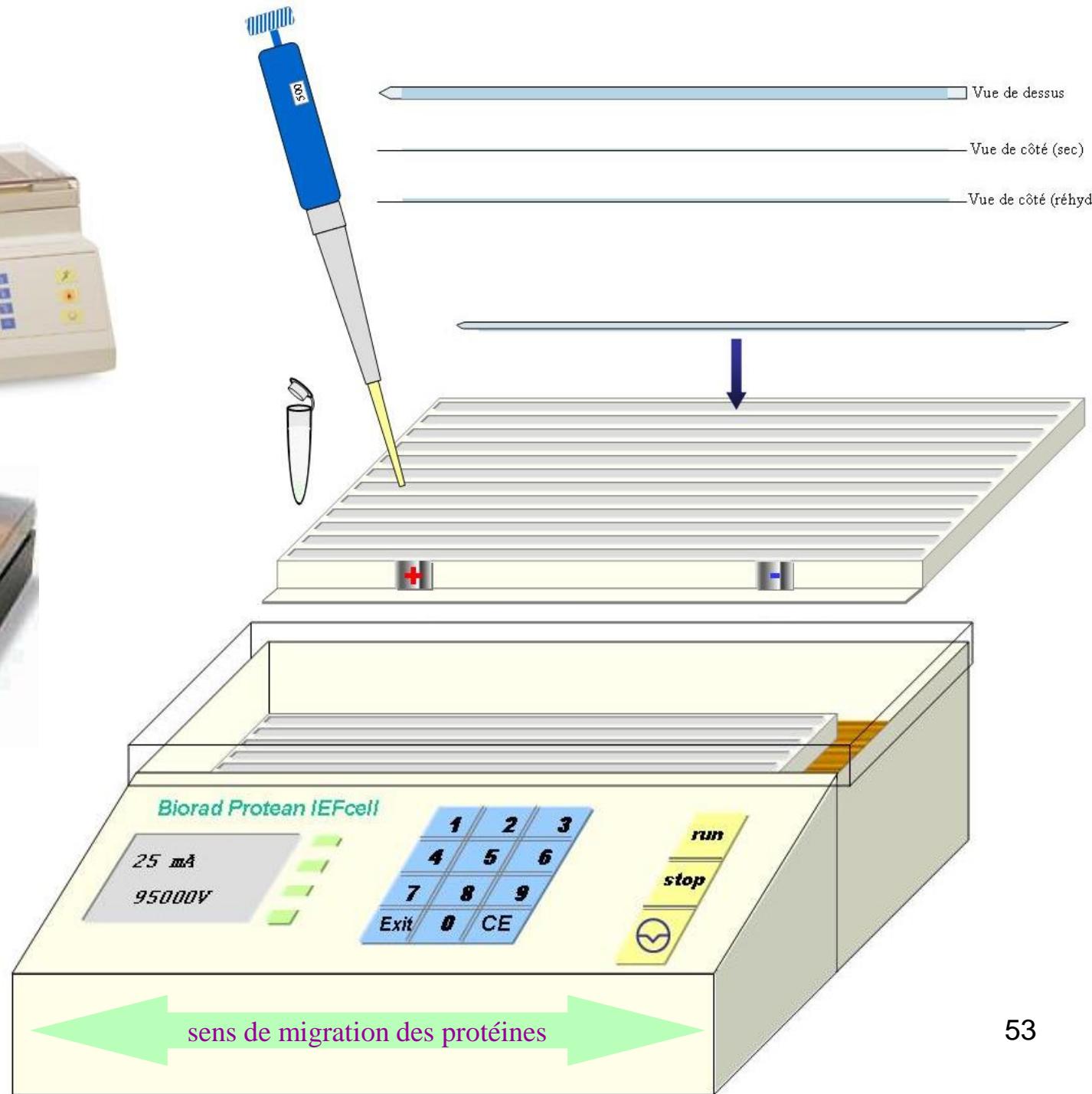


# IPG-IEF

BIO-RAD



Amersham  
GE Healthcare



## Avantages des IPG strips:

- Excellente reproductibilité d'un lot à l'autre
- Gradient stable (absence de dérive)
- Grande capacité de dépôt d'extrait protéique (500 µL)
- Possibilité de dépôt d'extrait protéique par réhydratation
- Les protéines les plus acides et les plus basiques sont résolues
- Système d'étiquetage autorisant le traçage des échantillons
- Prêts à l'emploi
- Variété de tailles et de gradients
- Facilité d'utilisation



## Avantages de l'IPG-IEF:

- Matériel résistant, très facile à utiliser et à entretenir
- Migration à très haut voltage (10000V) à sec (système Peltier) sécurisé



# CA-IEF vs. IPG-IEF

Non seulement plus facile à utiliser mais également bien plus performante !

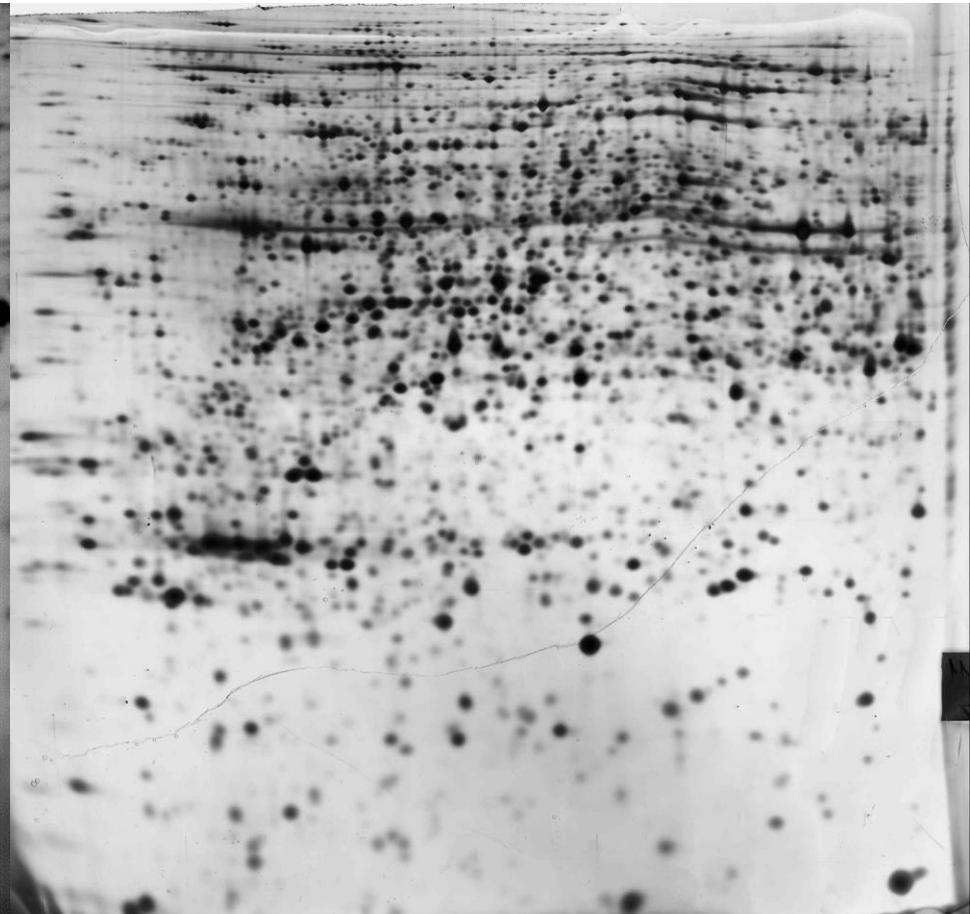


CA-IEF



1500 spots

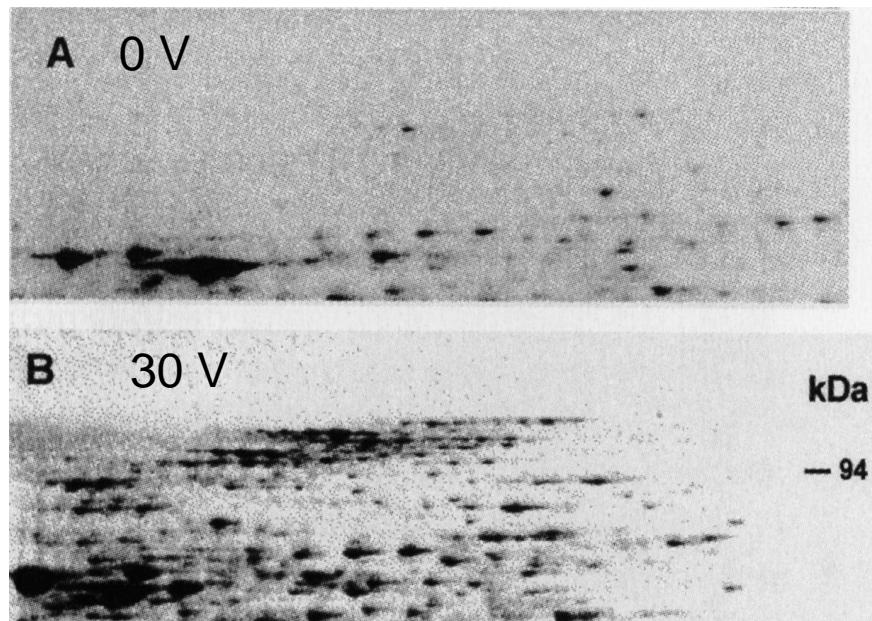
IPG-IEF



3000 spots

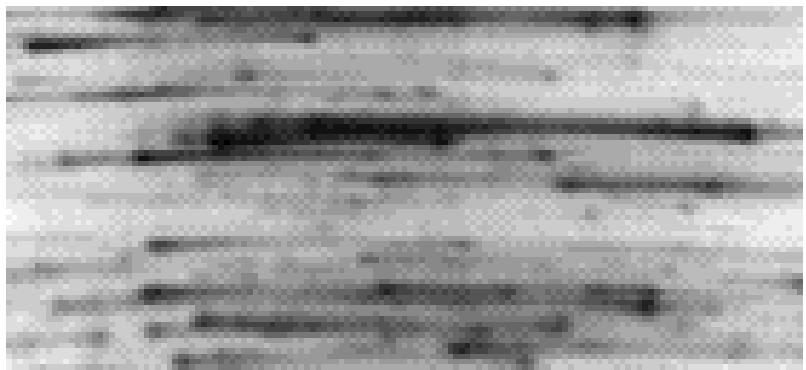
# IPG-IEF: quelques paramètres

Réhydratation passive (0V) vs. Réhydratation active (50V)

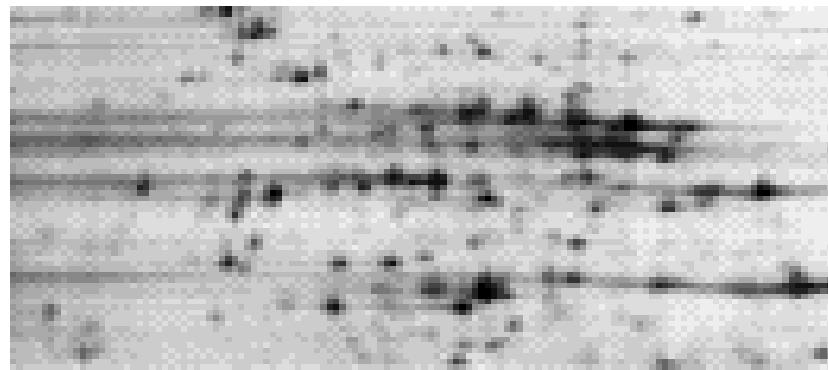


Temps de focalisation

pas assez !



trop !



# IPG-IEF: équilibration des strips

Une étape intermédiaire d'équilibration des strips est nécessaire entre IEF et SDS-PAGE.

## Equilibration buffer

Tris-HCl 1.5M, pH8.8	16.75 ml	50 mM
Urea	180.17 g	6 M
Glycerol (87% stock)	172.5 ml	30 %
SDS	10 g	2 %
Bromophenol blue	trace	trace
Di-ionized water		qsp 500 ml

1<sup>er</sup> bath/ DTT (100 mg/10 ml), 10ml per strip, 15min

2<sup>nd</sup> bath / iodoacetamide (250 mg/10 ml), 10ml per strip, 15min

Urée et glycérol minimisent les effets d'électroendosmose.

Les protéines sont enrobées de Sodium Dodecyl Sulfate (SDS, cation) donc deviennent chargées négativement.

Le DTT rompt les ponts disulfures.

L'iodoacétamide (IAA) alkyle les ponts rompus pour éviter qu'ils ne se reforment; IAA élimine l'excès de DTT limitant le phénomène de « streaking ».

1: [Electrophoresis](#). 1988 Jan;9(1):57-9.

Horizontal two-dimensional electrophoresis with immobilized pH gradients using PhastSystem.

[Görg A](#), [Postel W](#), [Günther S](#), [Friedrich C](#).

Lehrstuhl für Allgemeine Lebensmitteltechnologie, Technische Universität München, Freising-Weihenstephan, Federal Republic of Germany.

Protocols for horizontal two-dimensional electrophoresis with immobilized pH gradients in the first dimension were modified for horizontal micro two-dimensional electrophoresis using PhastSystem. Different equilibration conditions of the first-dimensional immobilized pH gradient gel strip prior to second-dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis were evaluated. Silver stained two-dimensional patterns were obtained within 3.5 h.

PMID: 3234335 [PubMed - indexed for MEDLINE]

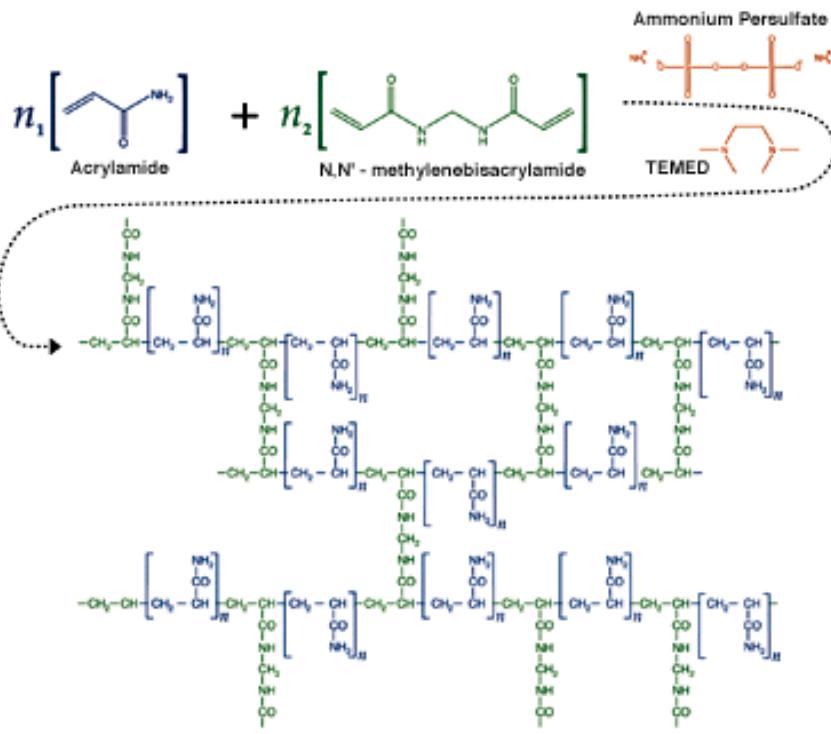
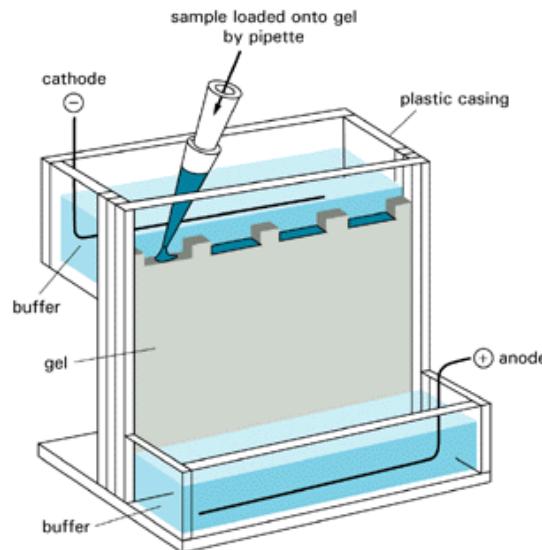
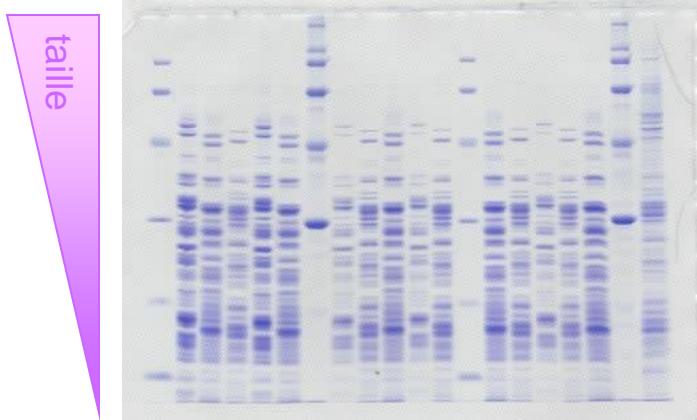
2ème dimension:

**SDS-PAGE**

(Sodium Dodecyl Sulfate-polyacrylamide gel electrophoresis)

# SDS-PAGE: principe

Séparation de protéines dont la charge est rendue négative par l'enrobage de molécules de SDS (d'autant plus de molécules de SDS que la protéine est volumineuse) le long d'un gel de polyacrylamide (plus concentré donc plus résolutif que les gels d'agarose).

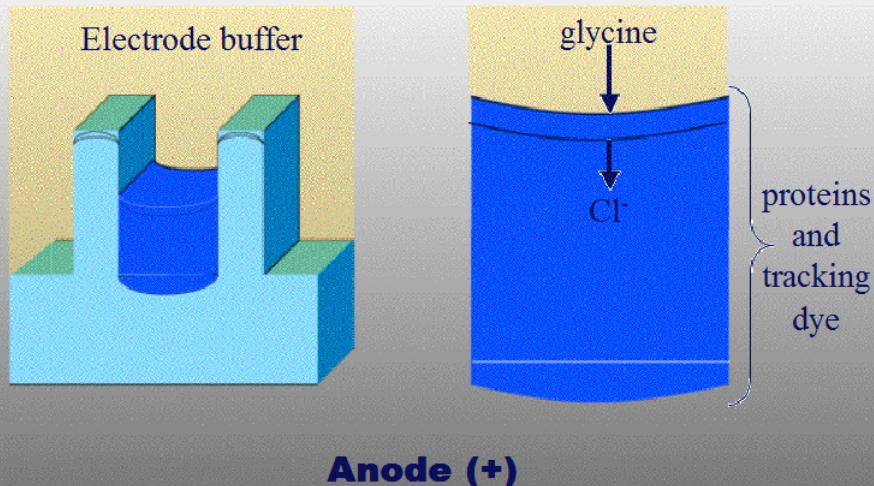


## Acrylamide solution

Acrylamide 40%	12%
Piperazine di-acrylamide/bis-acrylamide	0.5%
Tris 2M pH 8.8	0.5M
SDS (10%)	0.15%
Di-ionized water	qsp 2.5 L
TEMED	0.5%
Ammonium persulfate (10%)	2.5%

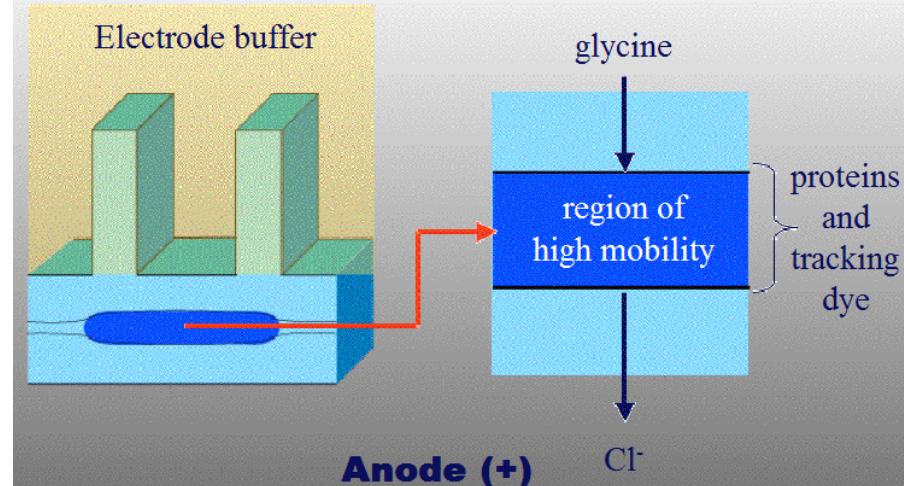
## SDS-PAGE: principe

Cathode (-)



Anode (+)

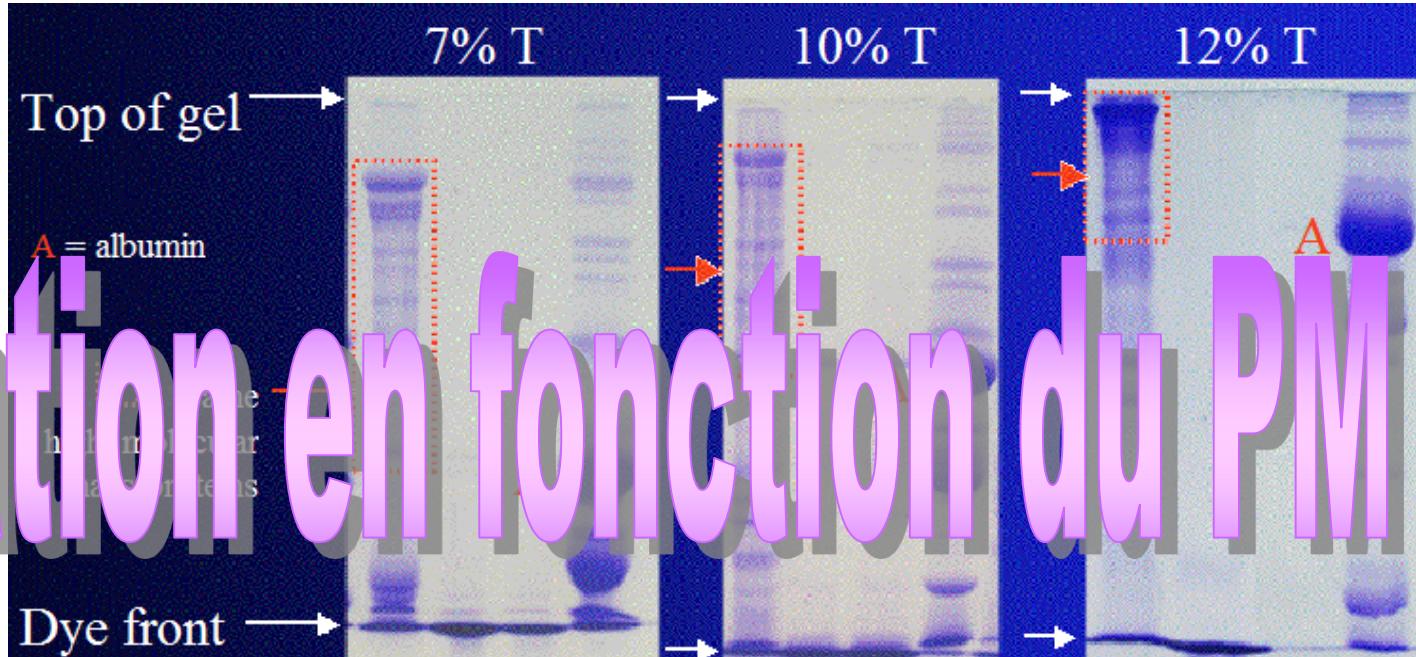
Cathode (-)



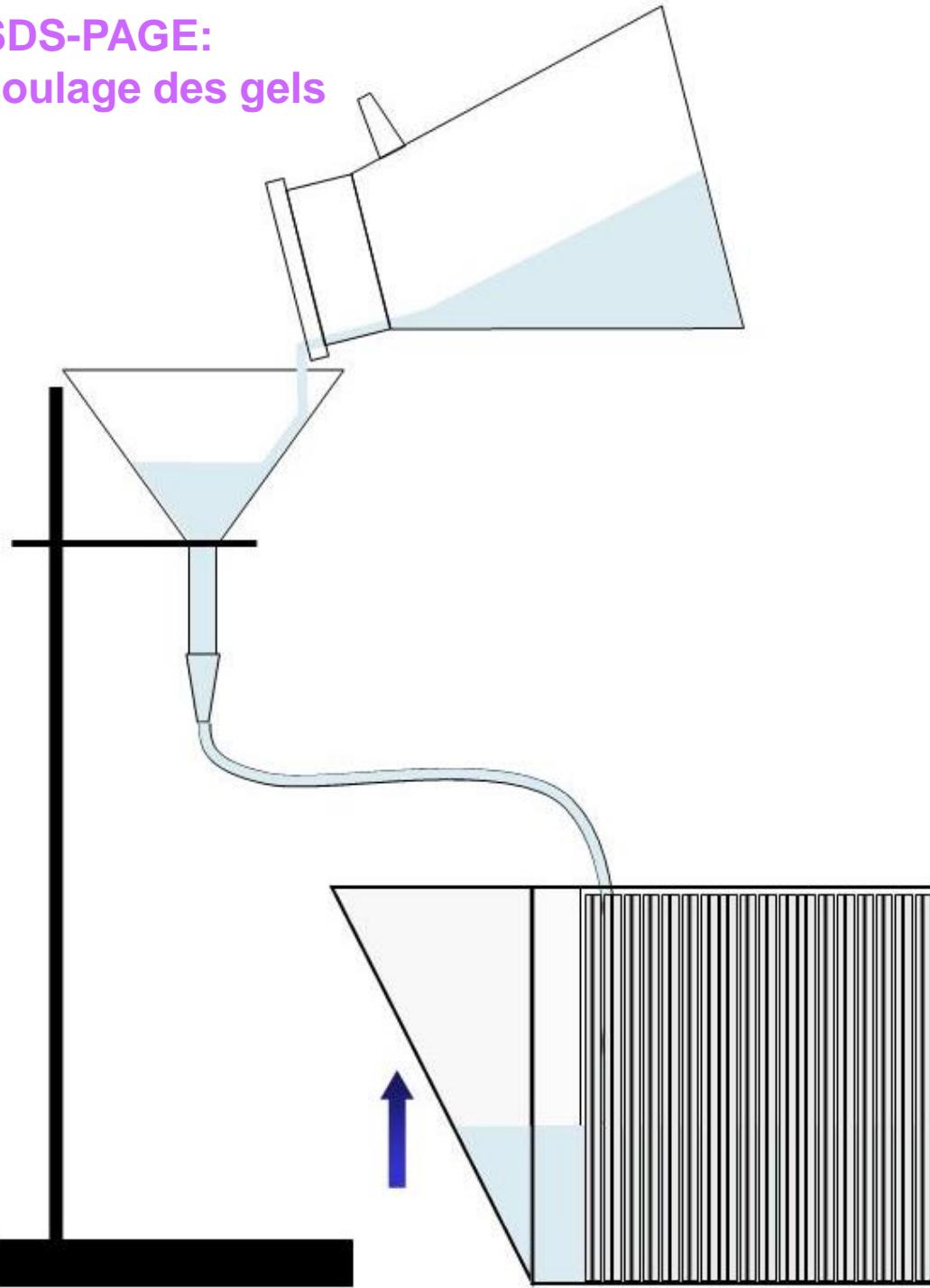
Tampon de migration:

Tris-HCl	25 mM
Glycine	0.2 M
SDS	0.1 M

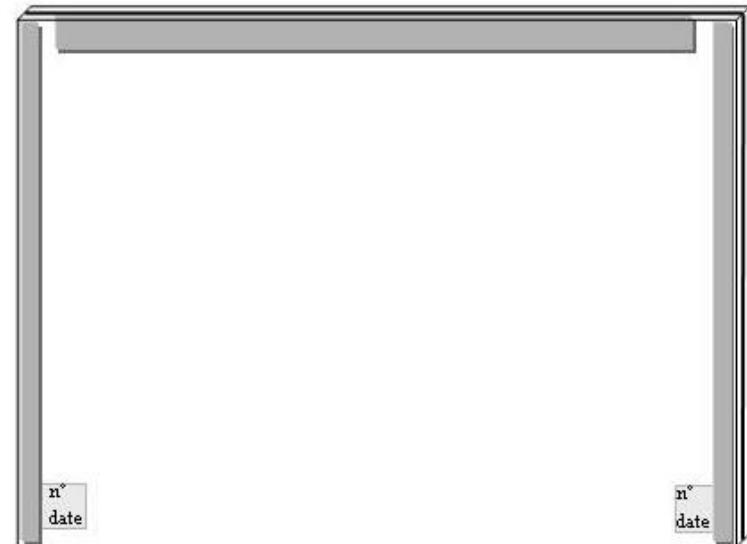
Séparation en fonction du PM



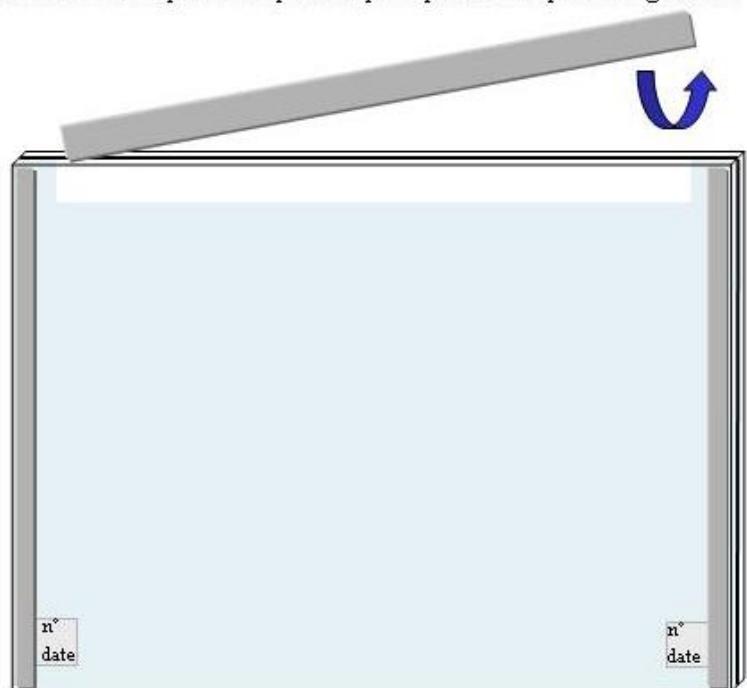
## SDS-PAGE: coulage des gels



Avant coulage du gel de 2ème dimension



Après coulage et polymérisation du gel de 2ème dimension.  
Retrait de l'espacement supérieur pour pouvoir déposer le gel IEF.



## SDS-PAGE:

## coulage des gels puis transfert des strips

Polymérisation de l'acrylamide: 2h minimum

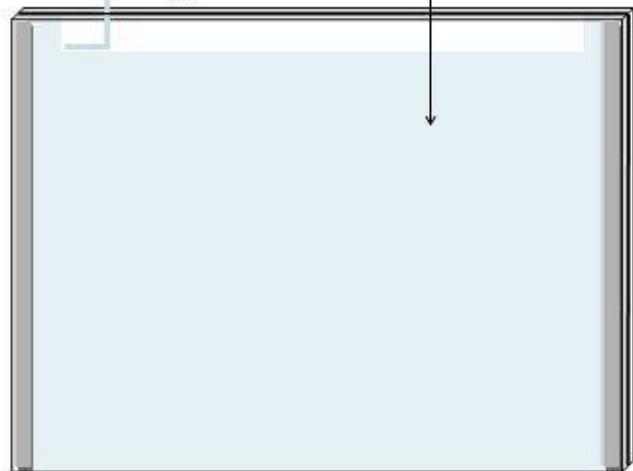


gel IEF-  
ampholites  
déposé sur  
plaqué 2D,  
sous pression  
exercée par la  
seringue.

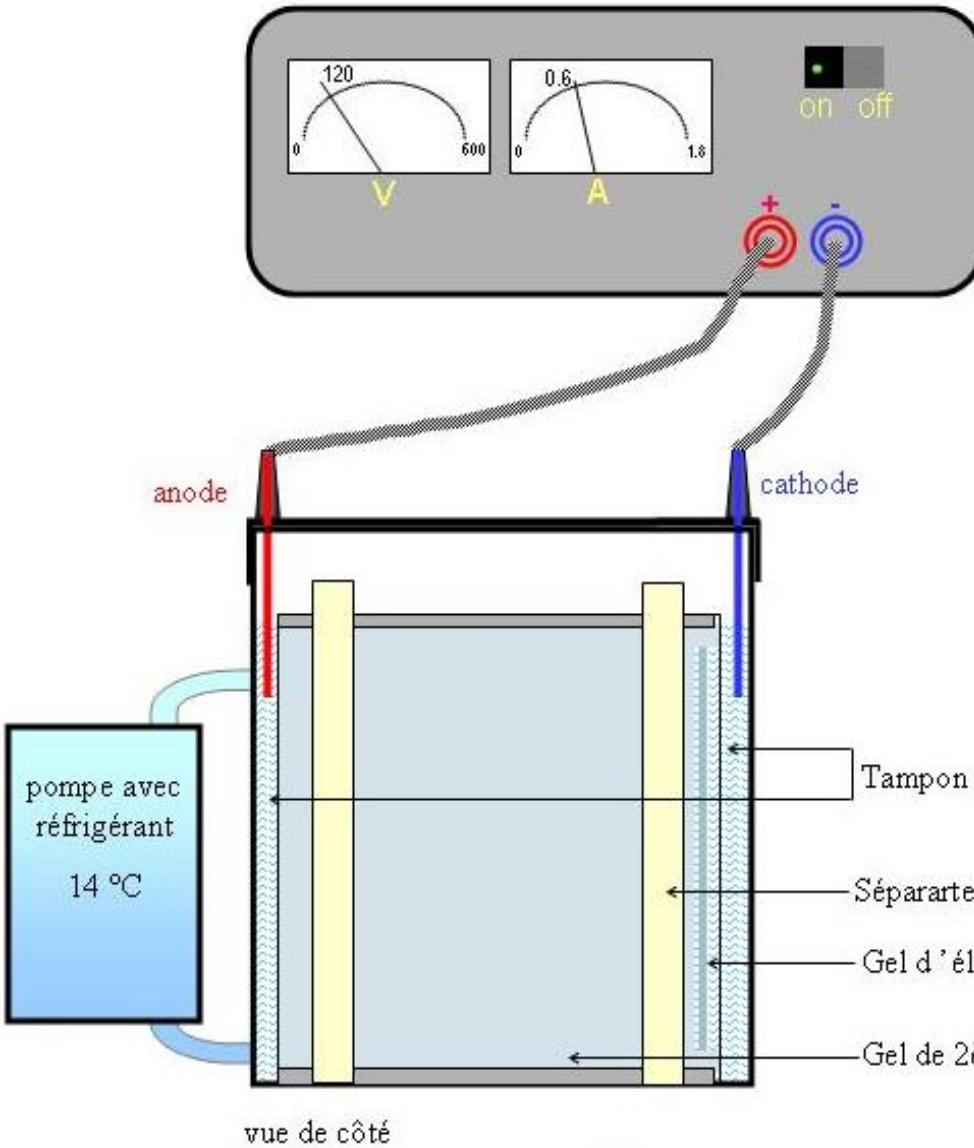
gel IEF-IPG  
déposé sur  
plaqué 2D.

Gel IEF

Gel 2D

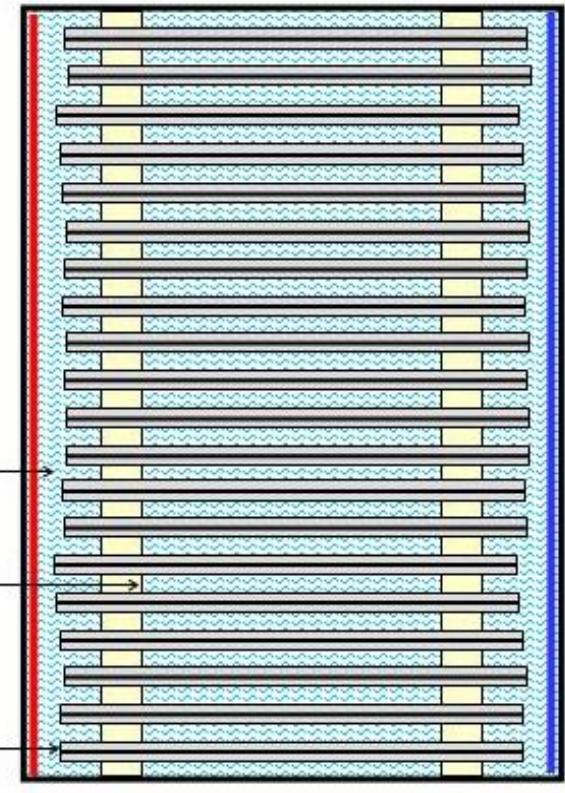


# SDS-PAGE: migration



Utilisation de bleu de bromophénol comme témoin de migration.  
Migration: 16h à 120V

→ 20 gels/migration



vue de côté

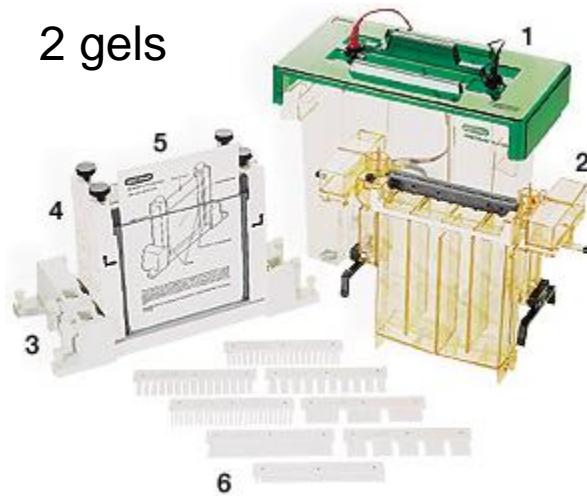
vue de dessus

Sens de migration des protéines

# SDS-PAGE: migration

Multitude de systèmes.

2 gels

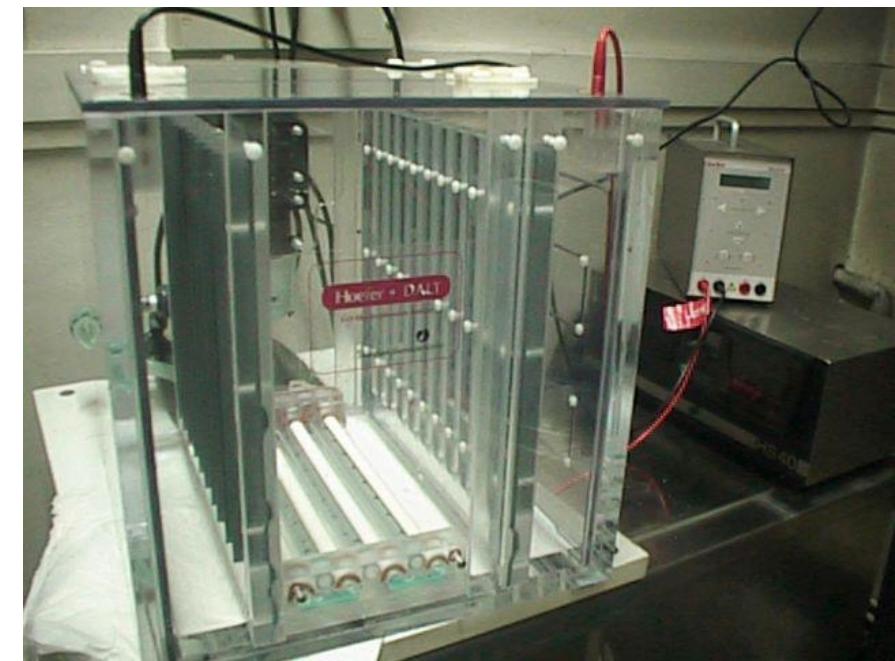


6 gels



20 gels

12 gels



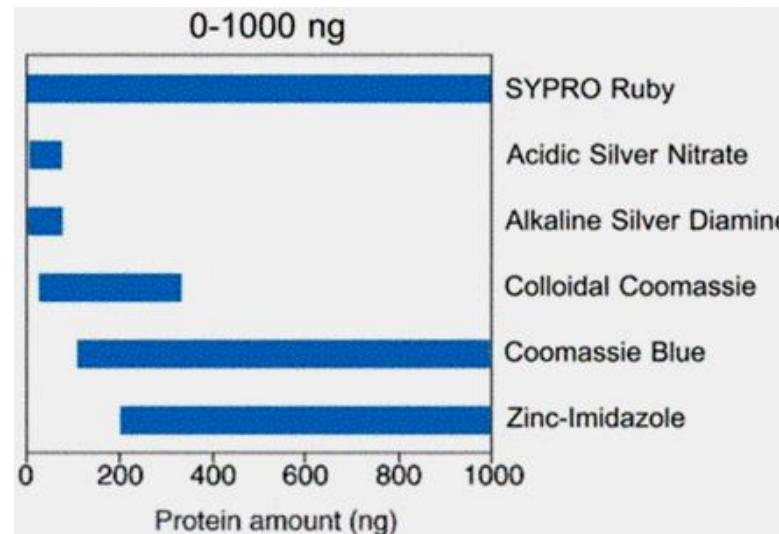
# Révélation des protéines

# Révélation des protéines: diverses méthodes de coloration

## Characteristics of Various Protein Stains

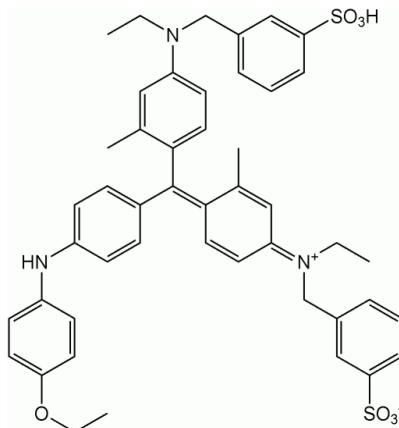
Stain	Sensitivity	Time Required	Advantage	Disadvantage	Recommended Application
Coomassie	100 ng	5–10 min staining 1–3 h destaining	Rapid Low cost After Coomassie stain, gel can be silver stained	Accumulates large volume of methanol present in staining and destaining solution	Routine work
EZBlue™ (Sigma)	2 ng	5–10 min	No fixation step	Expensive	Routine work
Colloidal Coomassie	10 ng	5–10 min	No destaining step	Fixation required	Routine work
CBB in acid	20 ng	30–60 min	No fixation step No destaining step	Longer staining time	Quantitation of protein in gel
Silver (alkaline method)	0.1 ng	2 h	Most sensitive, when sensitized with glutaraldehyde prior to staining	Complex reagent preparation, unstable reagent	To assess purity of protein preparation
Silver (acid method)	0.6–1.2 ng	90 min	Few steps Simple reagent preparation	Less sensitive than alkaline silver stain method	To assess purity of protein preparation
Zinc	5 ng	25–40 min	No fixation step Elution of unstained protein	Multiple steps when toning reaction intended	Peptide sequencing, antibody development
Nile Red	100 ng	As little as 6 min	Rapid	UV light box and camera required for documentation	Routine work
Calconcarboxylic acid	10 ng (during electrophoresis) 25 ng (post electrophoresis)	30–70 min	Migration of stained proteins during electrophoresis	Simultaneous staining is less sensitive than post staining	Routine work
Eosin Y	10 ng	30 min	Antigenicity of the stained protein retained	Transilluminator required	Antibody development
Procion blue	100 ng	1.5 h staining 48 h destaining	Proteins in gel can be quantitated	Time consuming	Quantitation of proteins
Amido Black	>100 ng	2–4 h	—	Less sensitive	—
Fast Green FCF	200 ng	2–4 h	—	Less sensitive	—

Détection	Limite de détection	Dynamique
Bleu de Coomassie	100ng	300
Argent	200pg	7
Fluorescence	10-100pg	10000
DIGE (pre-coloration)	250pg	10000
Autoradiographie	1pg	20

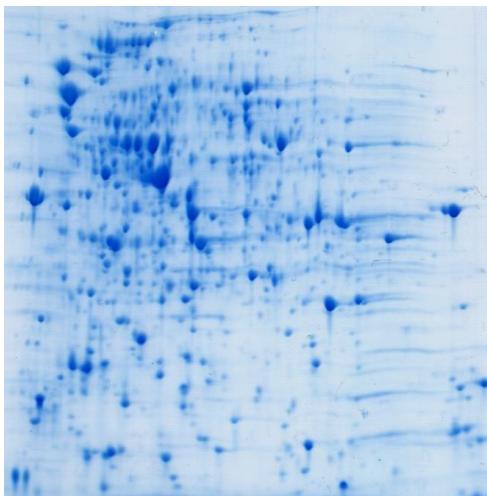


# Révélation des protéines: méthodes de coloration les plus courantes

CBB

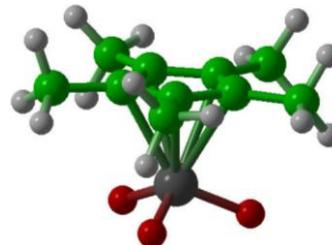


Le bleu de Coomassie se lie aux acides aminés basiques (principalement l'arginine, histidine) et aromatiques.



(+) pas de saturation  
(-) faible sensibilité

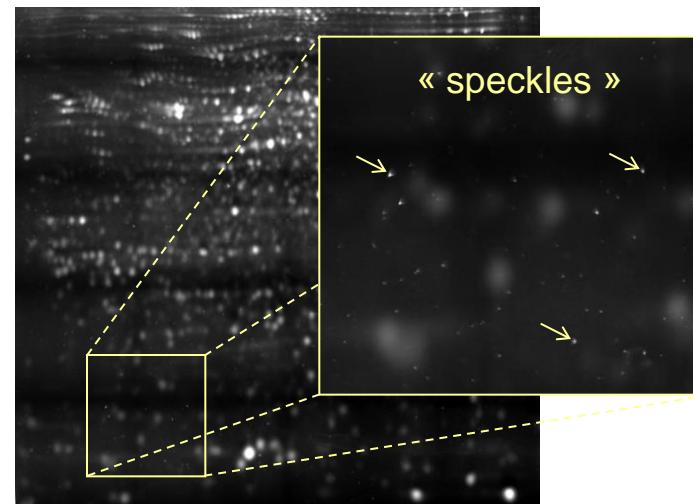
SYPRO Ruby



Fluorophore de type Ruthénium.

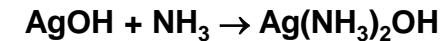
Interaction non-covalente (réversible), électrostatique et hydrophobe avec les protéines.

$\lambda = 610 \text{ nm}$



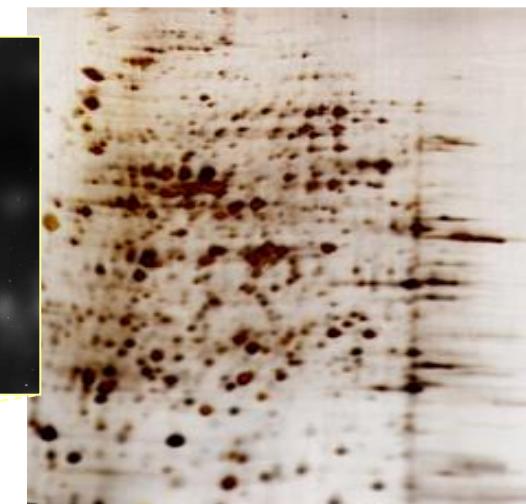
(+) pas de saturation, très sensible  
(-) onéreuse, « speckling », non visible (excision de spot), « photobleaching »

AgNO<sub>3</sub>



Réduction (formaldéhyde) des cations Ag<sup>+</sup> en argent métal.

Les groupes amines de la lysine et les résidus soufre des cystéines et méthionines réagiraient avec les cations Ag<sup>+</sup>.

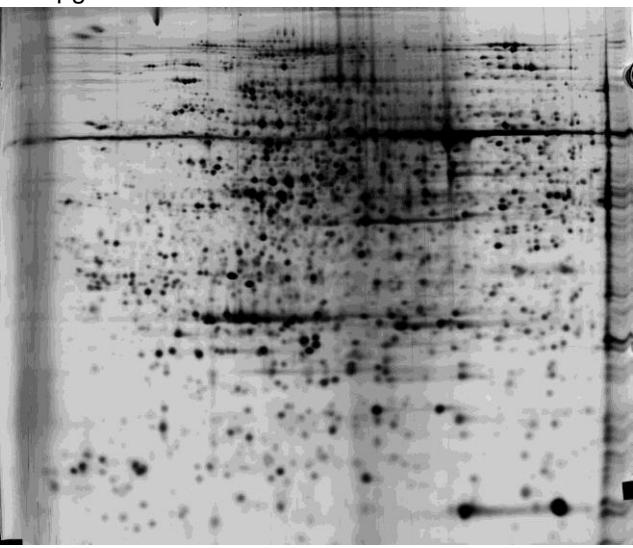


(+) sensibilité élevée  
(-) saturation 67

# Révélation des protéines: sensibilité des méthodes de coloration

Nitrate d'argent (AgNO<sub>3</sub>)

50 µg

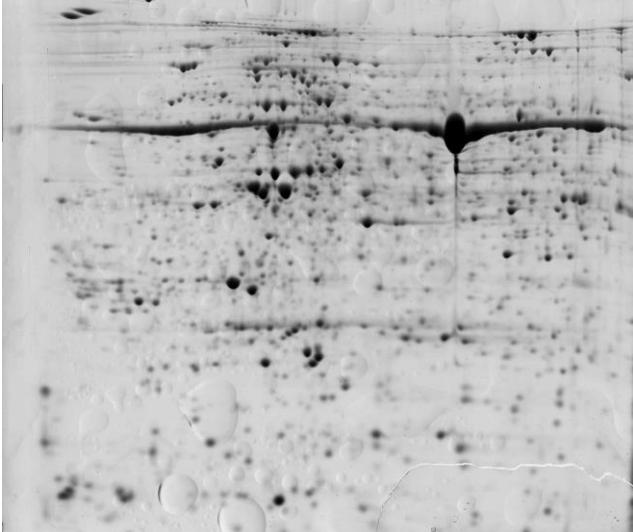


Bleu colloïdal (Coomassie Blue Brilliant CBB)

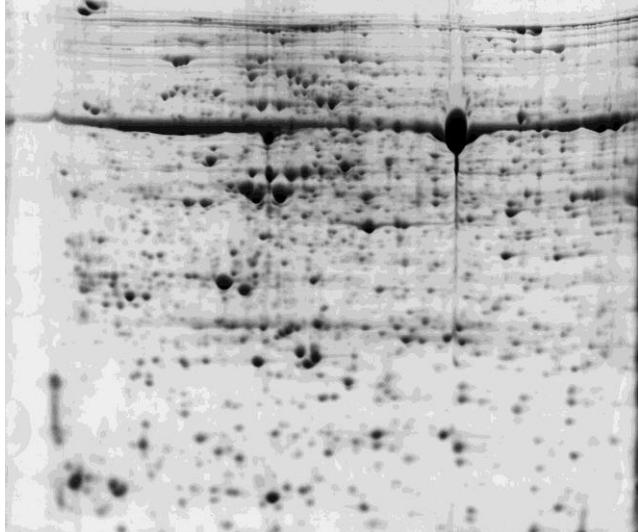
250 µg



500 µg



1000 µg



# Révélation des protéines: sensibilité des méthodes de coloration

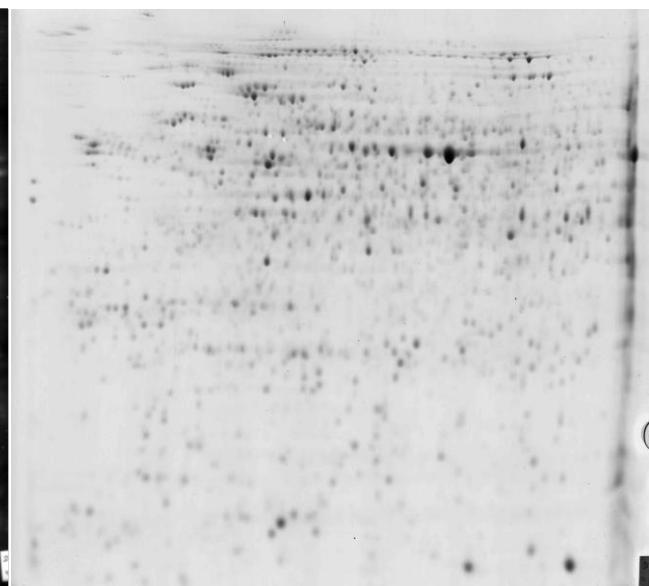
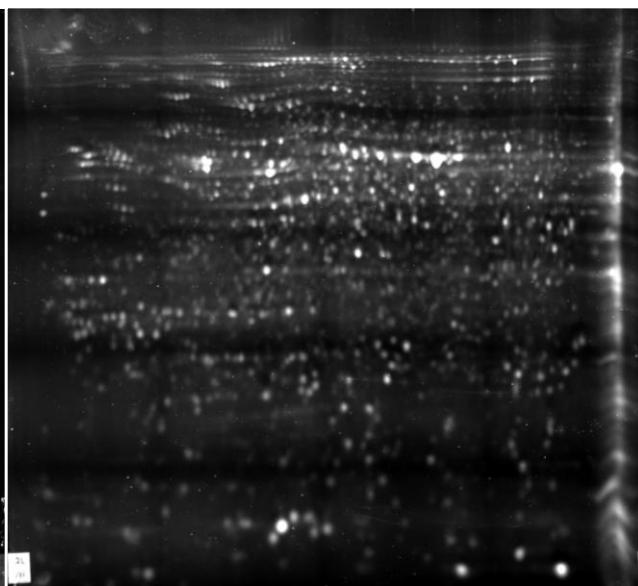
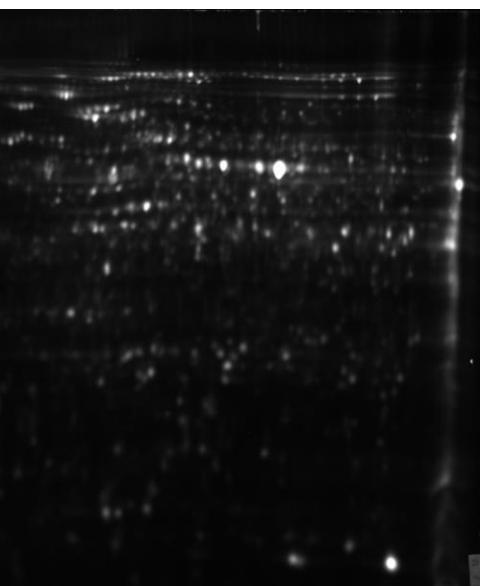


150 µg de protéines

DIGE cy5

sypro

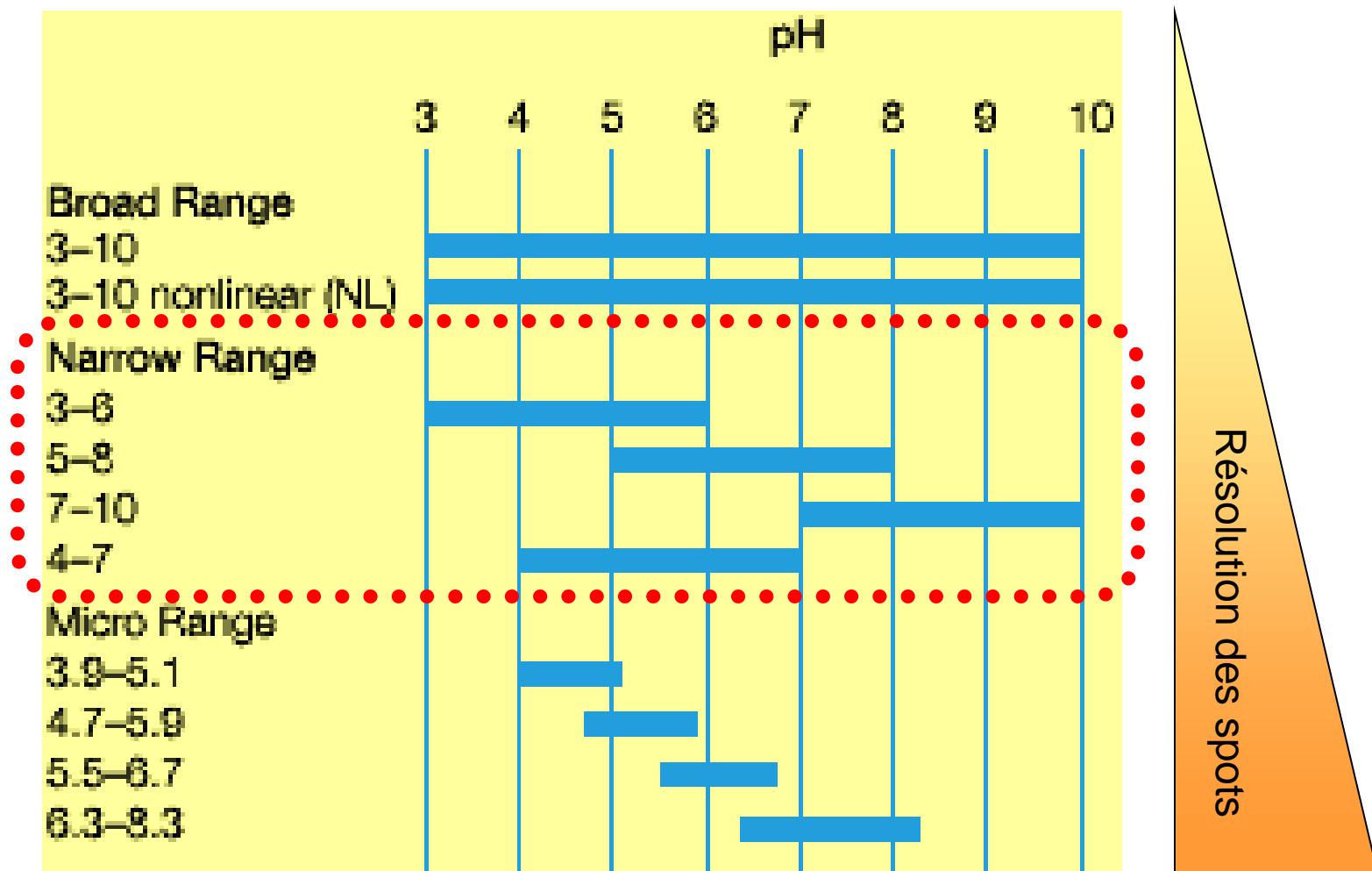
CBB



# Amélioration de la résolution des spots

## Gain de résolution: diminuer le gradient de pH

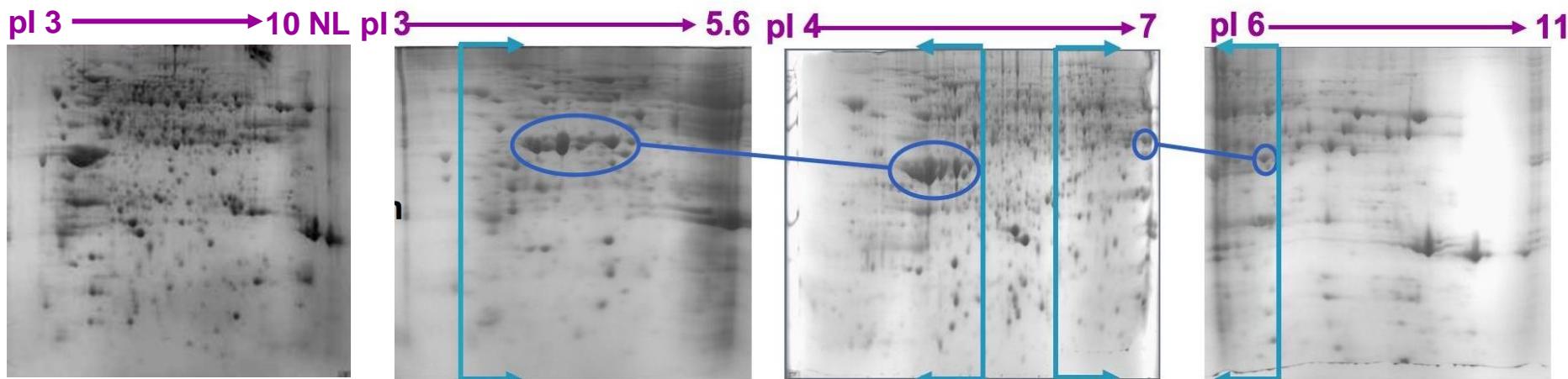
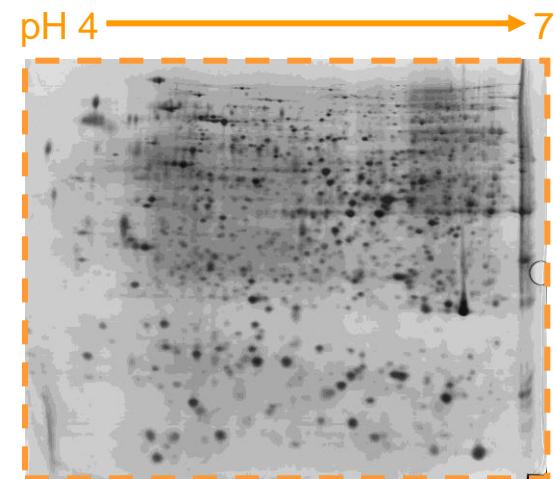
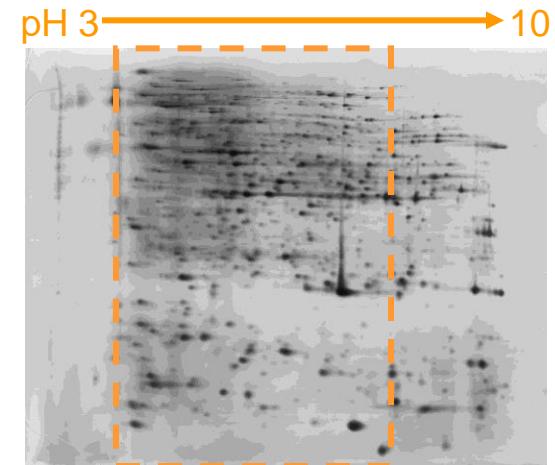
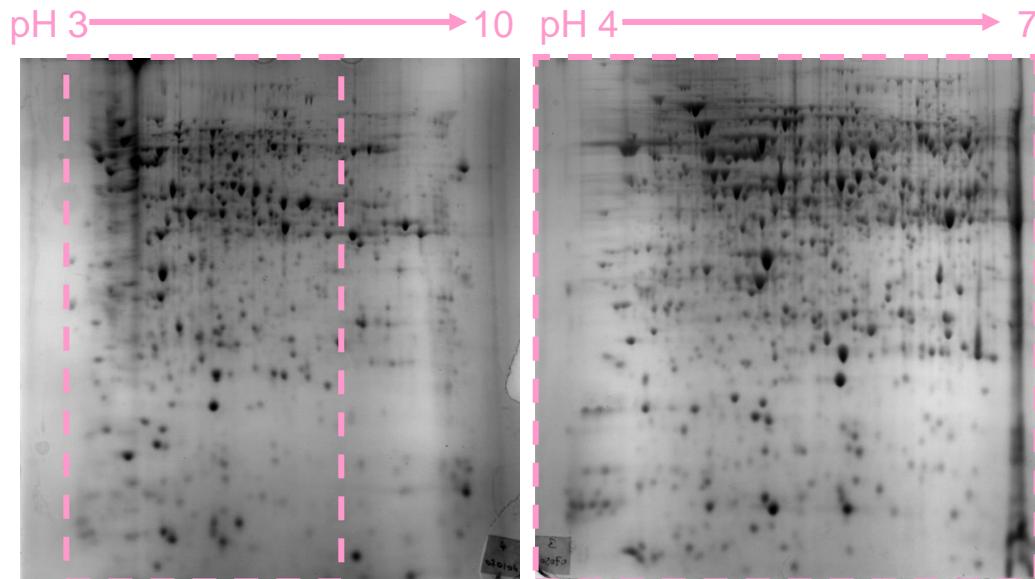
Gammes de pH commercialement disponibles:



**BIO-RAD**

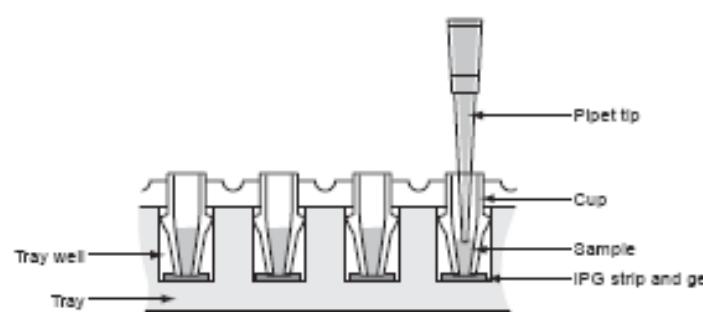
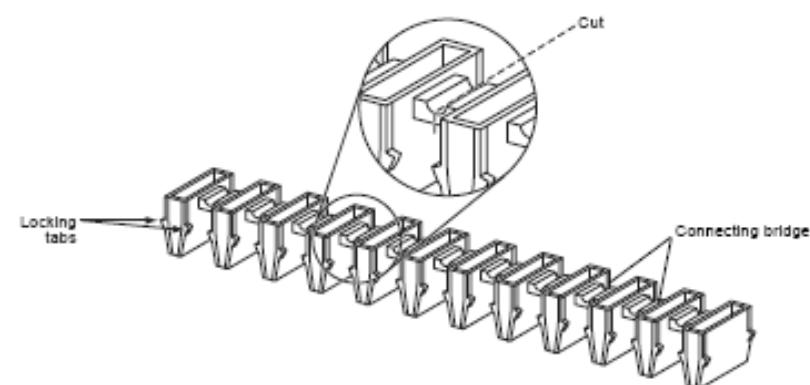
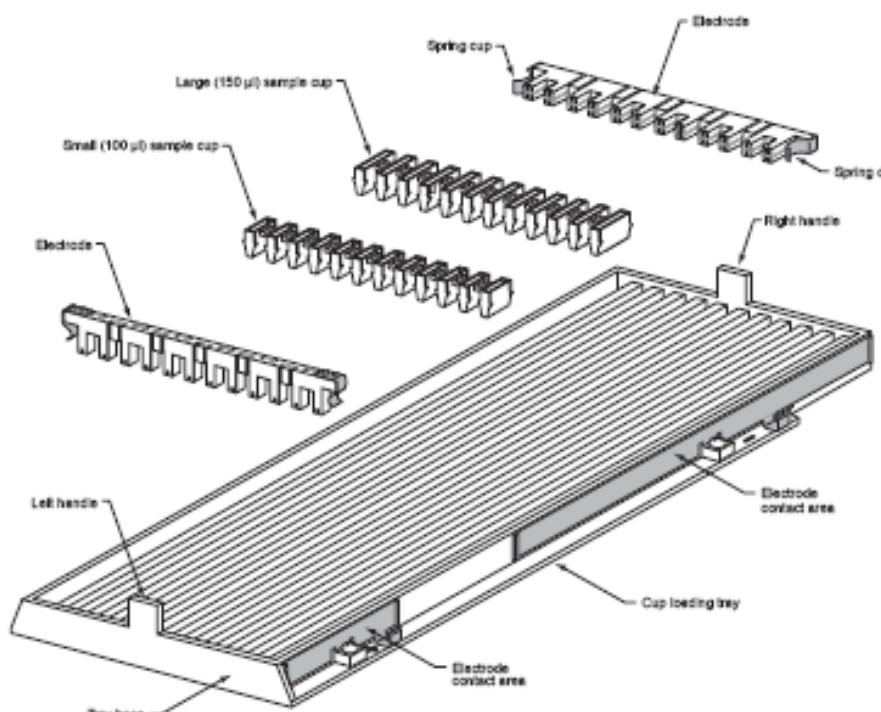
# Gain de résolution: diminuer le gradient de pH

Narrow-range IPG strips



# Gain de résolution: diminuer le gradient de pH

Pour les protéines basiques (pH 7-11): cup-loading



## Gain de résolution: diminuer le gradient de pH

Pour les protéines basiques (pH 7-11): cup-loading.  
Effet du Vh et de l'électroendosmose inversée.



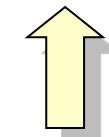
65000Vh

85000Vh

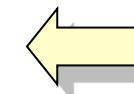
105000Vh

130000Vh

100000Vh



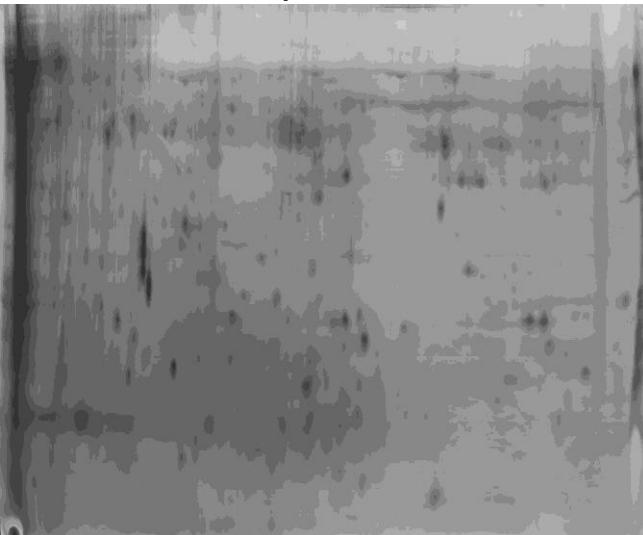
10% 2-propanol  
15% 2-propanol  
+ 5% glycerol



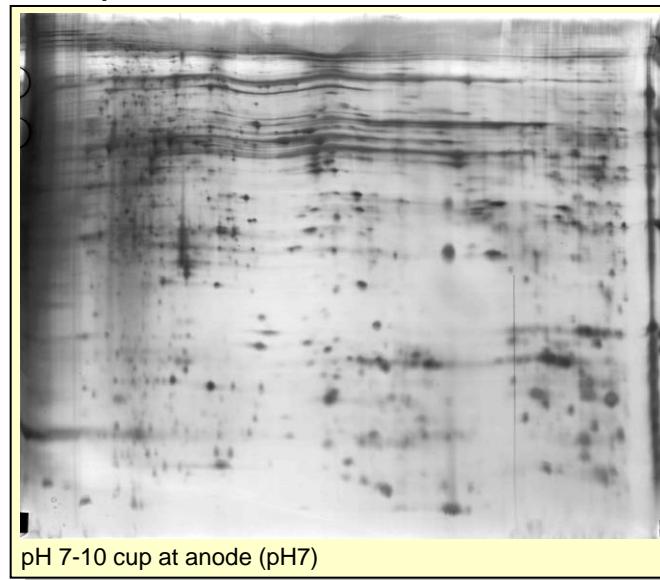
## Gain de résolution: diminuer le gradient de pH

Pour les protéines basiques (pH 7-11): cup-loading.

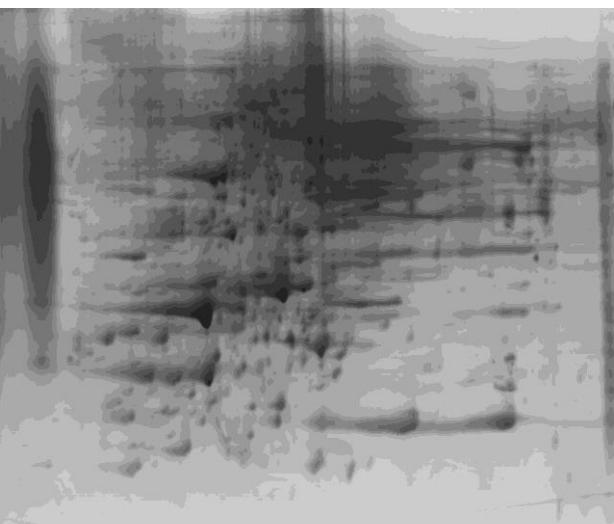
Effet de positionnement de la « cup ».



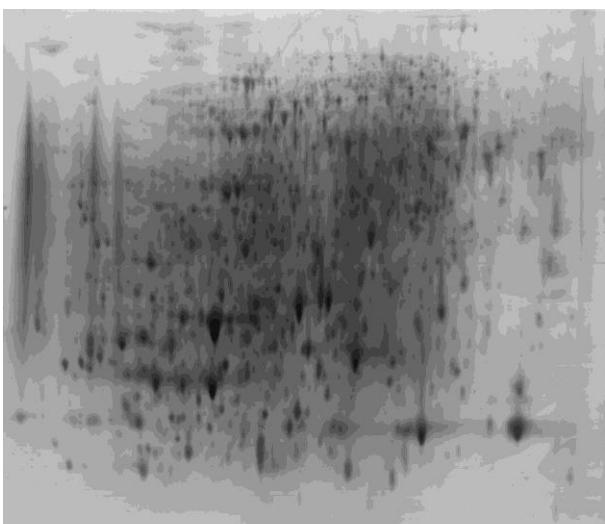
pH 7-10 cup at cathode (pH10)



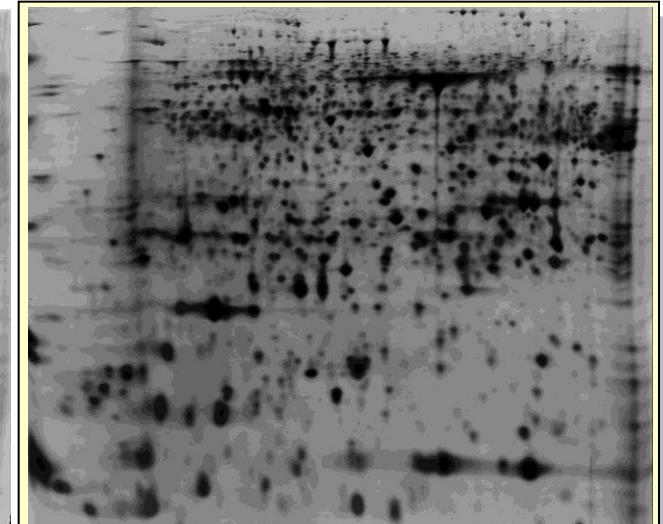
pH 7-10 cup at anode (pH7)



pH 4-7 cup at anode (pH4)



pH 4-7 cup at cathode (pH7)

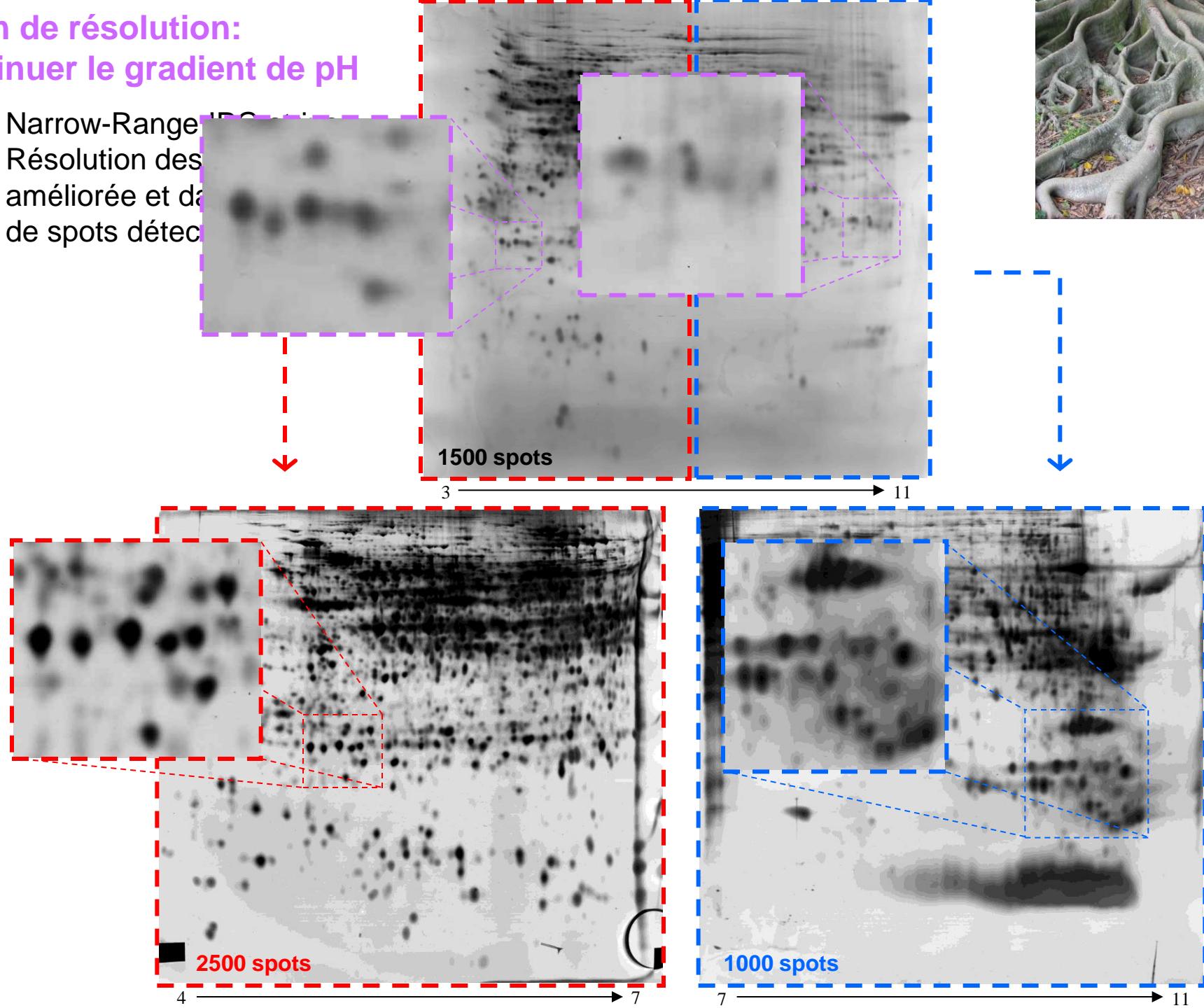


pH 4-7 in-gel rehydration

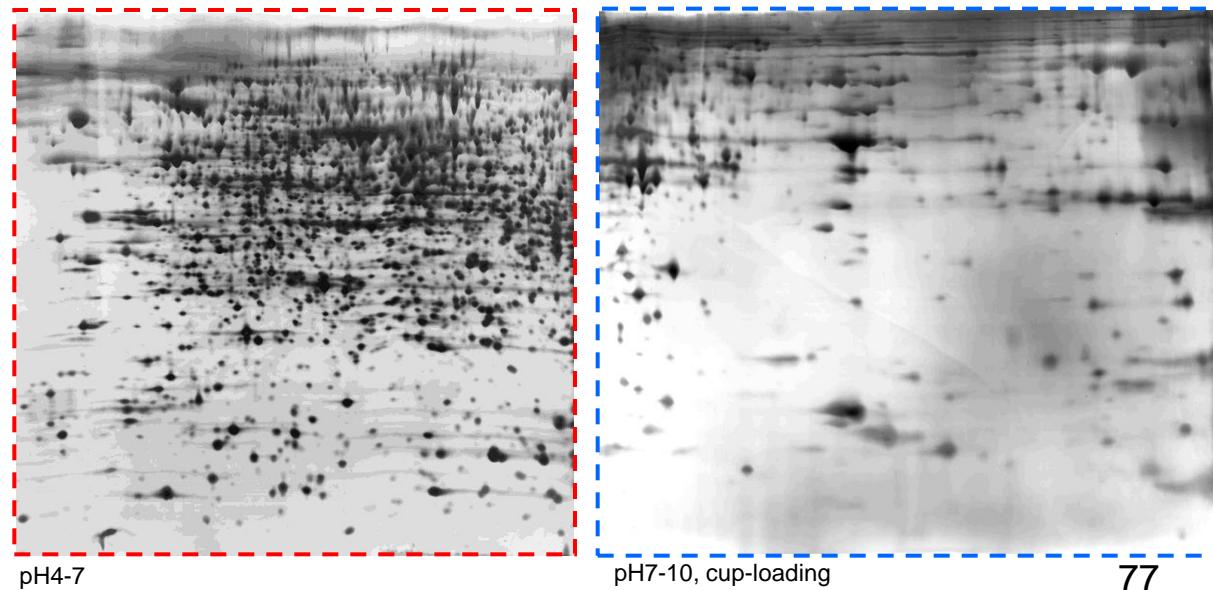
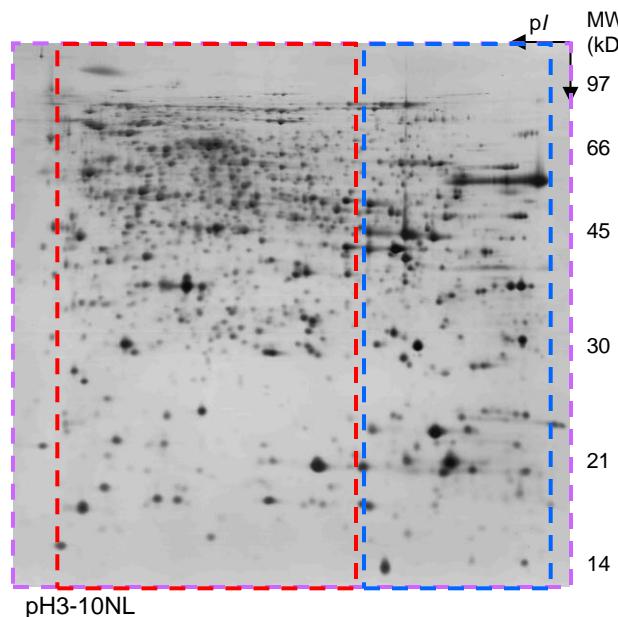
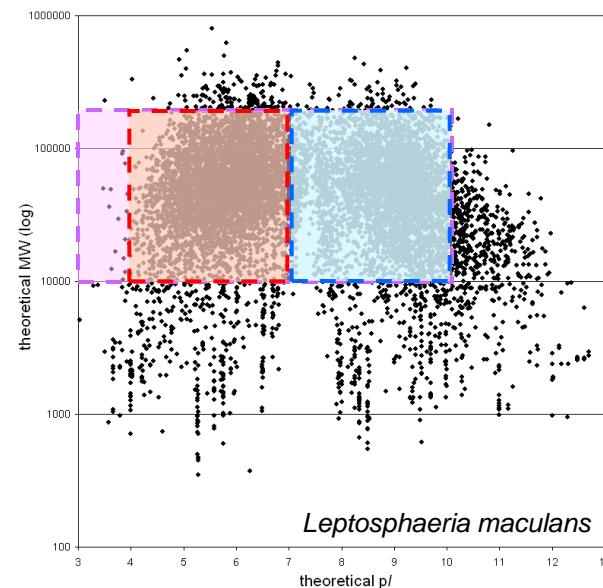


## Gain de résolution: diminuer le gradient de pH

Narrow-Range  
1D Gels  
Résolution des  
améliorée et da  
de spots détecte



# Gain de résolution: diminuer le gradient de pH

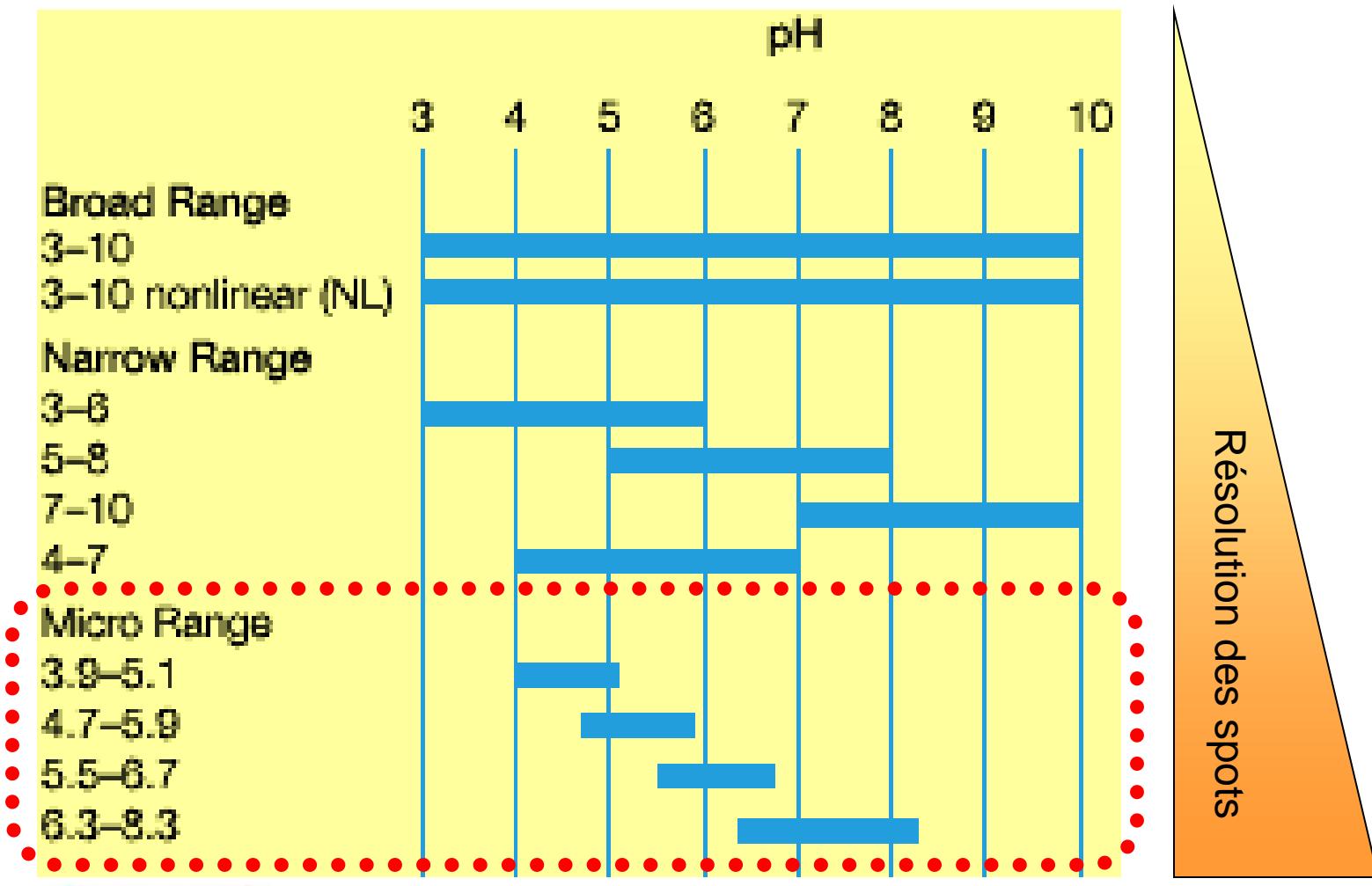


Net gain en résolution des protéines acides.

Protéines basiques récalcitrantes à la 2-DE.

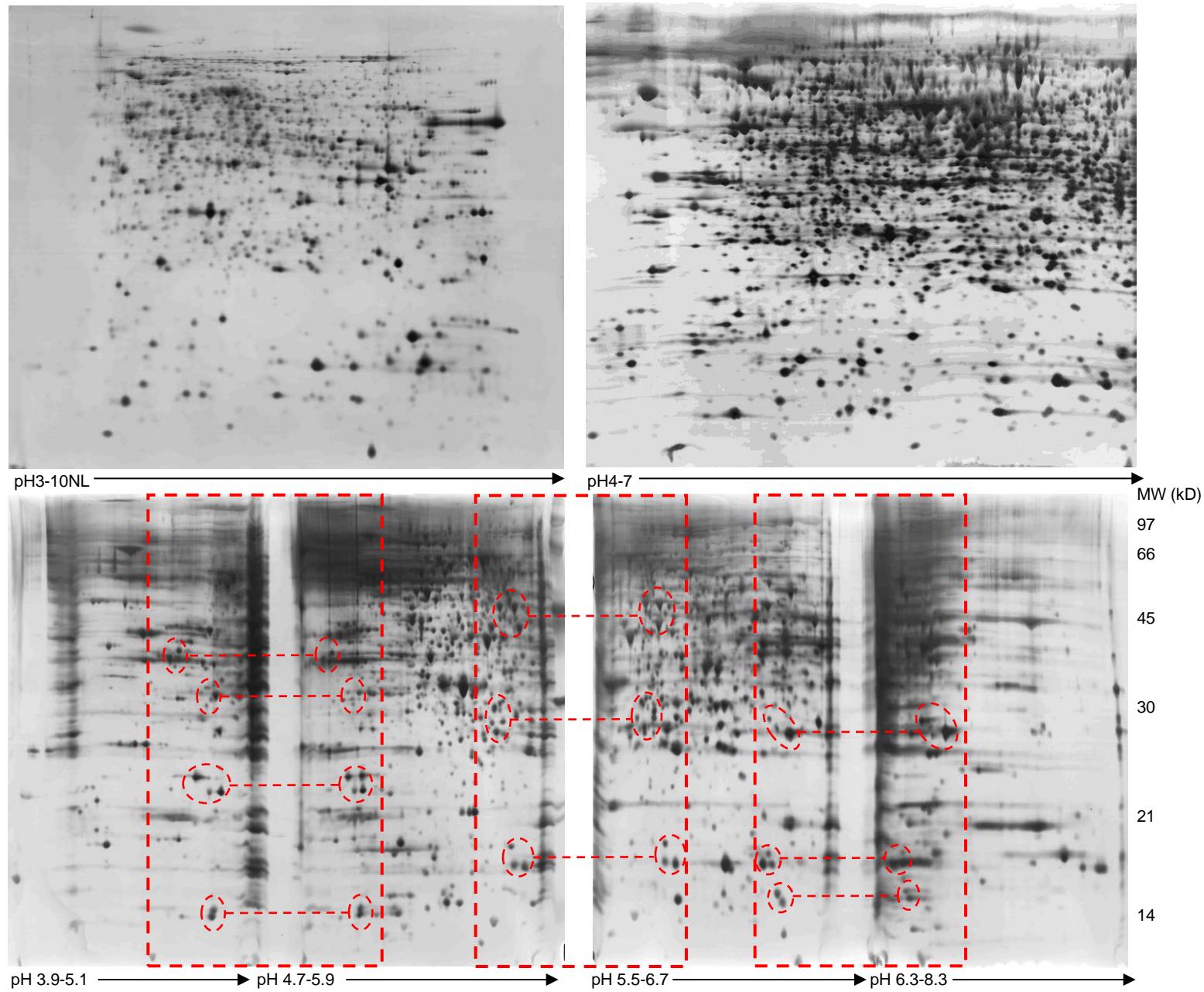
## Gain de résolution: diminuer le gradient de pH

Gammes de pH commercialement disponibles:



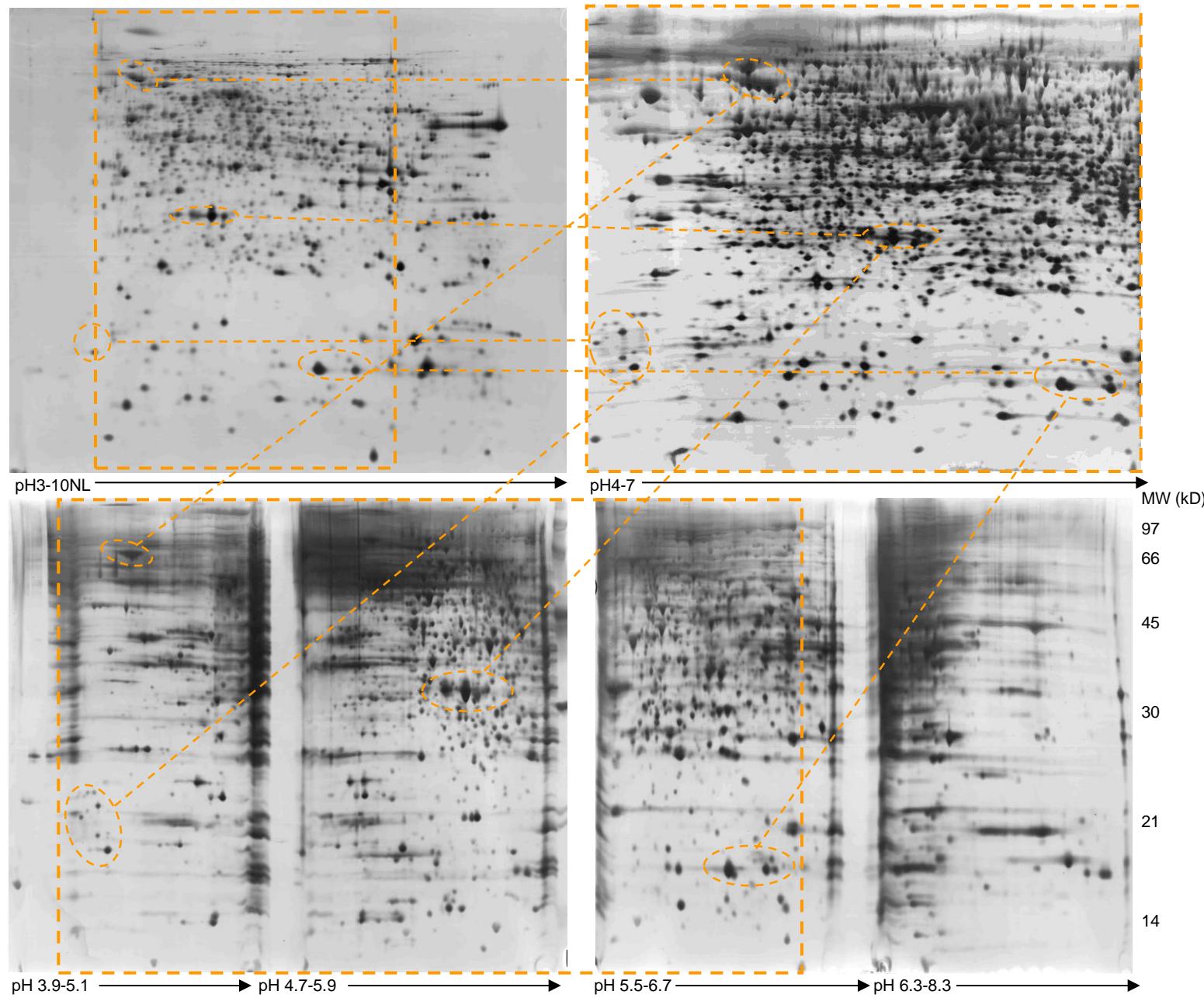
## Gain de résolution: diminuer le gradient de pH

### Micro-Range IPG strips



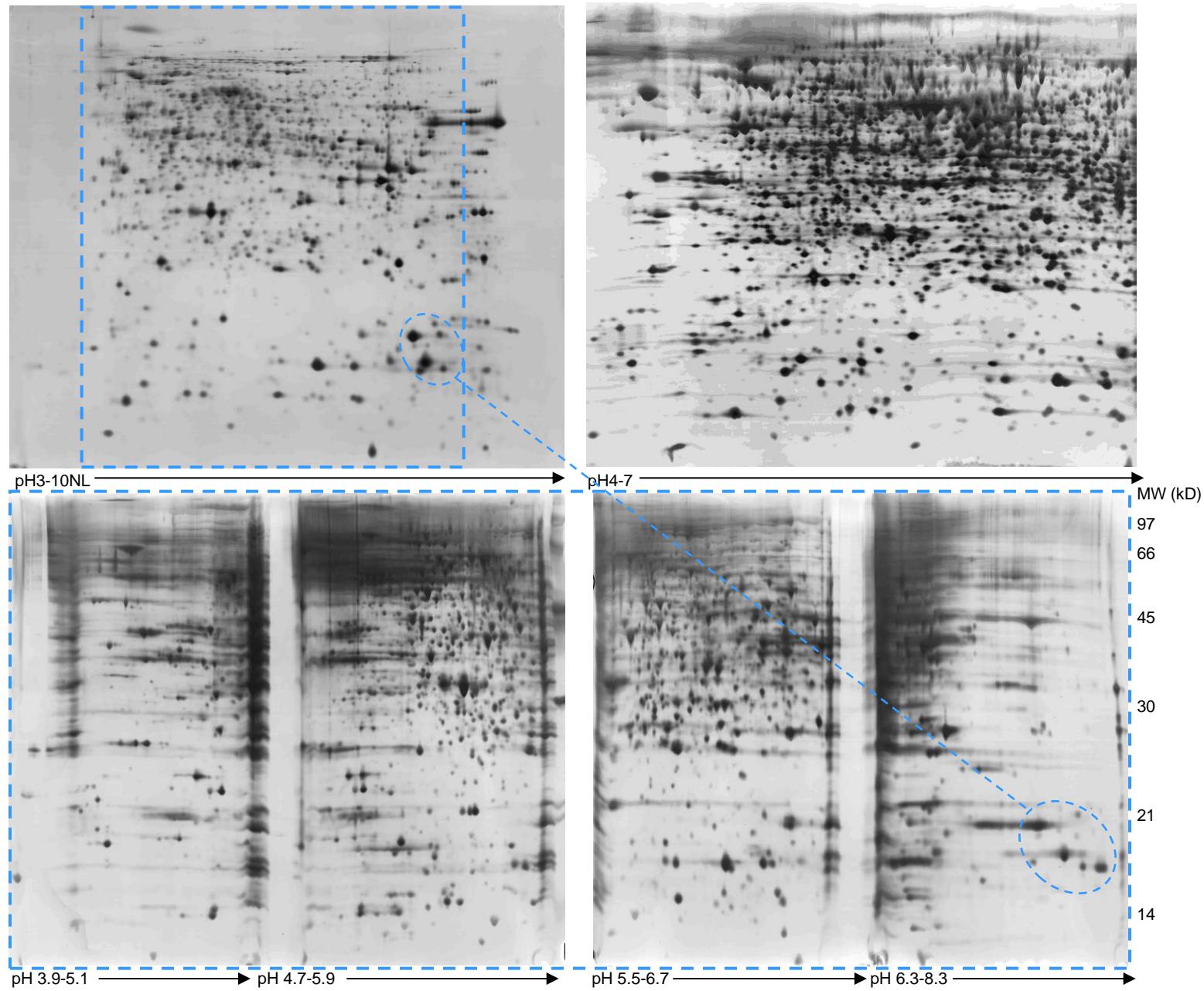
# Gain de résolution: diminuer le gradient de pH

## Micro-Range IPG strips

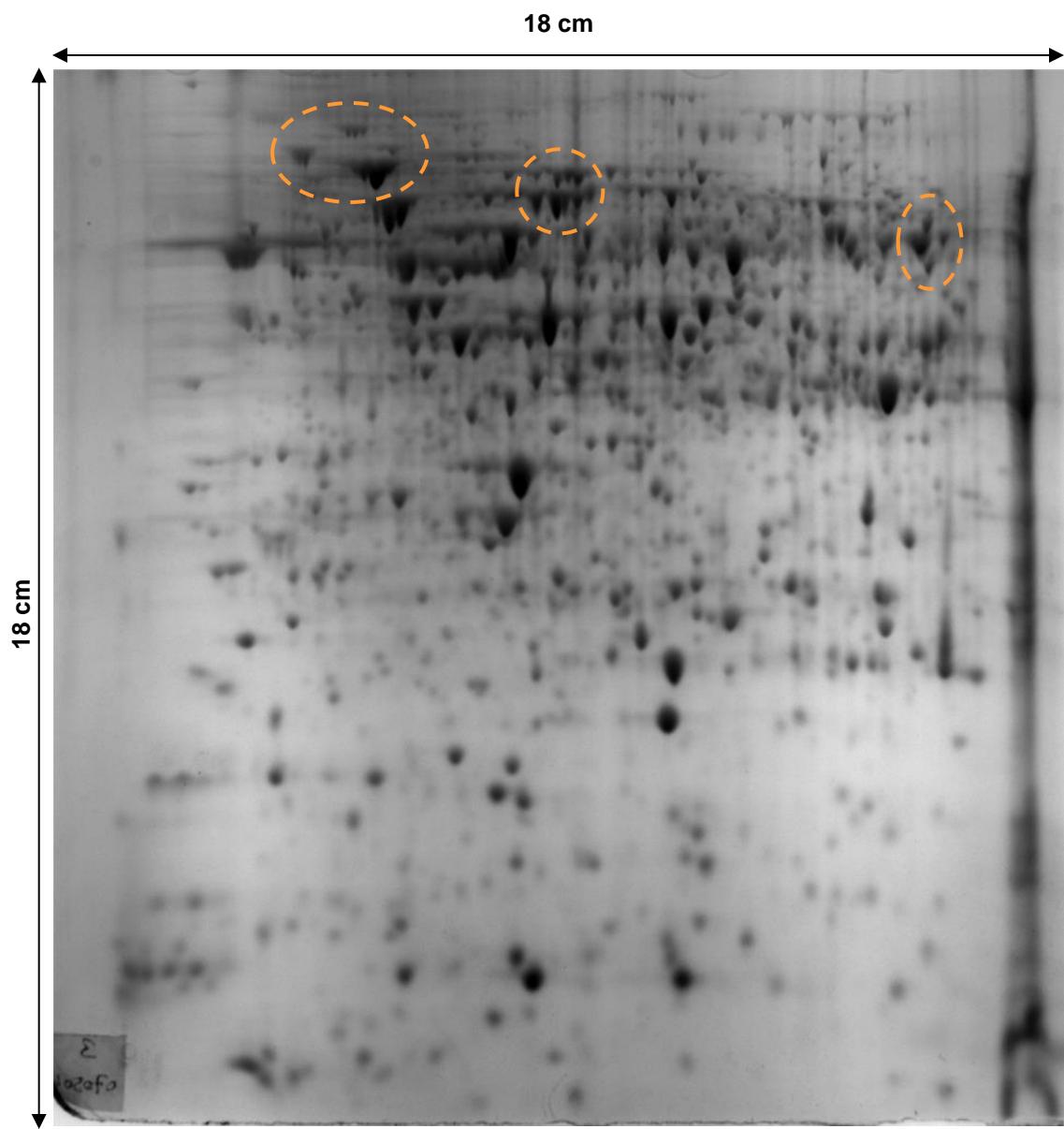
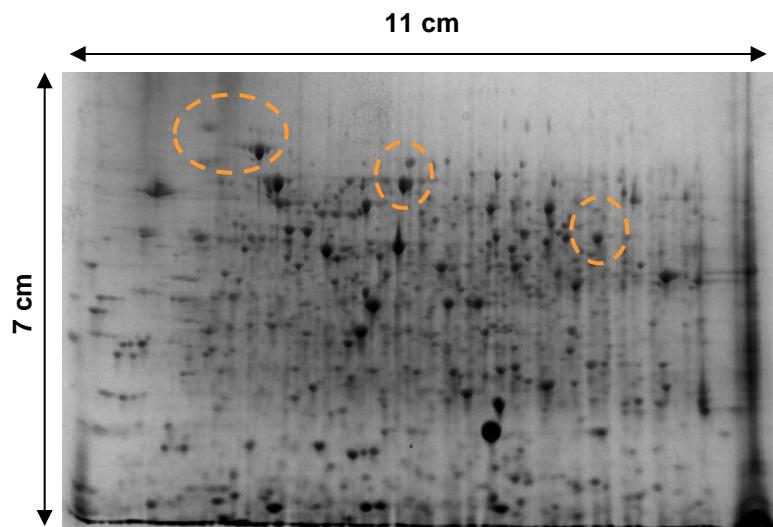


## Gain de résolution: diminuer le gradient de pH

Micro-Range IPG strips



## Gain de résolution: augmenter la taille des gels



Grands gels en routine:

24 x 20 cm

Possibilité de faire des très grands  
gels (mais pas en routine !):  
70 x 20 cm

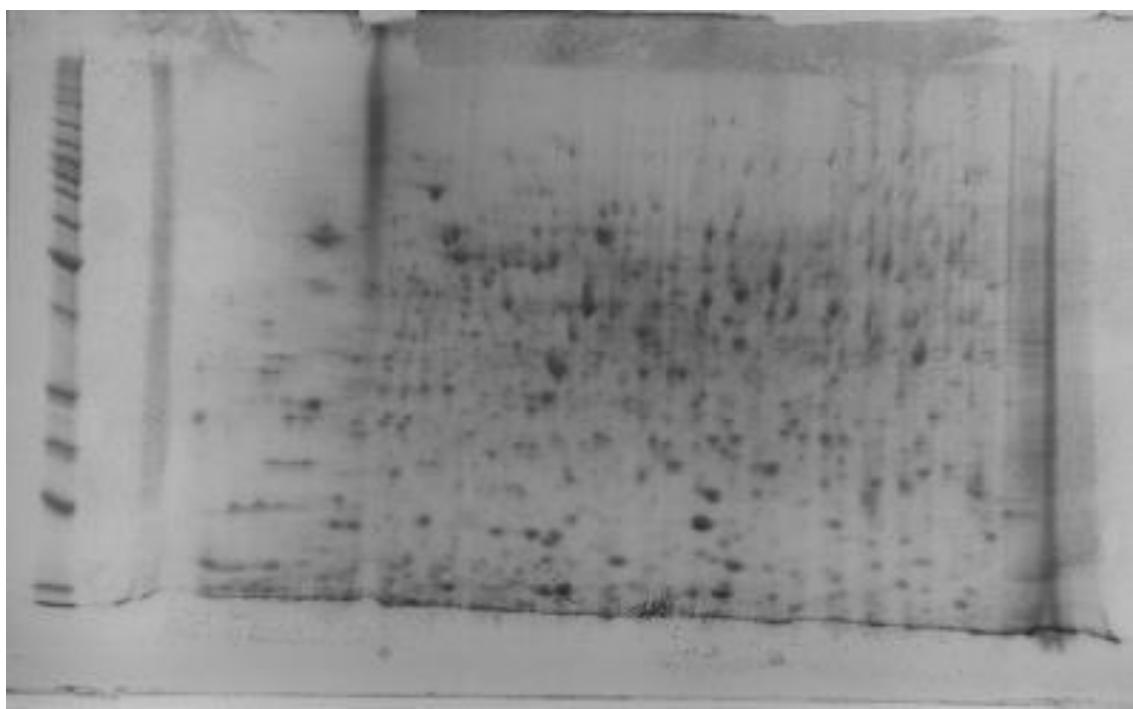
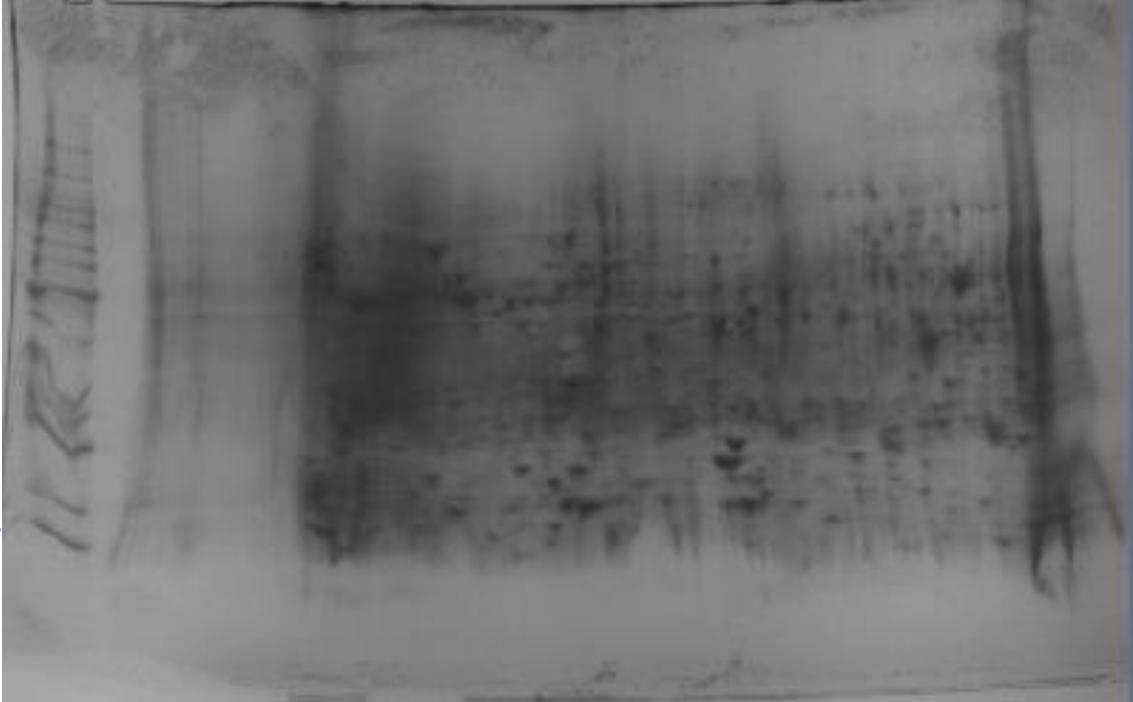
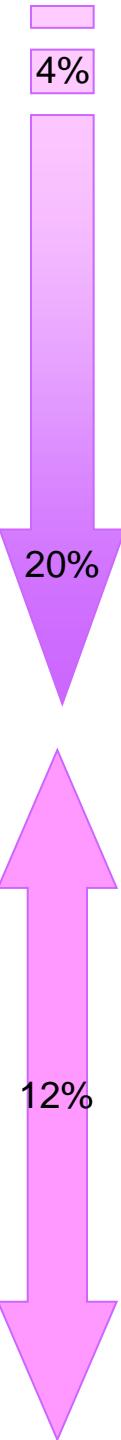


## Gain de résolution: utiliser des gels à gradient

### Objectif:

récupérer les protéines de PM très hauts ( $>100\text{kD}$ ) et très bas ( $<10\text{kD}$ ).

Toutefois, dans ce cas, ce n'est pas efficace.



# Avantages et limites de la 2-DE

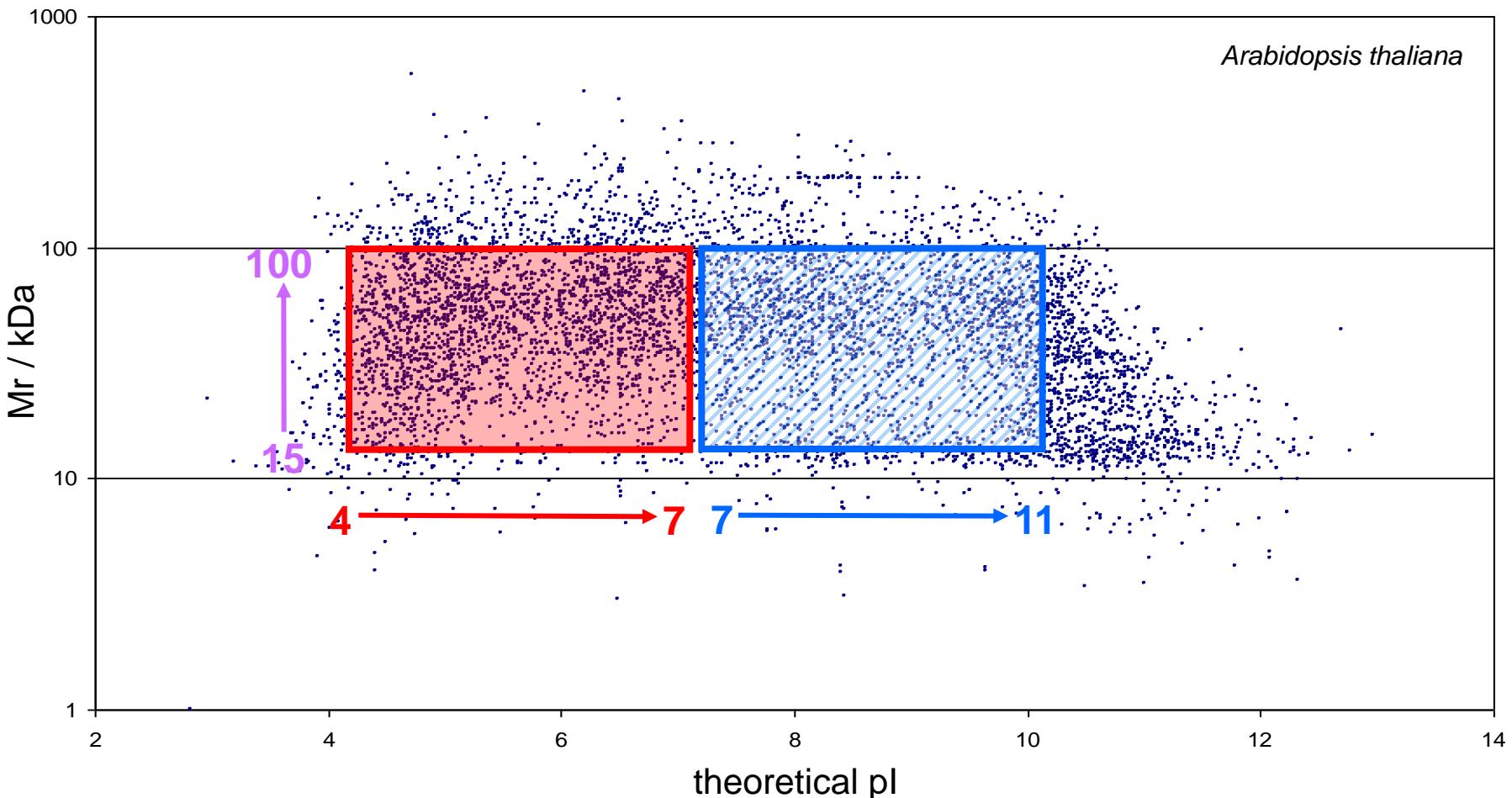
## Avantages et limites de la 2-DE

### Limites:

Faible représentation des protéines de haut PM ( $> 100$  kD)

Faible représentation des protéines de bas PM ( $< 10$  kD)

Faible représentation des protéines hydrophobes et/ou basiques



### Avantage:

Permet de séparer les isoformes de protéines si les modifications affectent la charge et/ou le poids.

# Avantages et limites de la 2-DE: isoformes de protéines

## Modification post-traductionnelle (PTM):

modification chimique d'une protéine, le plus souvent effectuée par une enzyme après sa traduction

### 1-Addition d'un groupe fonctionnel

l'acétylation, soit l'addition d'un groupe acide

l'alkylation, soit l'addition d'un group

la méthylation est l'additio

acides aminés lysine ou a

la biotinylation, qui est l'acylation d'u

la glutamylation, qui est la liaison co

tubuline ou un autre protéine.

la glycation, qui est la liaison cova

glycine sur la partie C-terminale de

la glycosylation, soit l'addition d'un g

hydroxyllysine, sérine, ou thréonine,

l'isoprénylation, soit l'addition d'un g

geranylgeranoli) :

la prénylation

la myristylation

la farnésylation

la géanyl-géryylation

la lipoylation

la phosphopantetheinylation, ou add

coenzyme A.

la phosphorylation, qui est l'addition

une séleine, tyrosine, thréonine ou hi

la sulfation, qui est l'addition d'un gr

la sélénation.

l'amidation en C-terminal.

### 2-Addition d'autres protéin

l'ubiquitination, qui est une liaison co

acceptrice.

la sumoylation, qui est une liaison co

sur une lysine acceptrice.

la Neddylation, qui est une liaison co

lysine acceptrice.

l'urmylation.

### 3-Changement de la nature chimique des acides aminés

la citrullination, ou déimination soit la conversion d'arginine en citrulline.

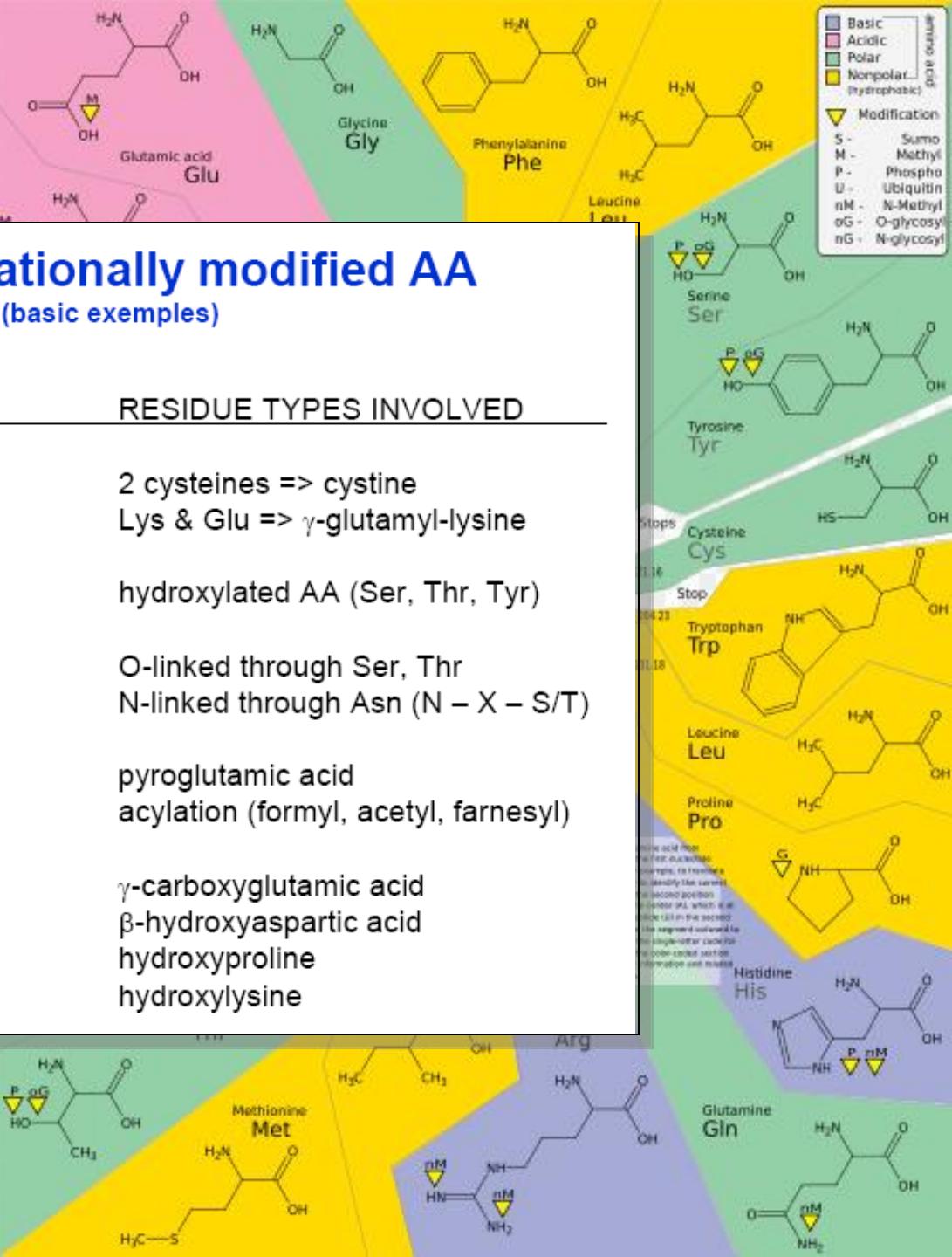
la déamidation, soit la conversion de glutamine en acide glutamique ou de

l'asparagine en acide aspartique

### 4-Changement de structure

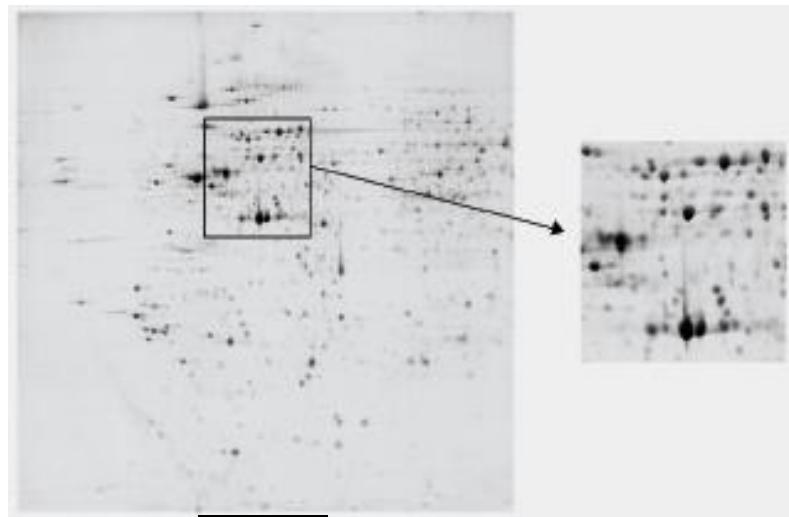
les ponts di-sulfures qui résultent dans la liaison chimique des groupement SH3 de deux cystéines.

Les clivages protéiques par une protéase d'un lien peptidique.

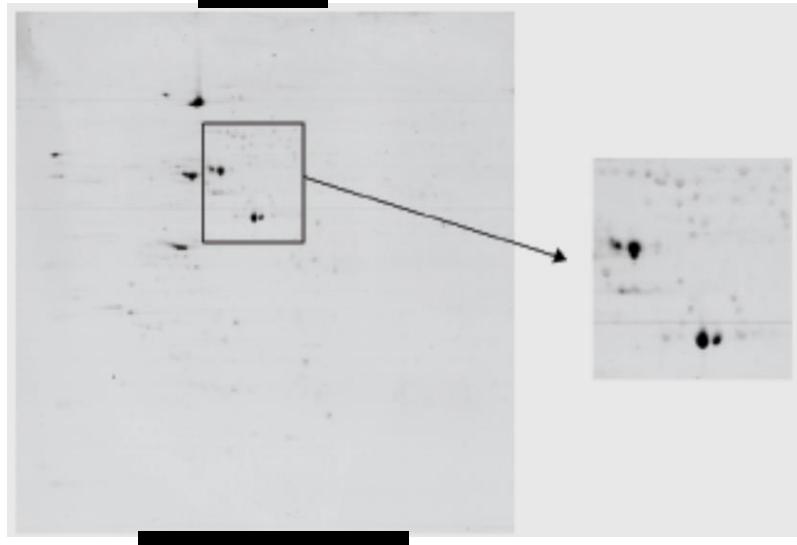


# Avantages et limites de la 2-DE: isoformes de protéines, cas des phospho- et glycoprotéines.

Phosphoprotéines: Pro-Q Diamond

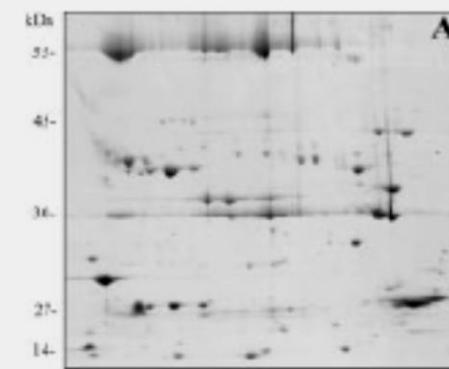


Sypro

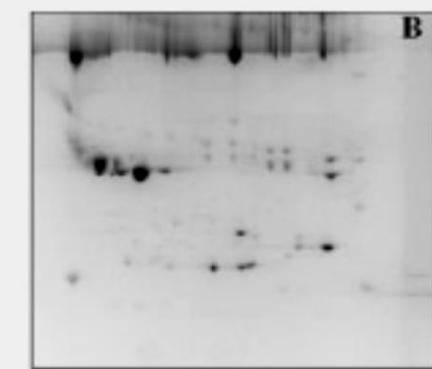


Pro-Q diamond

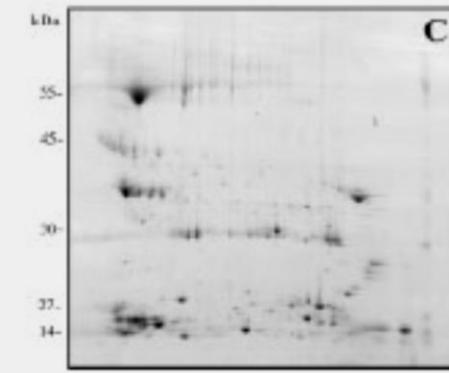
Glycoprotéines: Pro-Q Emerald



A

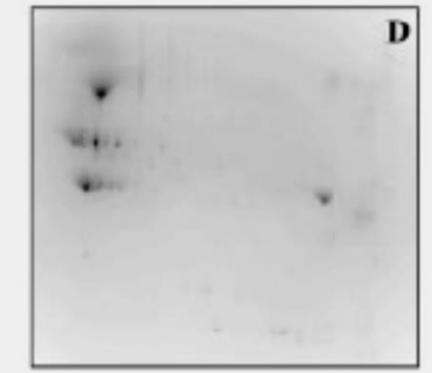


B



C

CBB



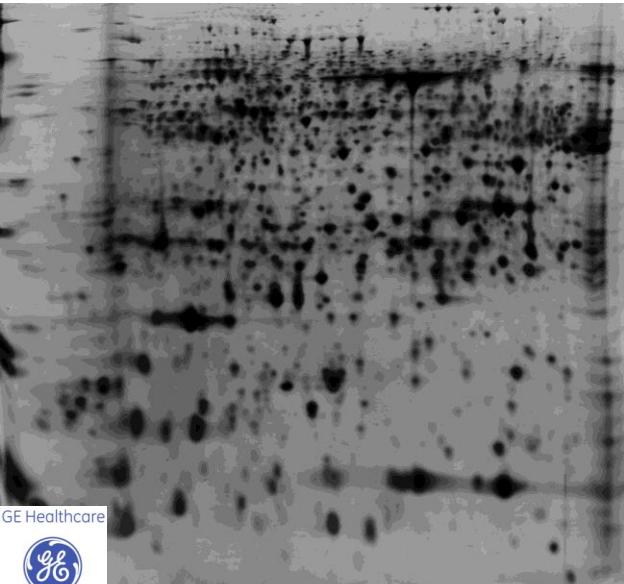
D

Pro-Q emerald

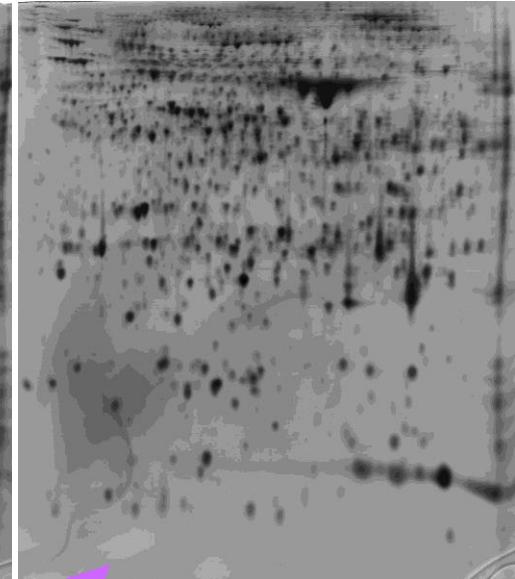
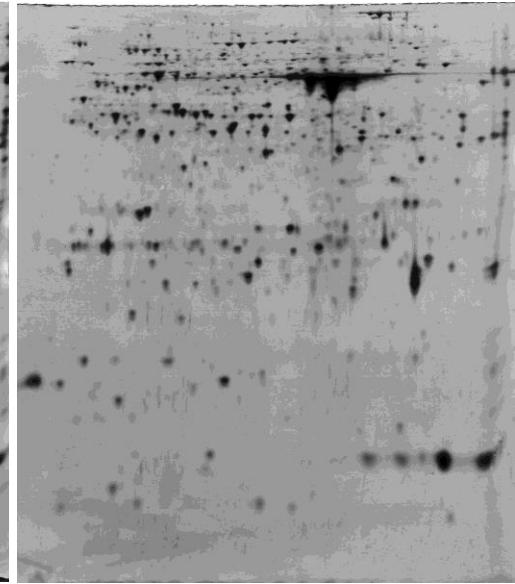
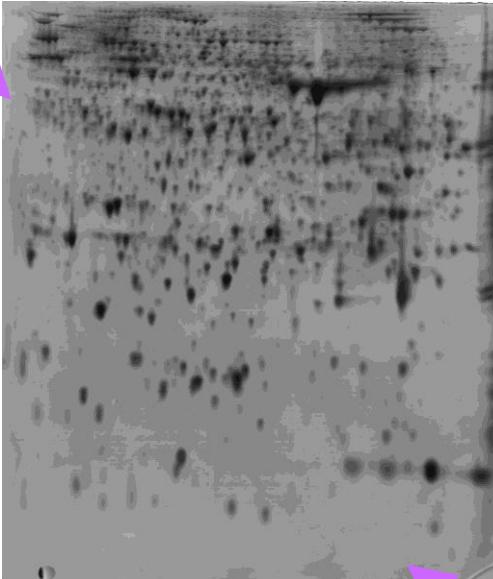
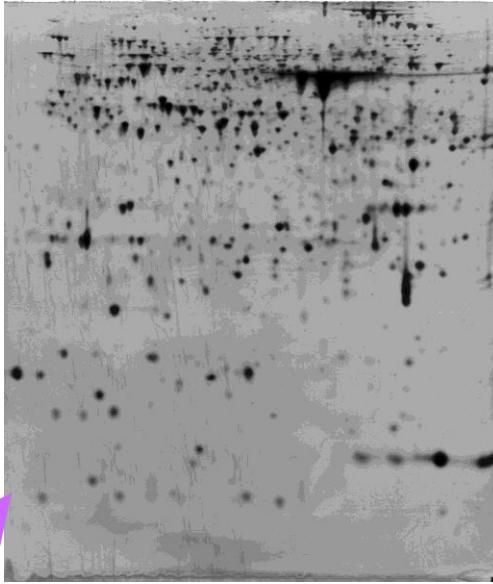
# Avantages et limites de la 2-DE: quantitatif mais reproductibilité limitée

Variations dues aux fournisseurs:  
même extrait, IPG strip pH 4-7, 24 cm

BIO-RAD



Répétitions biologiques/techniques

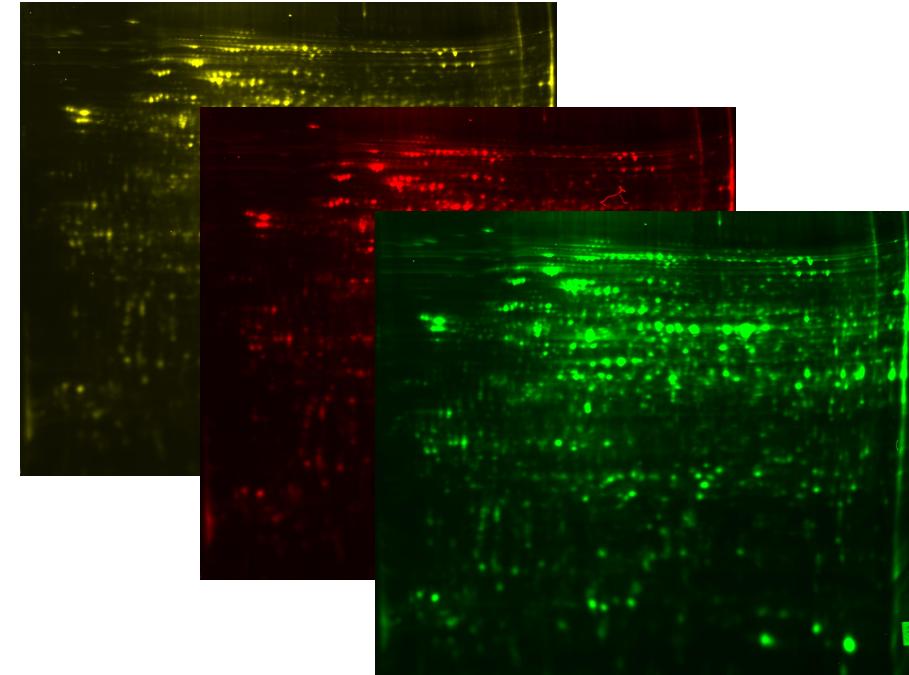


rep. biol.

## Avantages et limites de la 2-DE: sources de variation à l'origine du manque de reproductibilité

- Extraction de protéines (broyage plus ou moins fin, réchauffement possible, résolution plus ou moins complète)
- Stockage (-80°C, échantillon peut se dégrader au cours des mois)
- Dosage de protéines (pipetage)
- - IEF (pipetage, fabricants, lots, réhydratation)
- - SDS-PAGE (coulage, temps de polymérisation, temps de migration)
- - Coloration (AgNO<sub>3</sub> en particulier, temps de développement)

Solution:  
Difference In Gel Electrophoresis (DIGE)



# Marquage des protéines

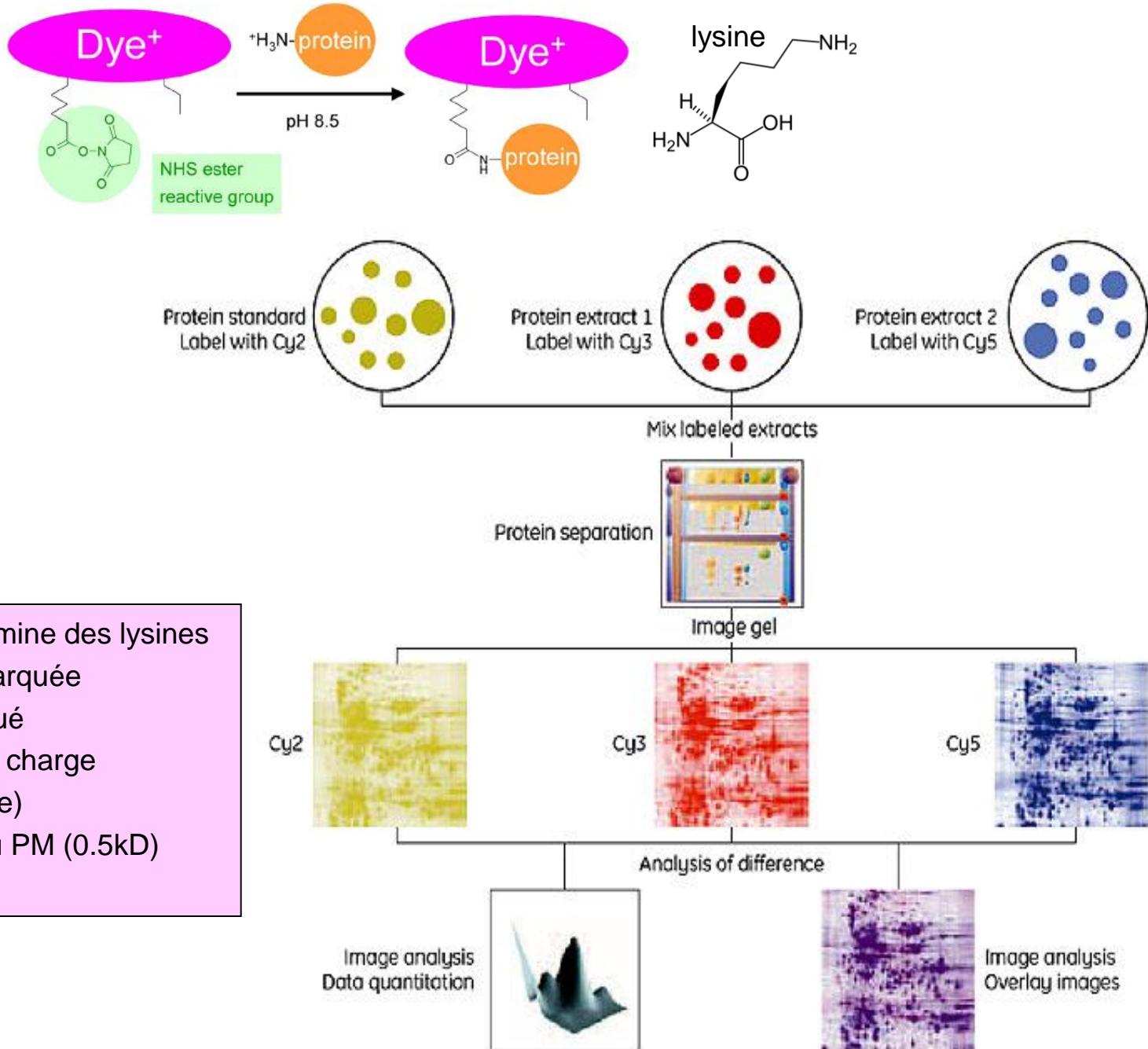
## 2-D DIGE

(Two-Dimensional  
Difference in Gel Electrophoresis)

## 2-D DIGE: principe (minimal labelling)



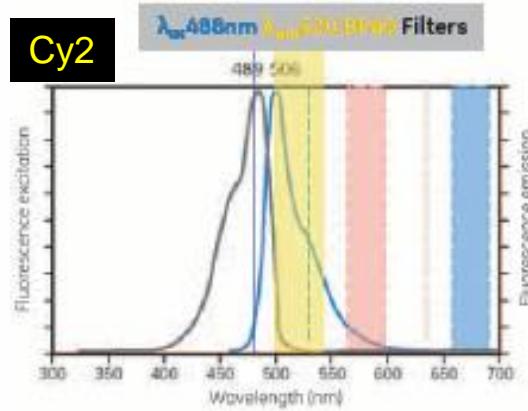
Cy2 Cy5 Cy3  
434D 466D 464D



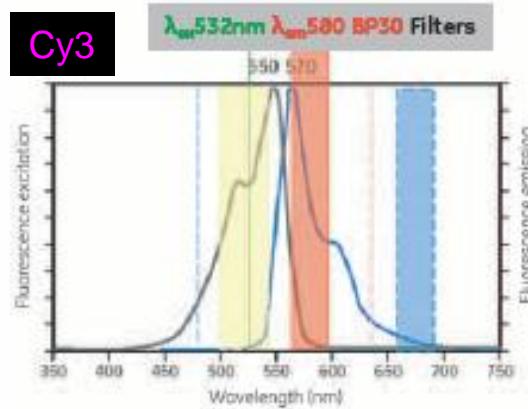
Marquage du groupe amine des lysines  
1 lysine/protéine est marquée  
1-2% du spot est marqué  
Pas de changement de charge  
(substitution équivalente)  
Faible augmentation du PM (0.5kD)  
Très onéreux !

## 2-D DIGE: principe (minimal labelling)

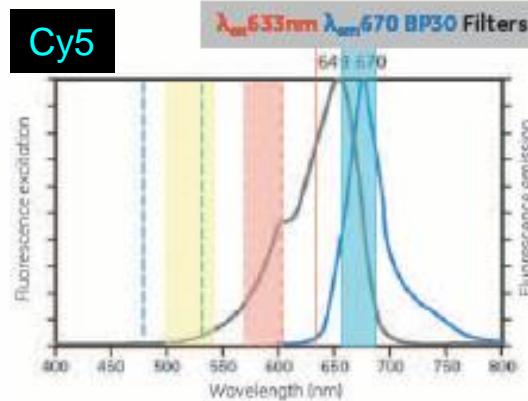
Cy2



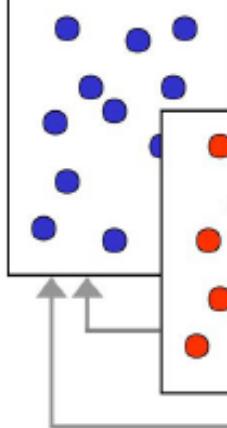
Cy3



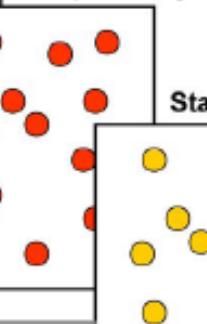
Cy5



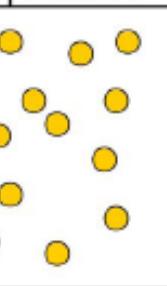
Sample 2 - Cy5



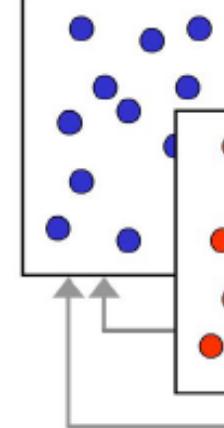
Sample 1 - Cy3



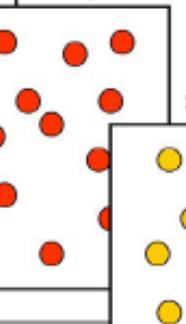
Standard - Cy2



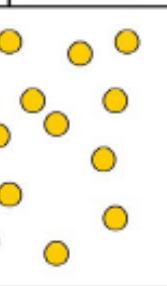
Sample 4 - Cy5



Sample 3 - Cy3



Standard - Cy2

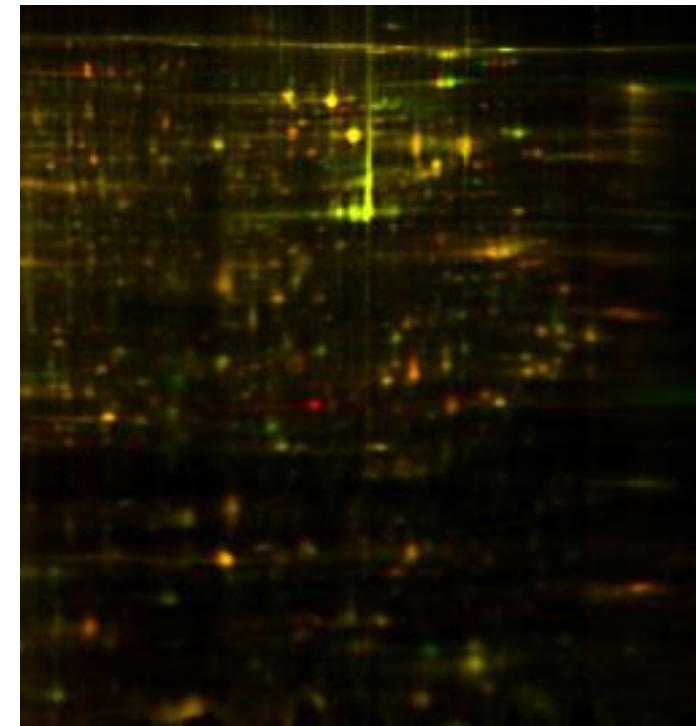


Gel A

MATCHING

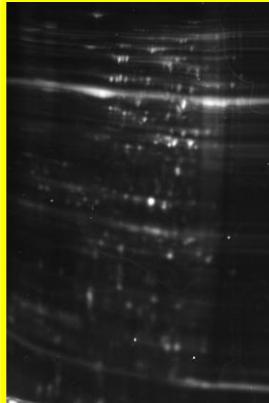
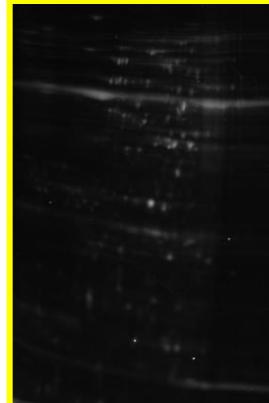
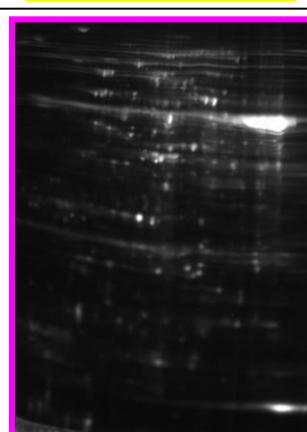
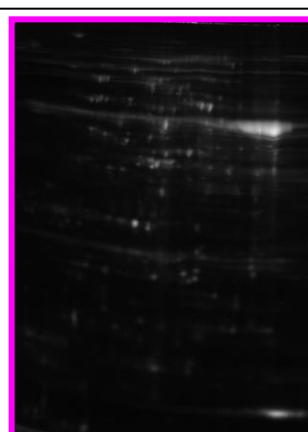
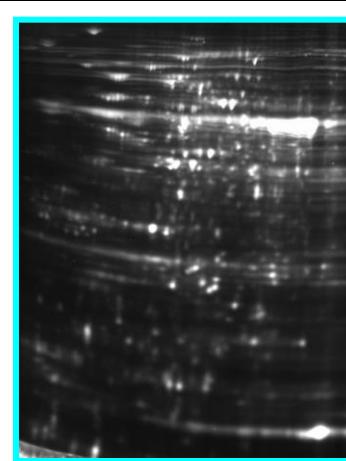
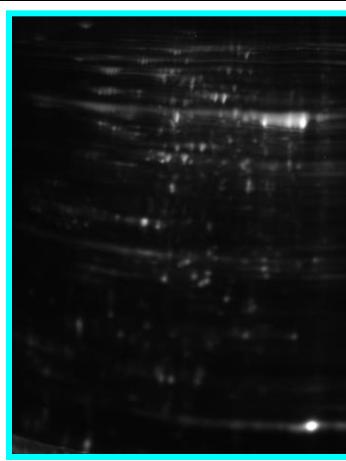
Gel B

Comparaison:  
Cy3/Cy2  $\leftrightarrow$  Cy5/Cy2



## 2-D DIGE

Effet du temps d'exposition

	10 sec	5 sec	3 sec	2 sec
Gel 7 DIGE <b>CyDye2</b>				
Gel 8 DIGE <b>CyDye3</b>				
Résolution 100 µm				

## 2-D DIGE

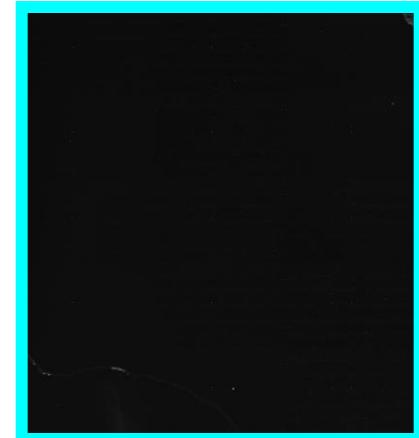
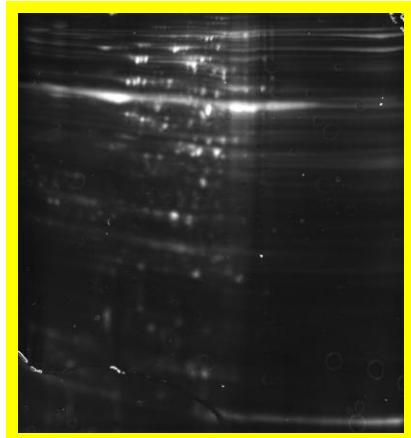
$\lambda$  488/520

Gel 7  
DIGE  
**CyDye2**

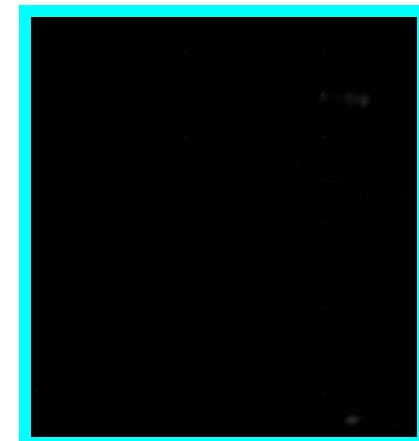
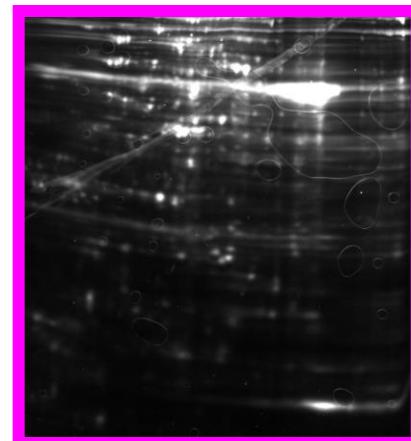
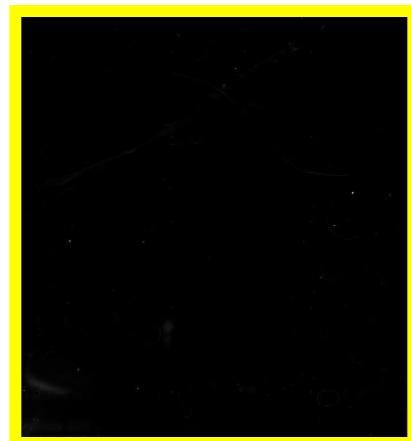
Interférence  
entre  
CyDyes

$\lambda$  532/580

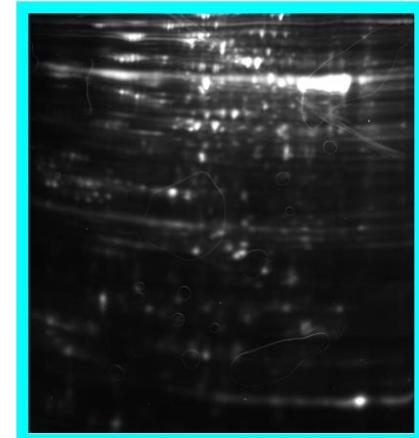
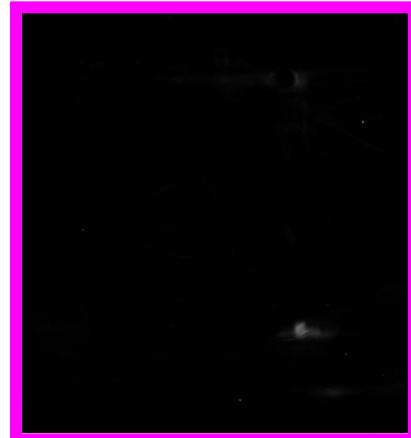
$\lambda$  635/670



Gel 8  
DIGE  
**CyDye3**



Gel 9  
DIGE  
**CyDye5**

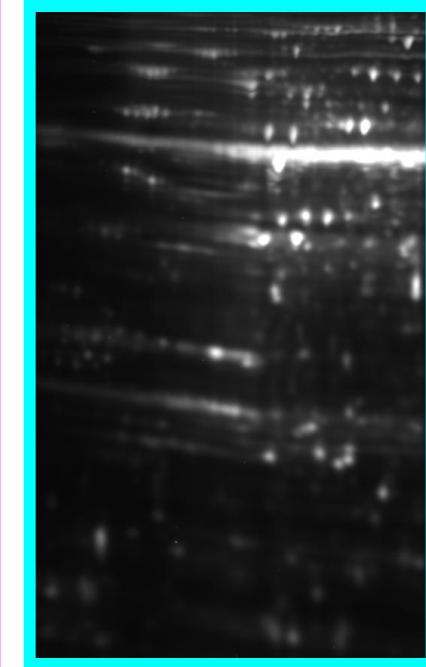
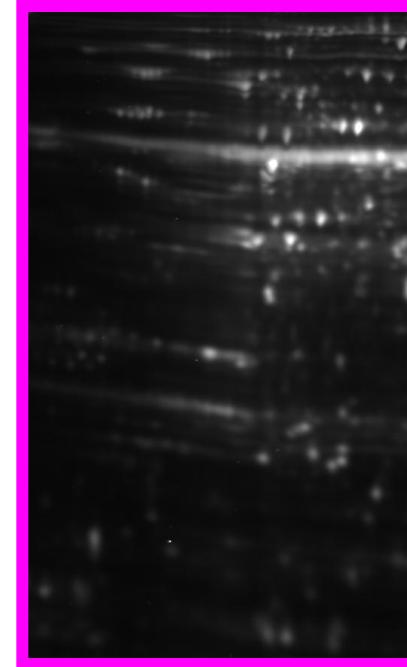
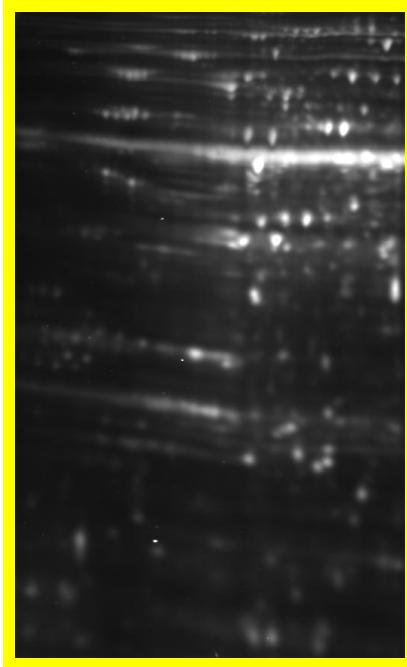


Résolution  
100  $\mu$ m

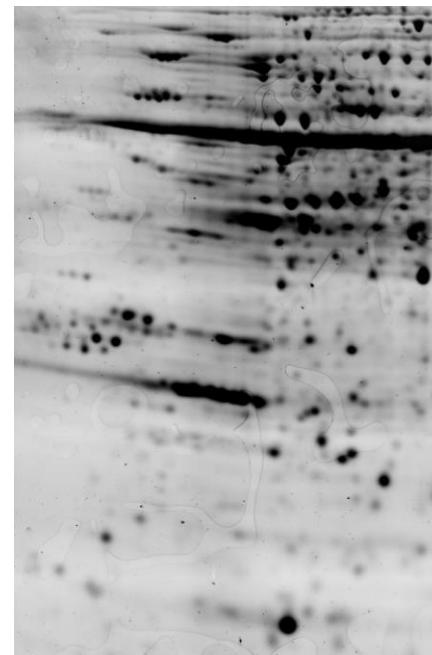
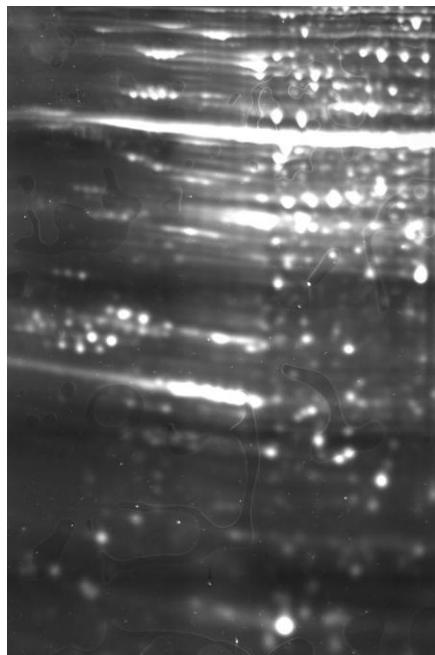
## 2-D DIGE

Sensibilité

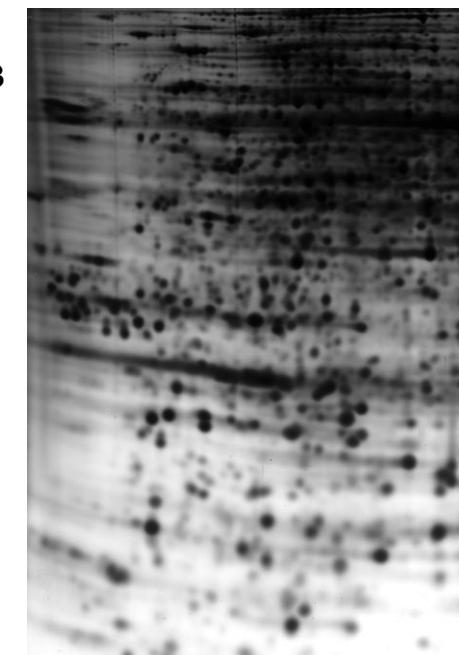
**Gel 2**  
(tech rep gel1)  
**DIGE**  
50 ug/CyDye



**Gel 2**  
**Ruby Sypro**  
150 ug



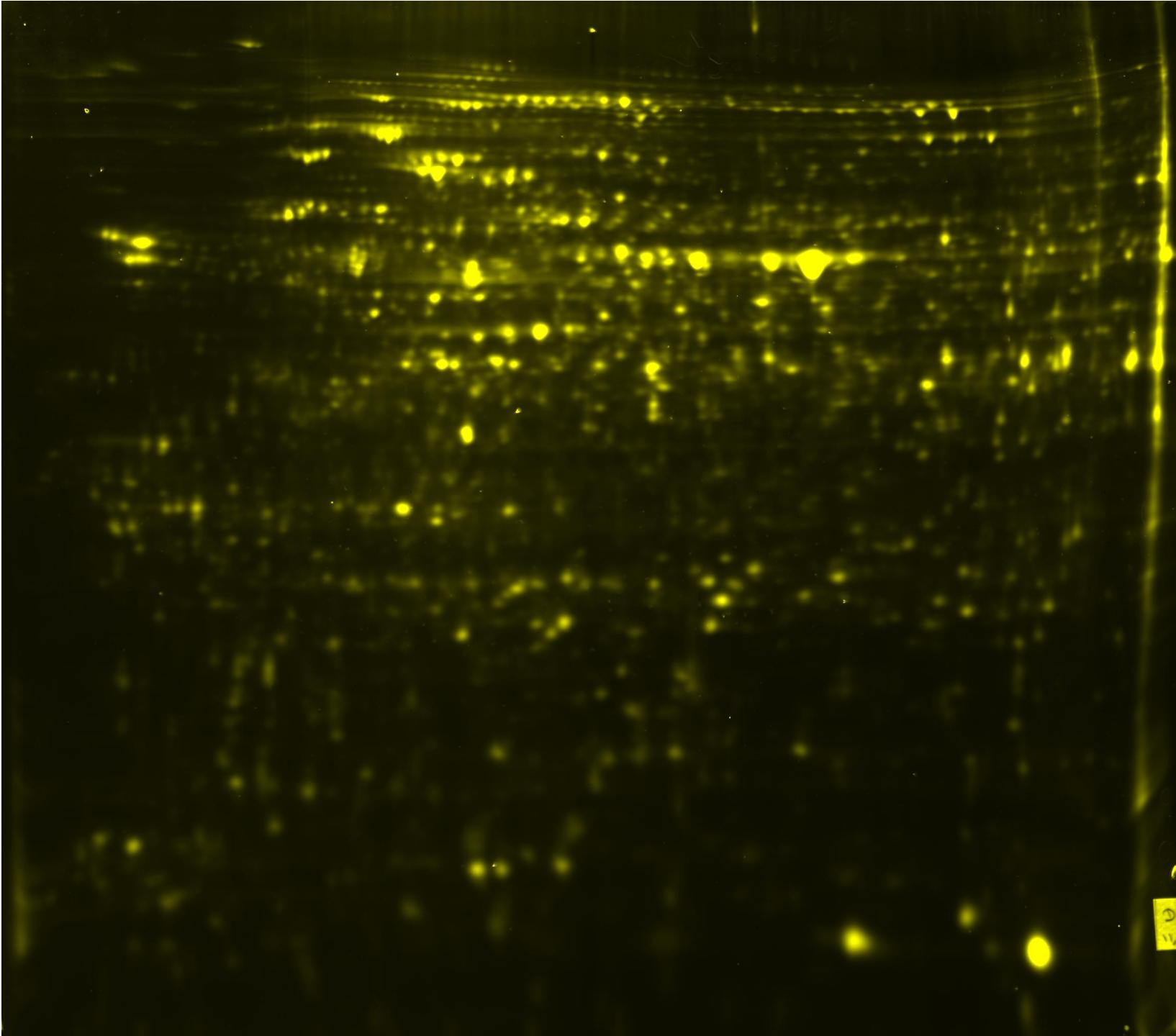
**Gel 12**  
**AgNO<sub>3</sub>**  
50 ug



Résolution  
100 µm

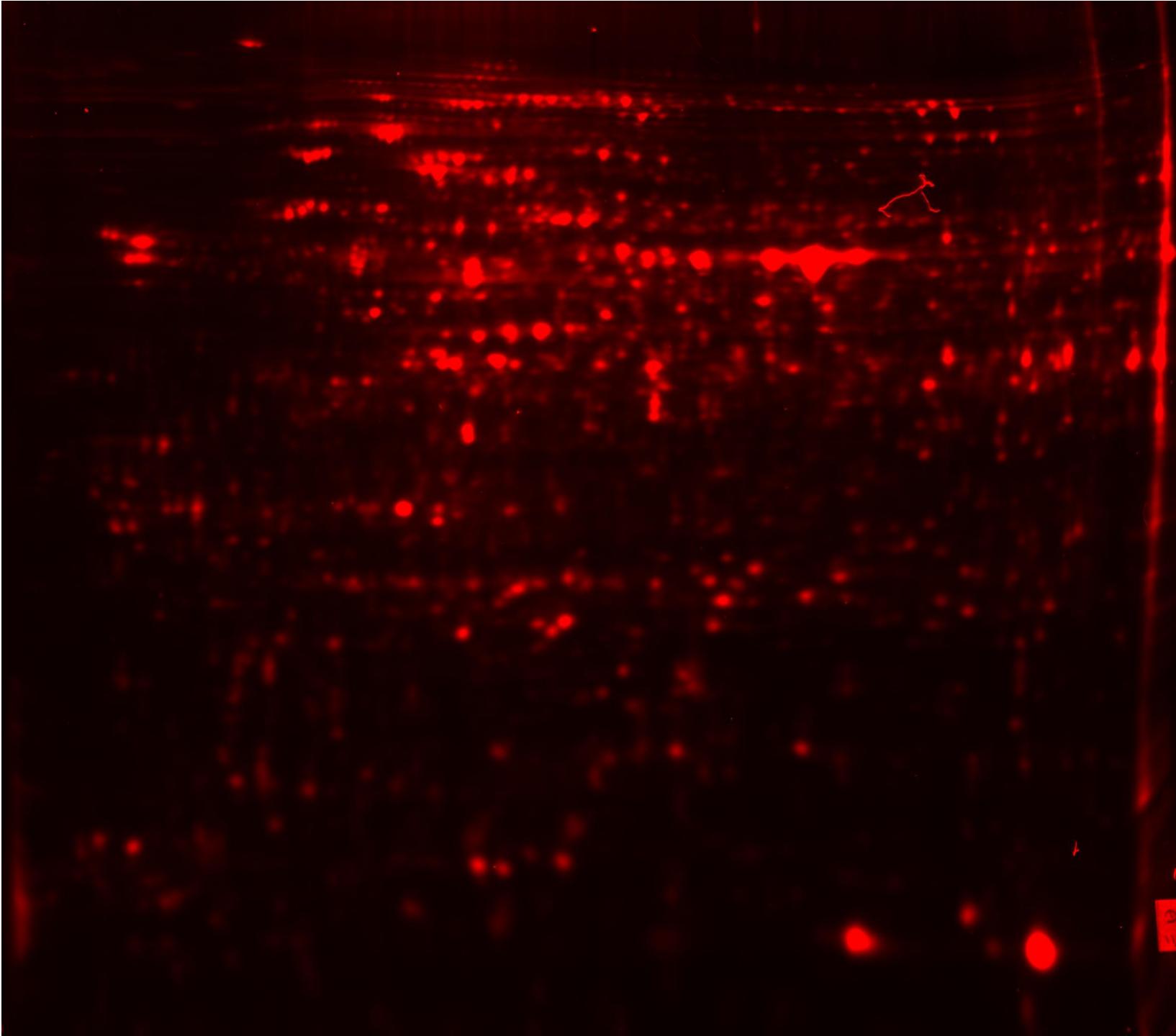
# 2-D DIGE

Cy2



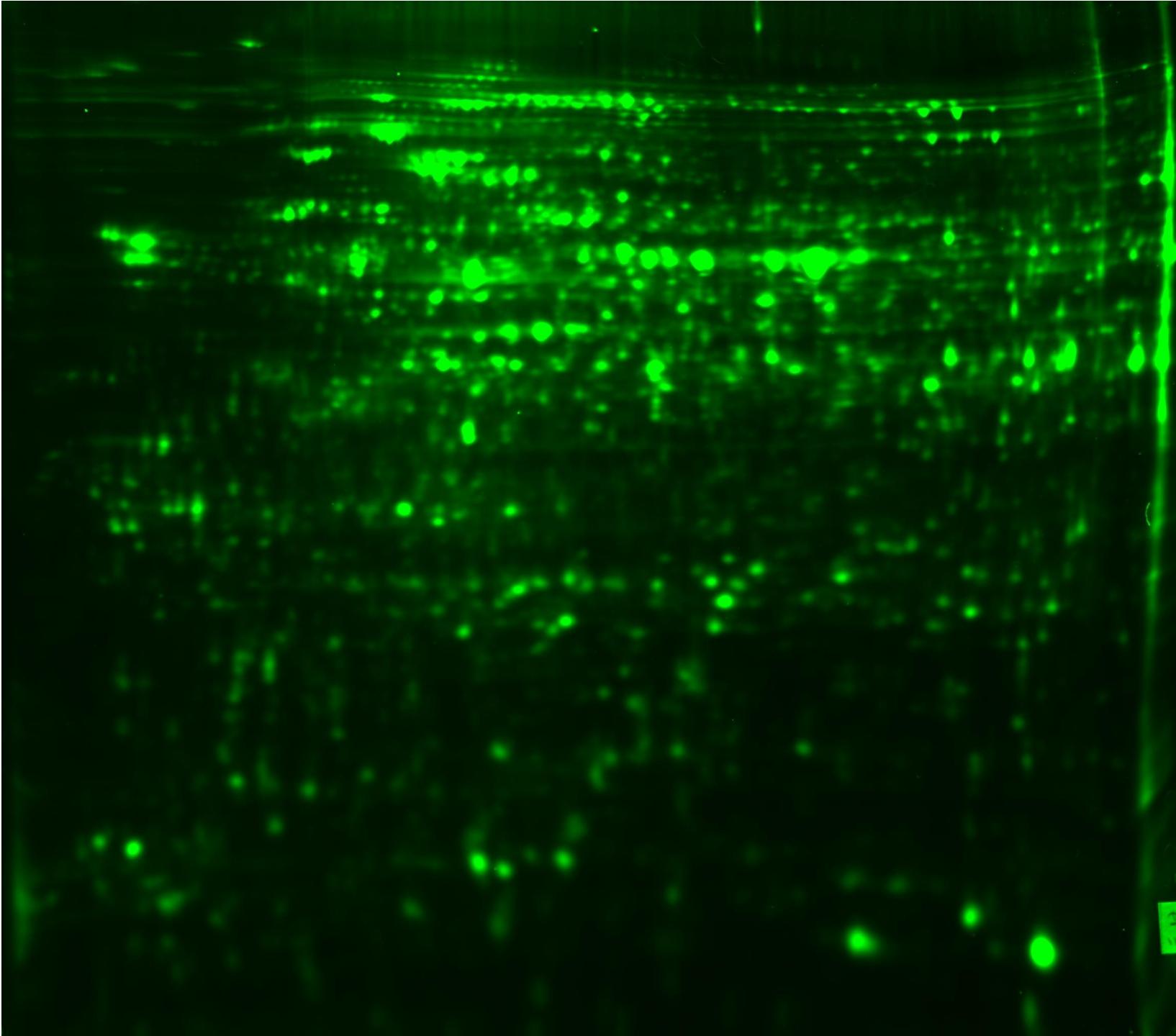
# 2-D DIGE

Cy3



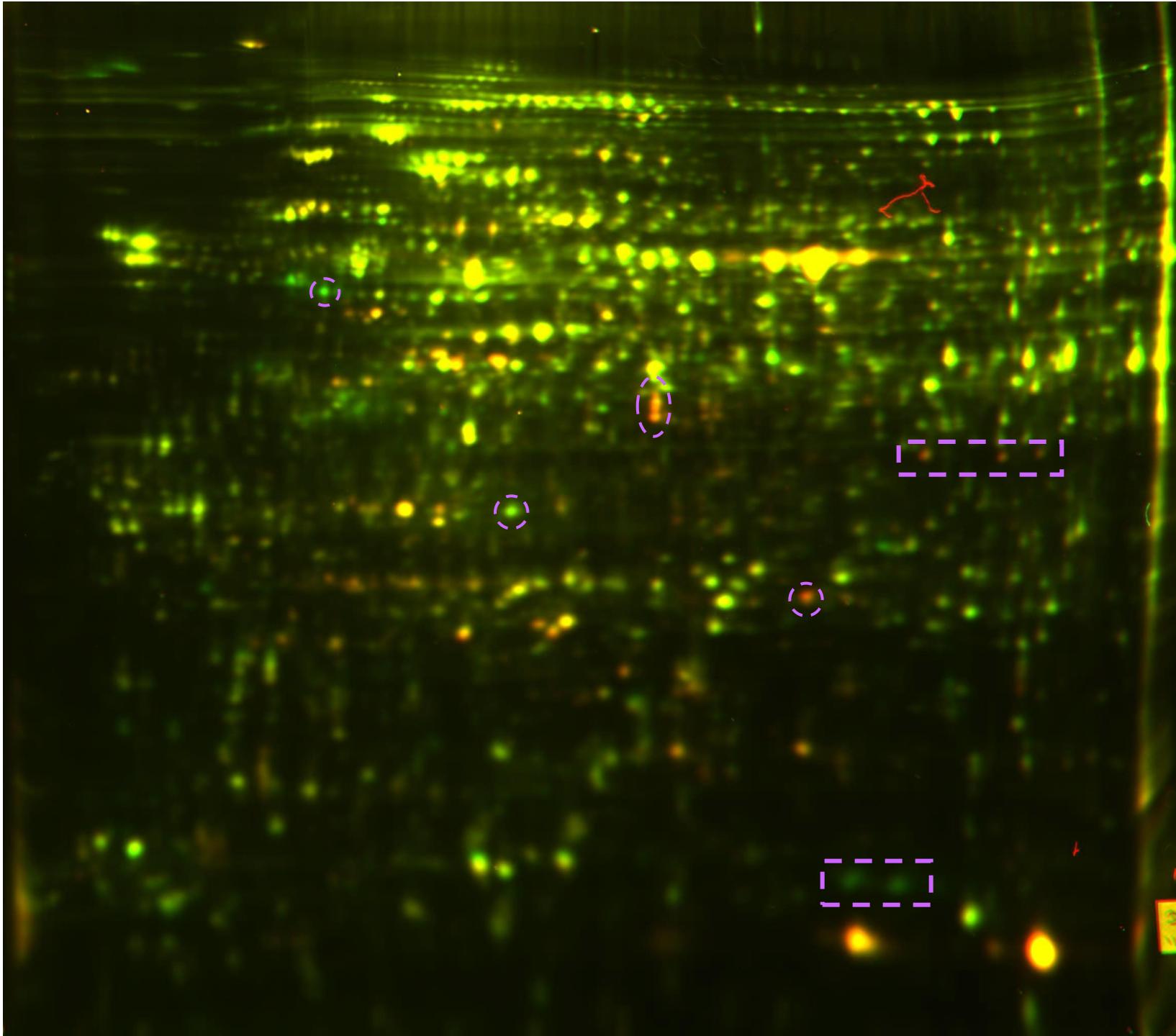
## 2-D DIGE

Cy5

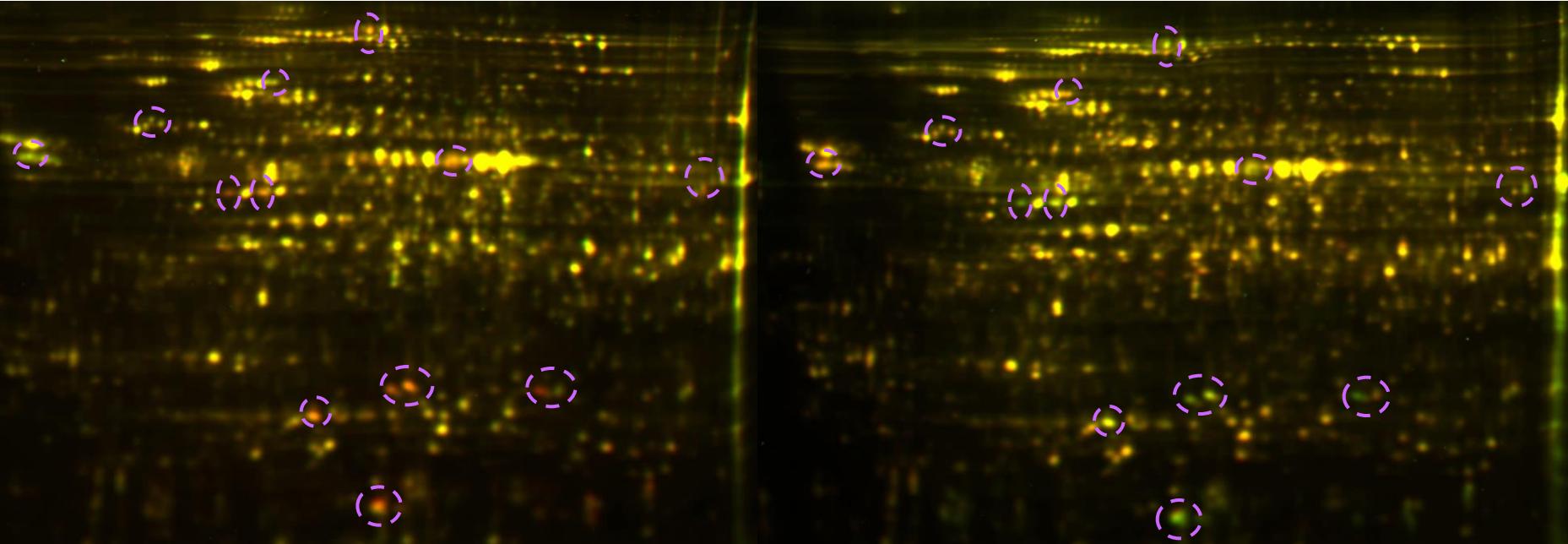
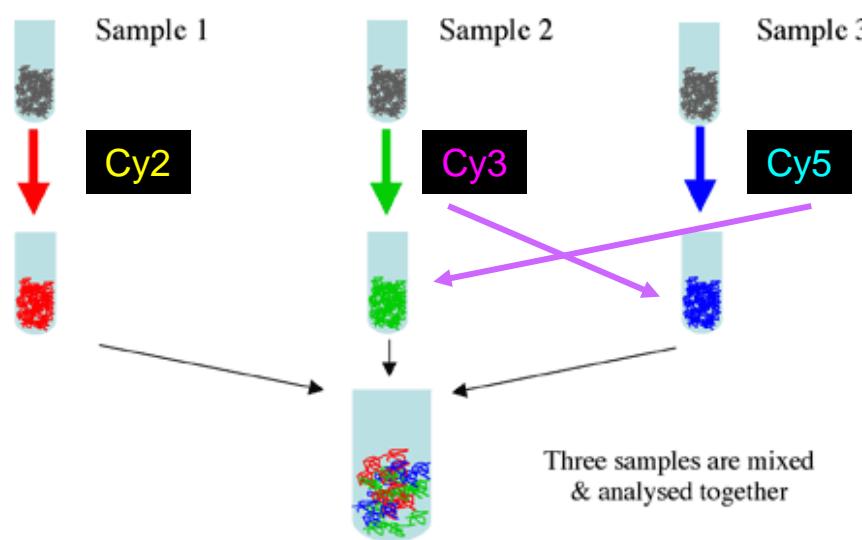


# 2-D DIGE

Cy3/Cy5



## 2-D DIGE: « dye swap »

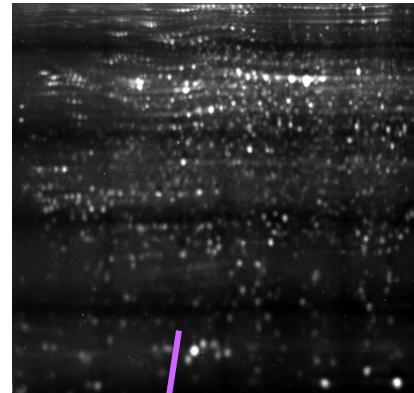
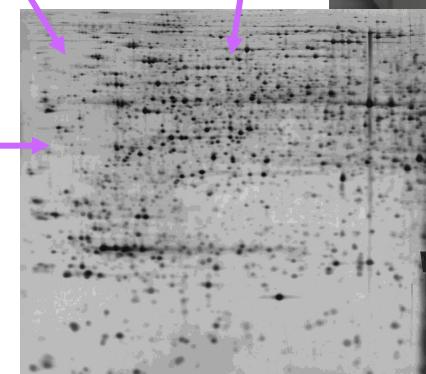
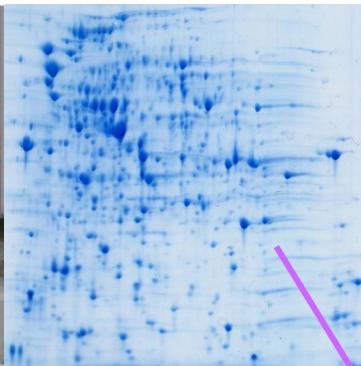


# Acquisition et analyse d'images

# Acquisition de l'image



visible



fluorescent



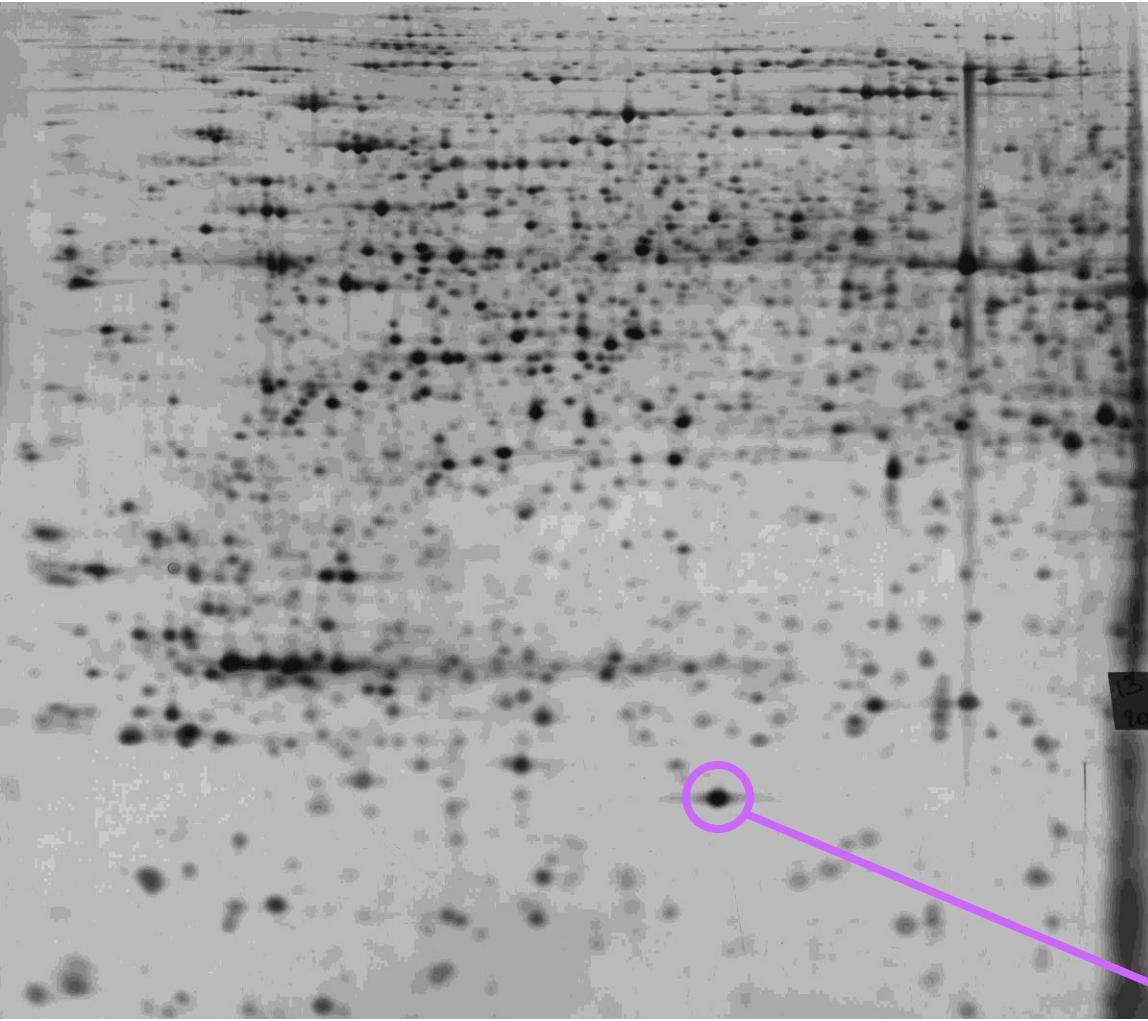
Calibration (visible seulement):  
conversion en DO  
proportionnelle aux quantités

Résolution de l'image (24x20cm):  
visible, 300 dpi  
fluo, 100 µm

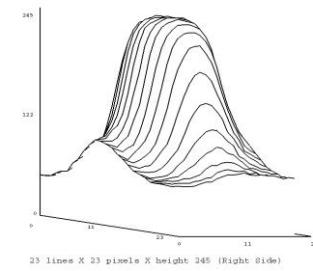
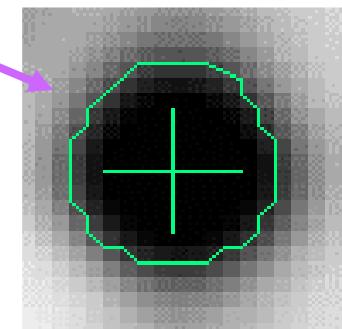
Nombre de niveaux de gris de l'image:  
image 8-bit, 256 niveaux  
image 16-bit, 65536 niveaux

Resolution (dpi)	Resolution (micron)	Image Dimensions (pixels)	Total no. of pixels per image	Bit Depth	File Size (Mb)
100	254	776 x 775	$5.97 \times 10^4$	8	0.6
200	127	1553 x 1550	$2.41 \times 10^6$	8	2.7
300	84.67	2329 x 2325	$5.41 \times 10^6$	8	5.3
100	254	776 x 775	$5.97 \times 10^4$	16	1.2
200	127	1553 x 1550	$2.41 \times 10^6$	16	4.8
300	84.67	2329 x 2325	$5.41 \times 10^6$	16	10.6

# Traitement d'image

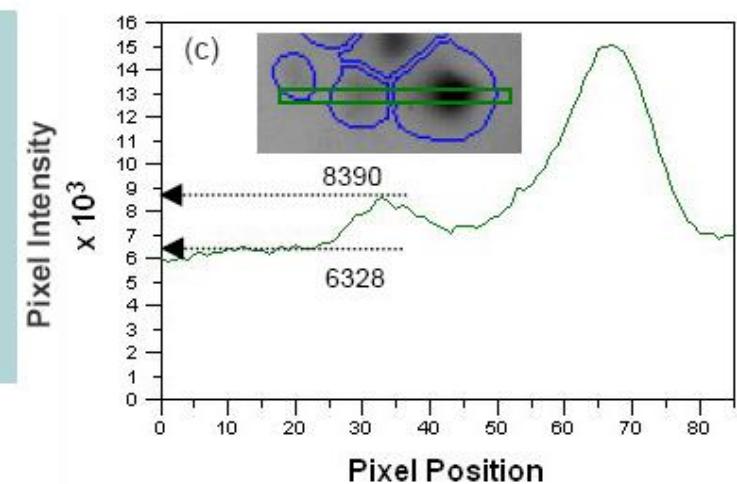
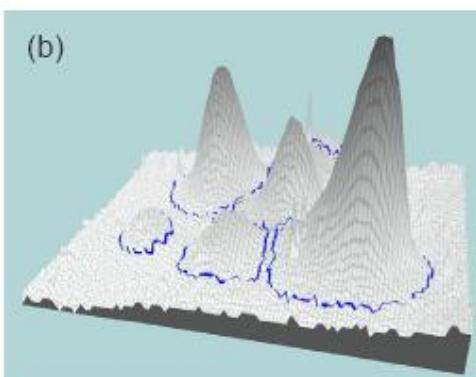
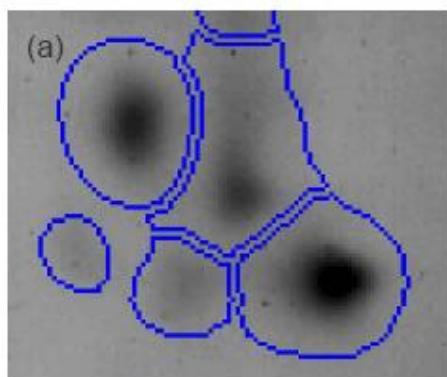
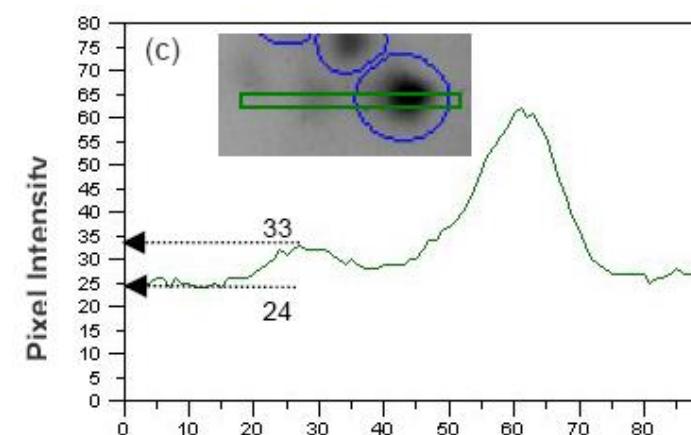
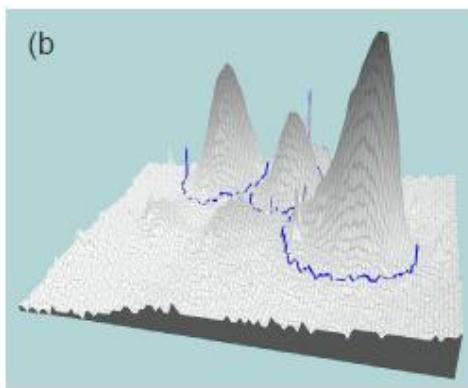
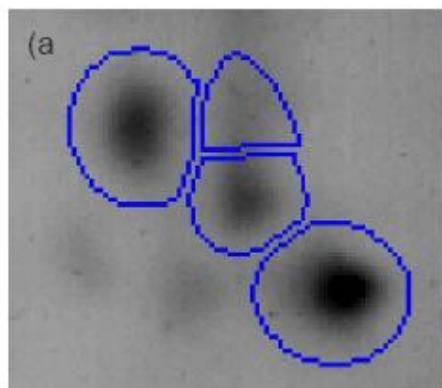


Mesurer les quantités de protéines.



# Traitement d'image: effet de la profondeur

8 Bit (256 niveaux) vs. 16 Bit (65536 niveaux)

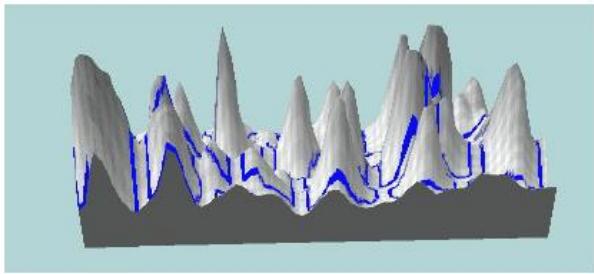
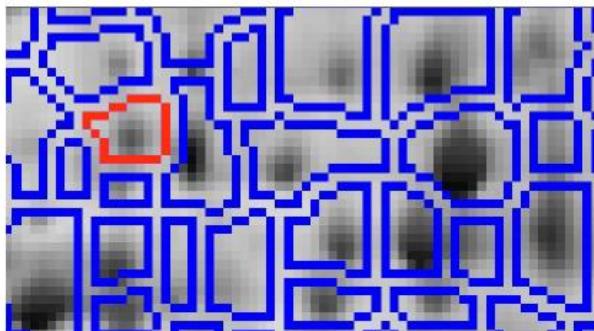
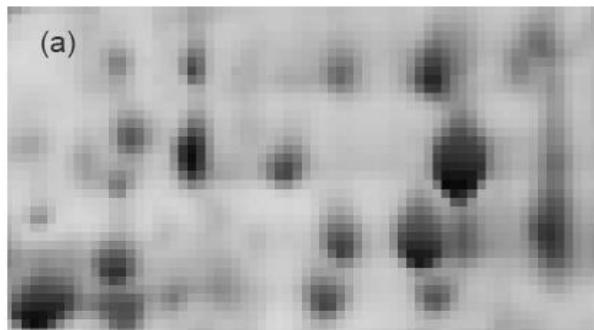


Une profondeur plus importante permet à un pixel d'être représenté par un plus grand nombre de couleurs ou de niveaux de gris.

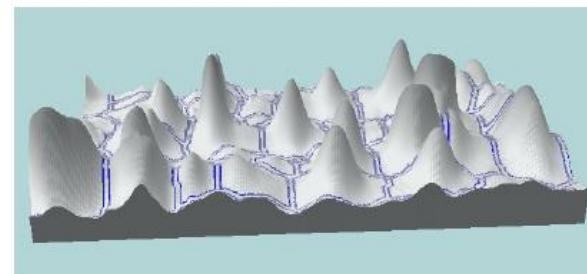
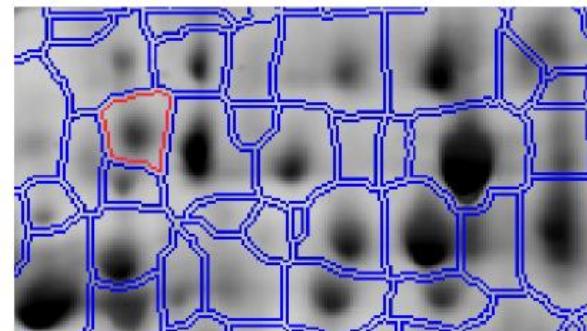
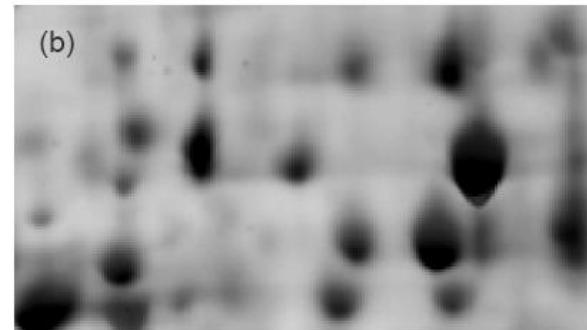
Plus les niveaux de gris seront élevés, meilleure sera la séparation des spots du bruit de fond, et meilleure sera la détection/quantification.

## Traitement d'image: effet de la résolution

**100dpi**



**300dpi**

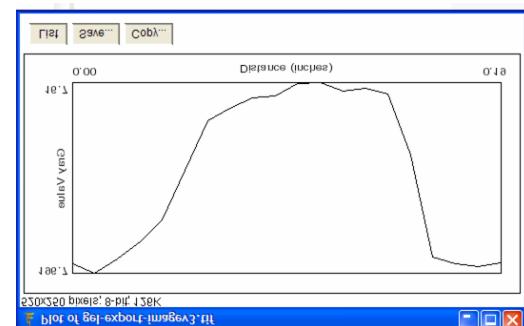
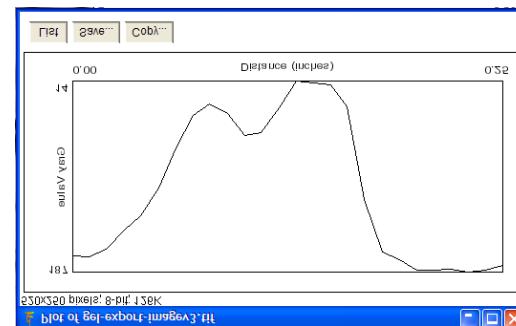
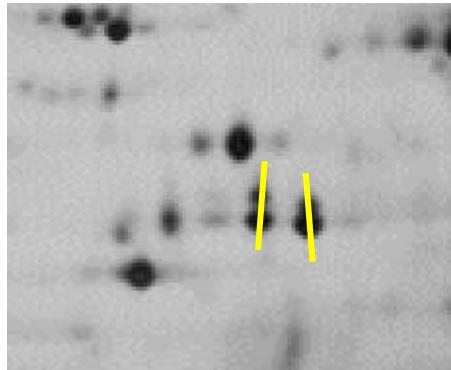


Une résolution plus élevée signifie que plus de pixels et donc plus de données sont disponibles pour l'analyse .

# Traitement d'image: effet de la résolution

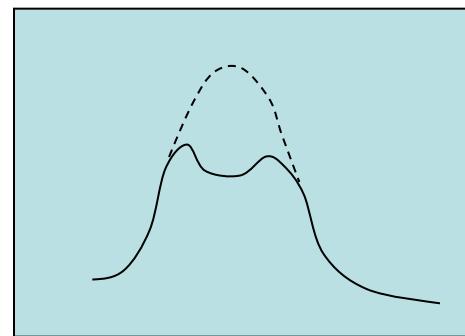
## Chevauchement:

- Mauvaise reproductibilité de détection
- Mauvaise quantification
- Multiples protéines dans le même spot (« cross-contamination », identification MS)



Favoriser:

- grands gels
- gamme de pH étroite

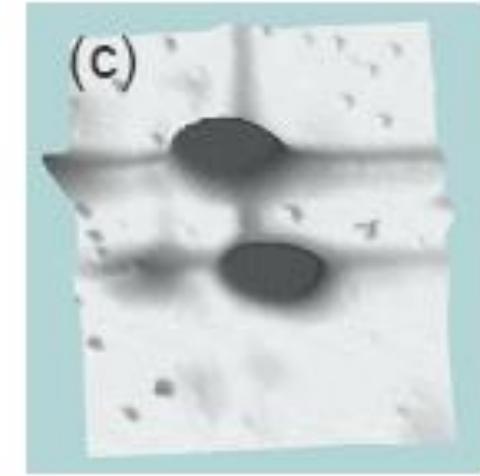
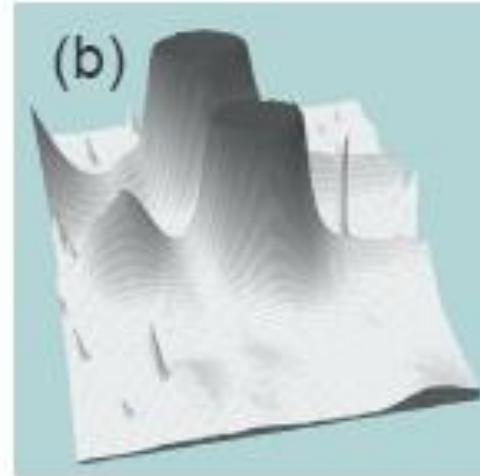
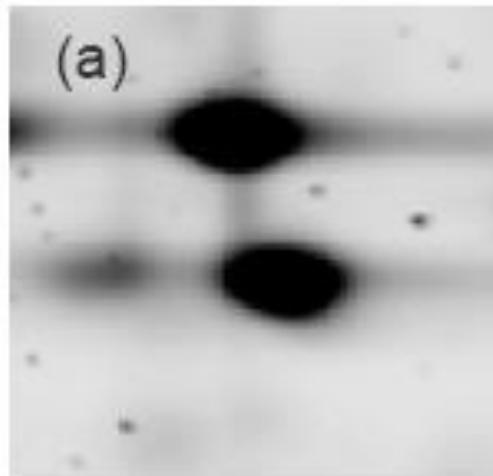


Révélation: pas de saturation (AgNO<sub>3</sub>)

## Traitement d'image: effet de la saturation, optimiser la dynamique

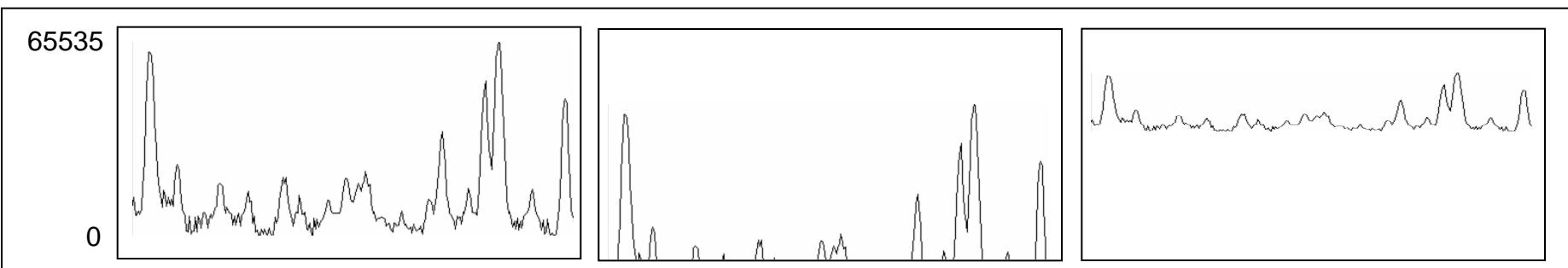
En optimisant la dynamique, il est important d'éviter les effets de saturation

- la saturation apparaît lorsque les niveaux de gris atteignent le maximum
- aucune donnée quantitative valable ne peut être générée à partir d'un spot saturé
- il y a un effet sur la normalisation.

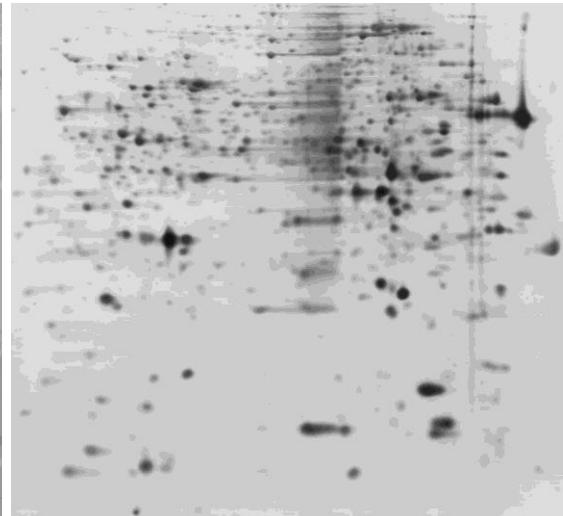
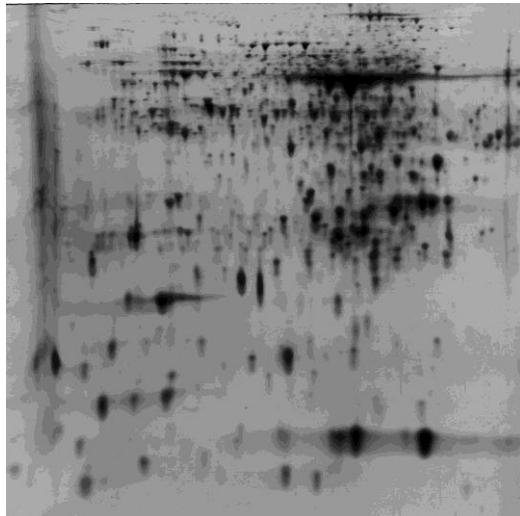
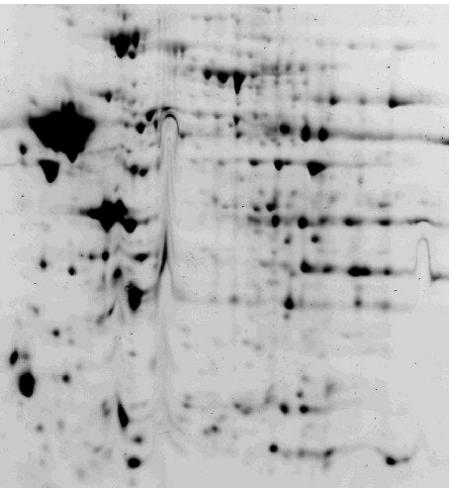
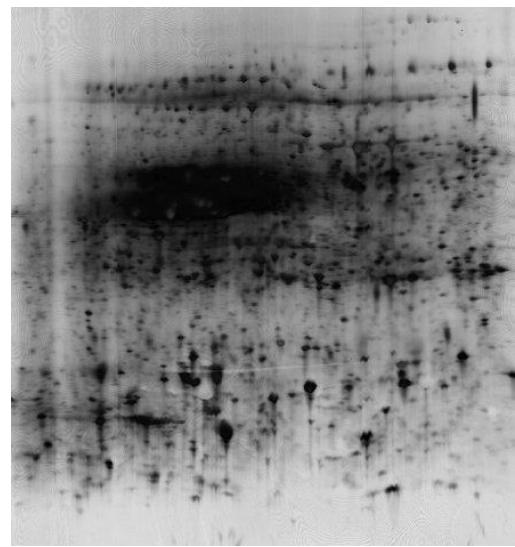
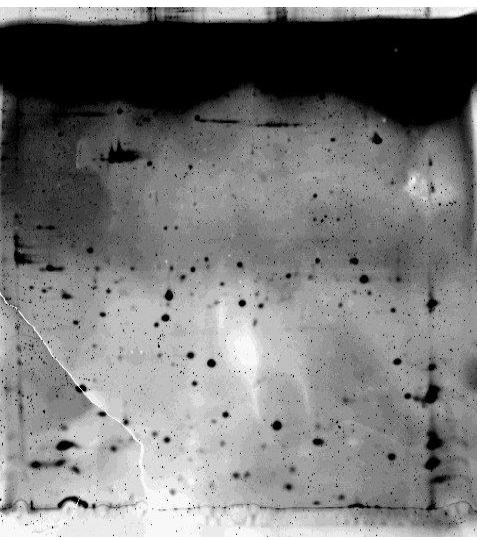


### Dynamique de l'image

Utiliser au maximum la dynamique de l'image sans aller jusqu'à 0, ni 100%



## Traitement d'images: exemples de gels inexploitables

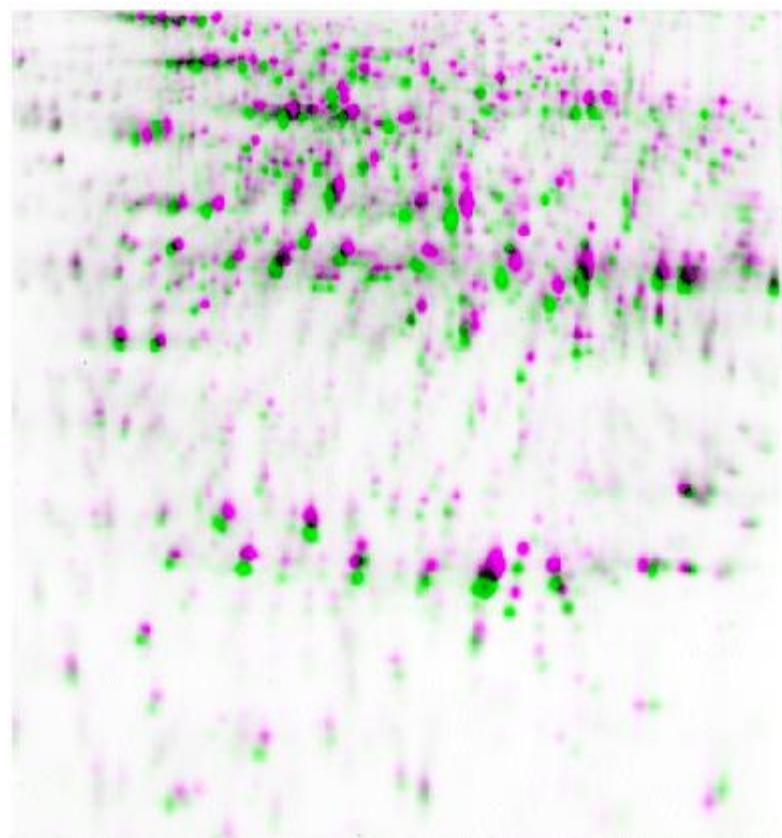


# Traitement d'images: les différents logiciels

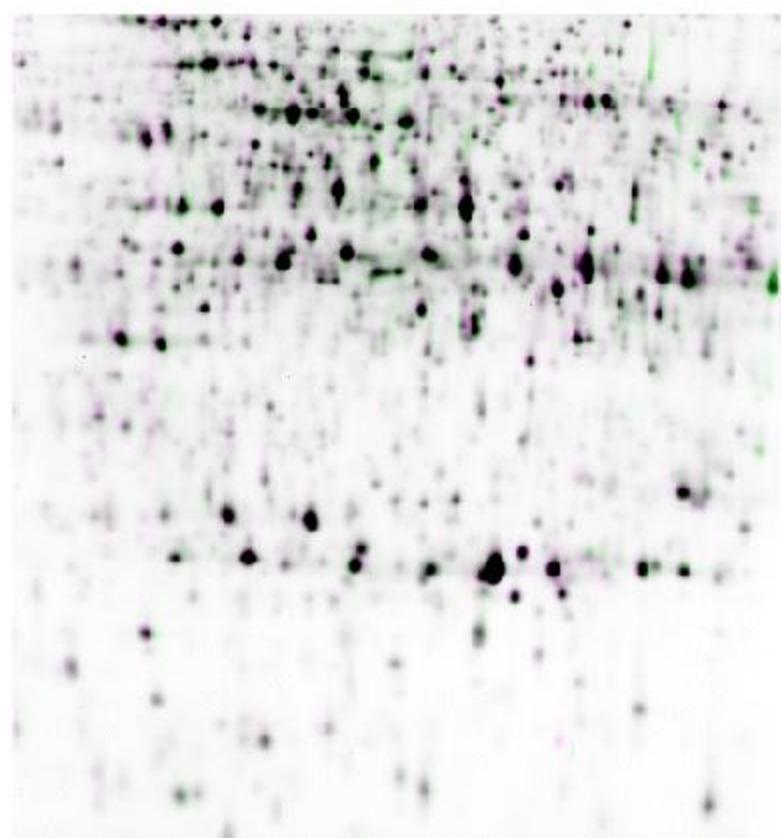
#	Software	Company <a href="http://www.decodon.com">http://www.decodon.com</a>	Year of arrival	Comments	Platforms	Images supported
1	Delta 2-D	DECODON <a href="http://www.decodon.com">http://www.decodon.com</a>	2000	Save-disabled evaluation version available	PC (Windows 98, ME, 2000, NT), Linux, Sun Solaris, Mac OS X	TIFF (8, 12 and 16 bit), JPEG, BMP, GIF, PNG.
2	GELLAB II+	Scanalytics <a href="http://www.scanalytics.com/">http://www.scanalytics.com/</a>	1999	Trial version available	PC (Windows 95, NT)	TIFF (8 bit)
3	Melanie	Geneva Bioinformatics <a href="http://www.genebio.com">http://www.genebio.com</a>	N/A <sup>d</sup>	30 day fully functional trial version available	PC (Windows 95, 98, 2000, NT)	TIFF (8, 16 bit), GIF, Biorad Scan
4	PD Quest	Bio-Rad Labo <a href="http://www">http://www</a>		<u>Procédure</u> Alignment des images (Conversion des pixels en DO) Elimination du bruit de fond Détection des spots Quantification des spots Appariement (« matching ») des spots Normalisation Exportation des quantités pour analyses statistiques	NT), or PC	TIFF (8, 16 bit)
5 <sup>e</sup>	Phoretix 2-D Advanced	Nonlinear Dy <a href="http://www">http://www</a> <a href="http://www">http://www</a>				TIFF (8, 12 and 16 bit)
5.1	AlphaMatch 2-D	Alpha Innotech <a href="http://alph">http://alph</a>				TIFF (8,12 and 16 bit)
5.2	Image Master 2-D Elite	Amersham Bi <a href="http://www">http://www</a>				TIFF (8,12 and 16 bit)
5.3	Investigator HT Analyzer	Genomic Sol <a href="http://www">http://www</a>				TIFF (8, 12 and 16 bit)
6	Progenesis	Nonlinear Dynamics <a href="http://www.nonlinear.com">http://www.nonlinear.com</a> <a href="http://www.phoretix.com">http://www.phoretix.com</a>	2001	Special hardware and software requirements	PC (Windows 2000)	TIFF (8, 12 and 16 bit)
7	Z3	Compugen <a href="http://www.2dgels.com">http://www.2dgels.com</a>	2000	21 day fully functional trial version available	PC (Windows 98, 2000, NT)	TIFF (8, 12 and 16 bit), JPEG, BMP, GIF, PNG, GEL, FLT
8	Proteome-Weaver	Definiens (Munich, Germany)	2002	21 day fully functional trial version available	PC (Windows 2000, XP)	TIFF (8, 12 and 16 bit), JPEG, BMP, GIF, PNG, GEL, FLT

# Traitement d'images: alignement des gels

Etape semi automatisée.



avant



Après

# Traitement d'images: alignement des gels

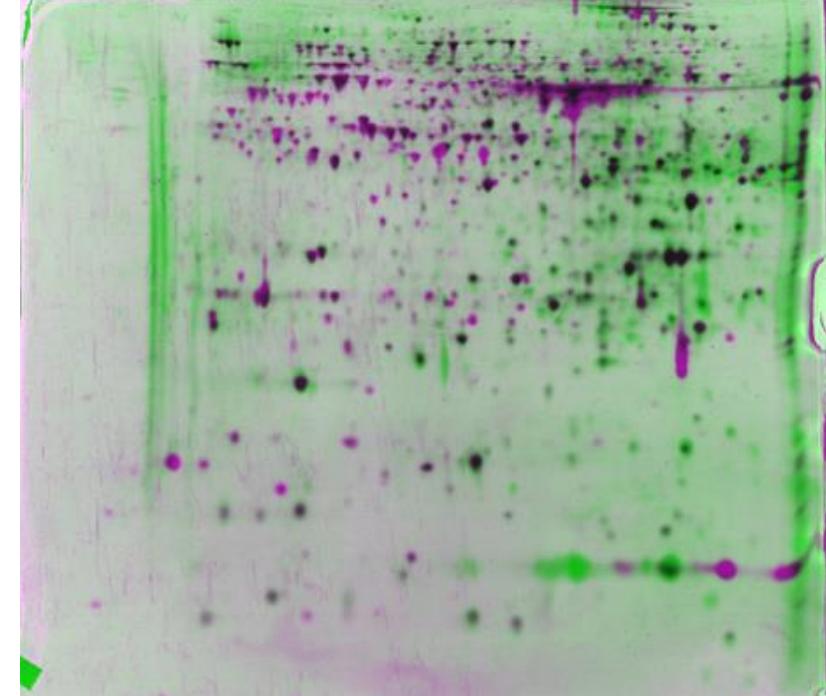
Etape semi automatisée.



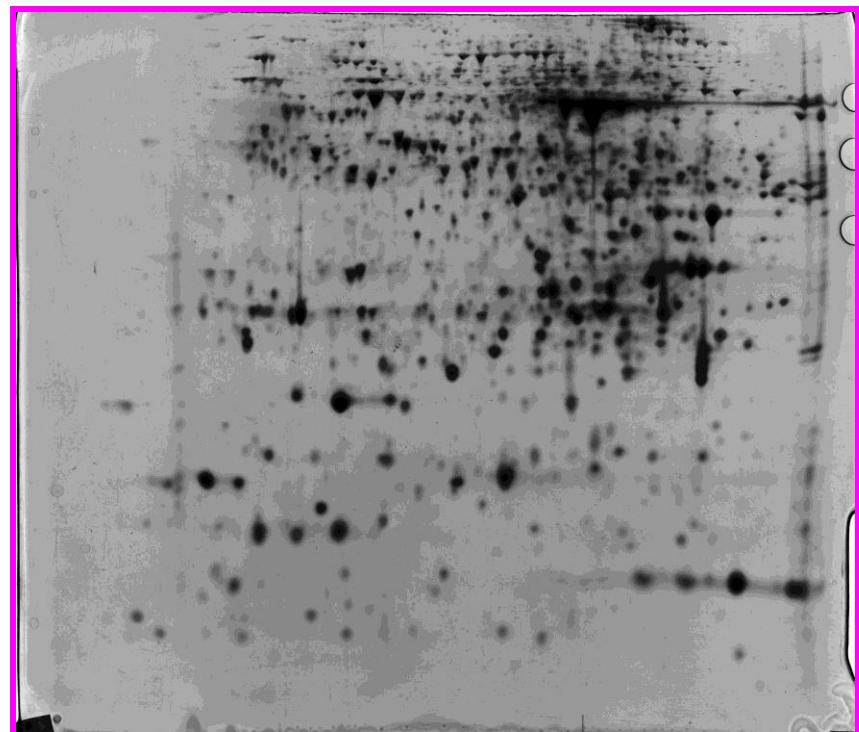
Progenesis  
SameSpot

Variations qualitatives (absence/présence)  
et quantitatives (augmentation/diminution).

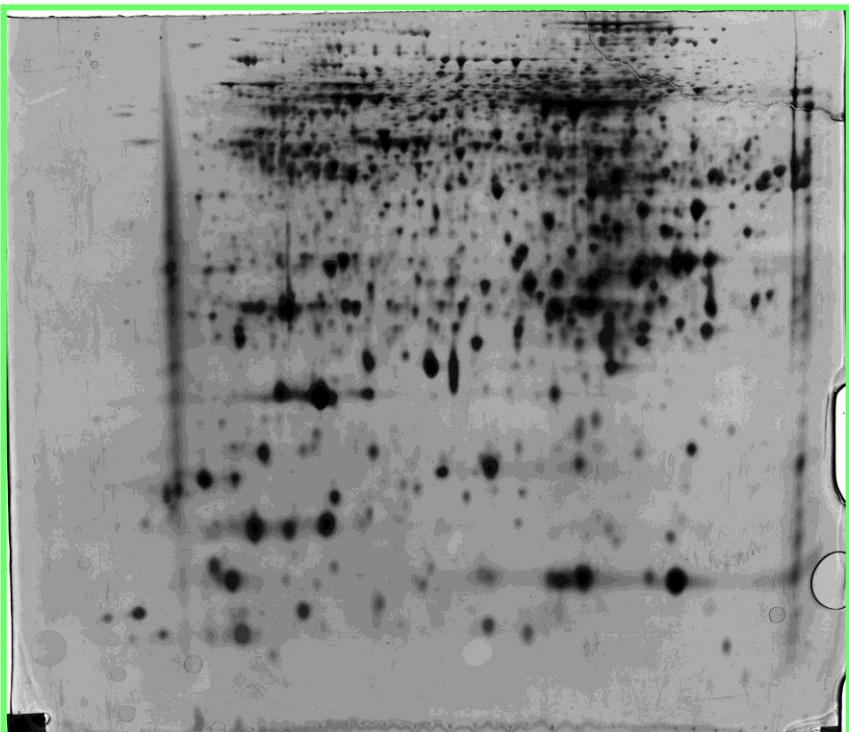
Ex.: variations entre espèces



*Populus deltoides*



*Populus trichocarpa*



# Traitement d'images: alignement des gels

Etape semi automatisée.



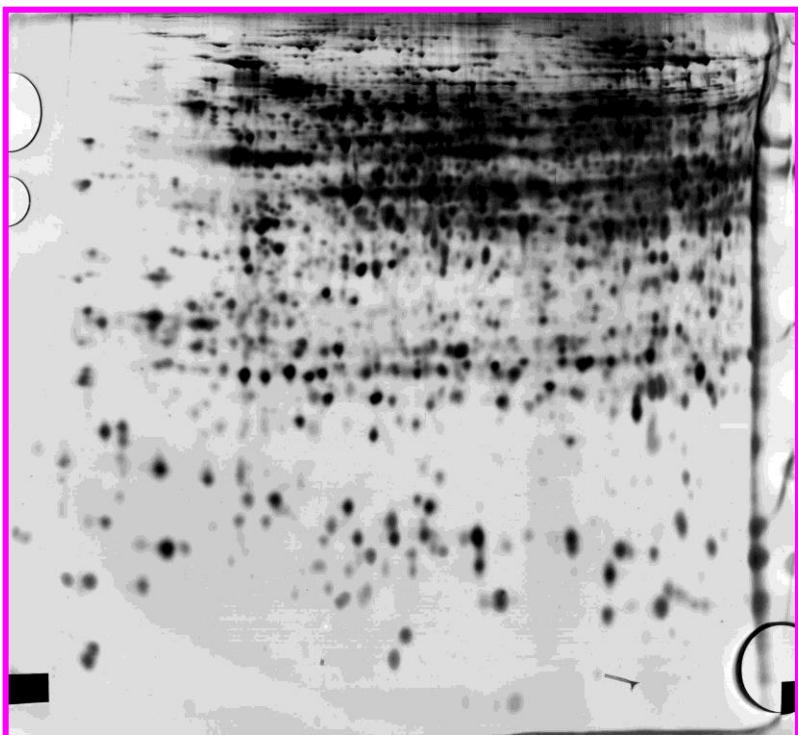
Progenesis  
SameSpot

Variations qualitatives (absence/présence)  
et quantitatives (augmentation/diminution).

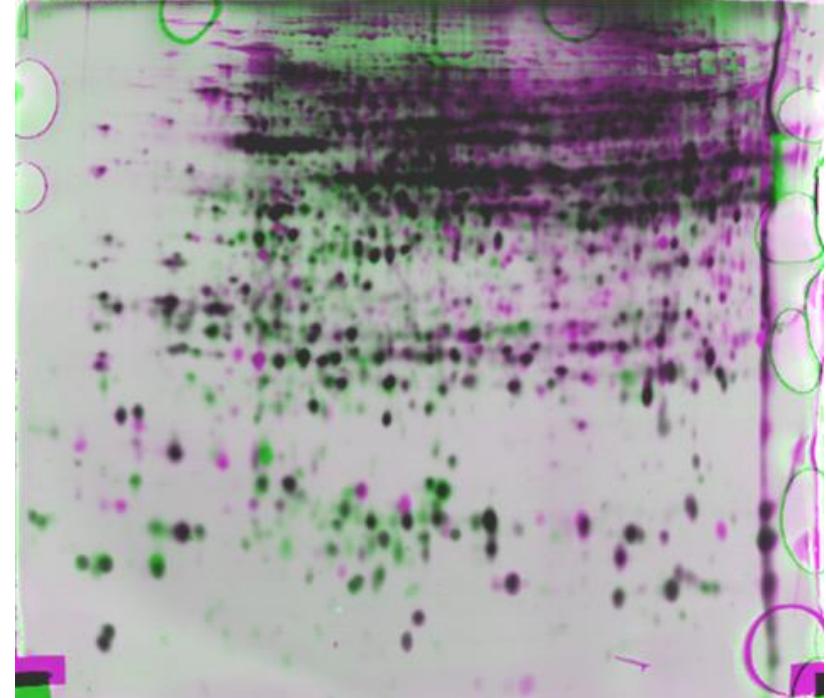
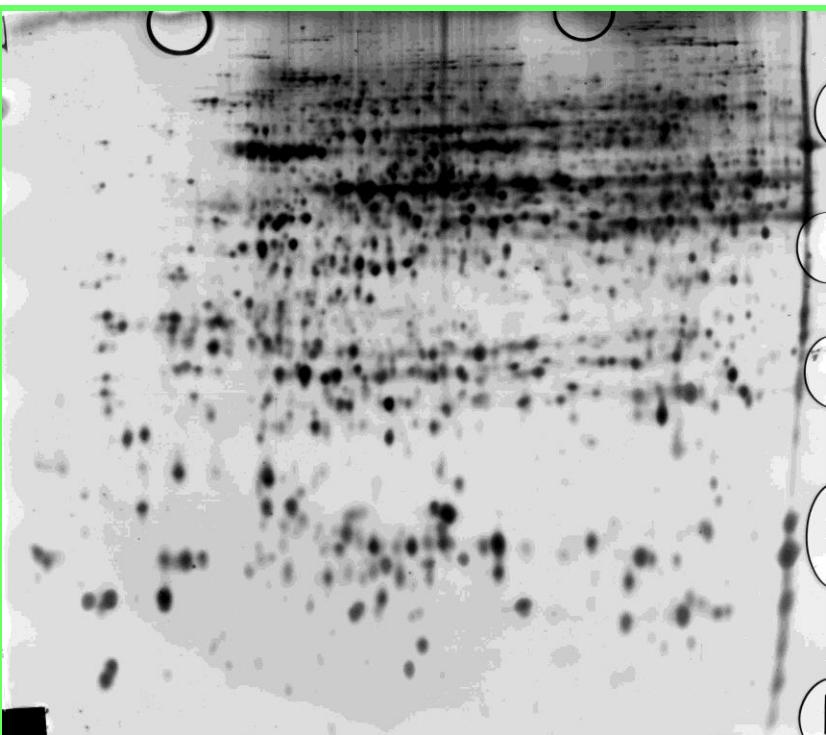
Ex.: variations entre traitements



Témoin



Stressé

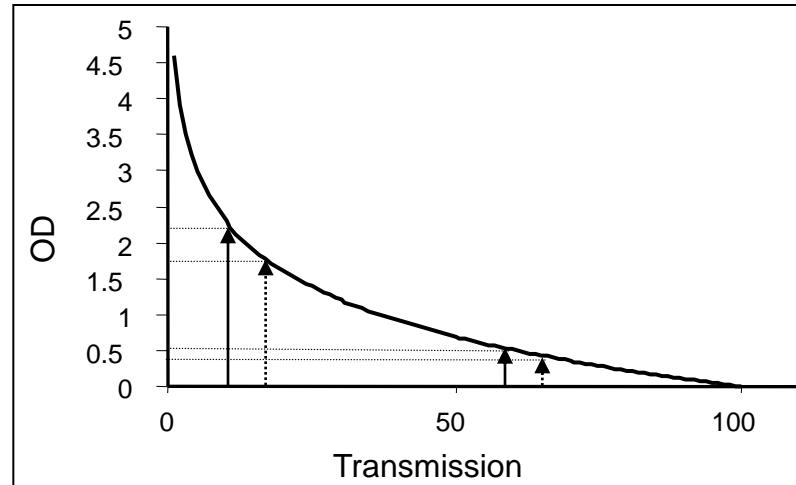


# Traitement d'images: conversion en densité optique (DO)

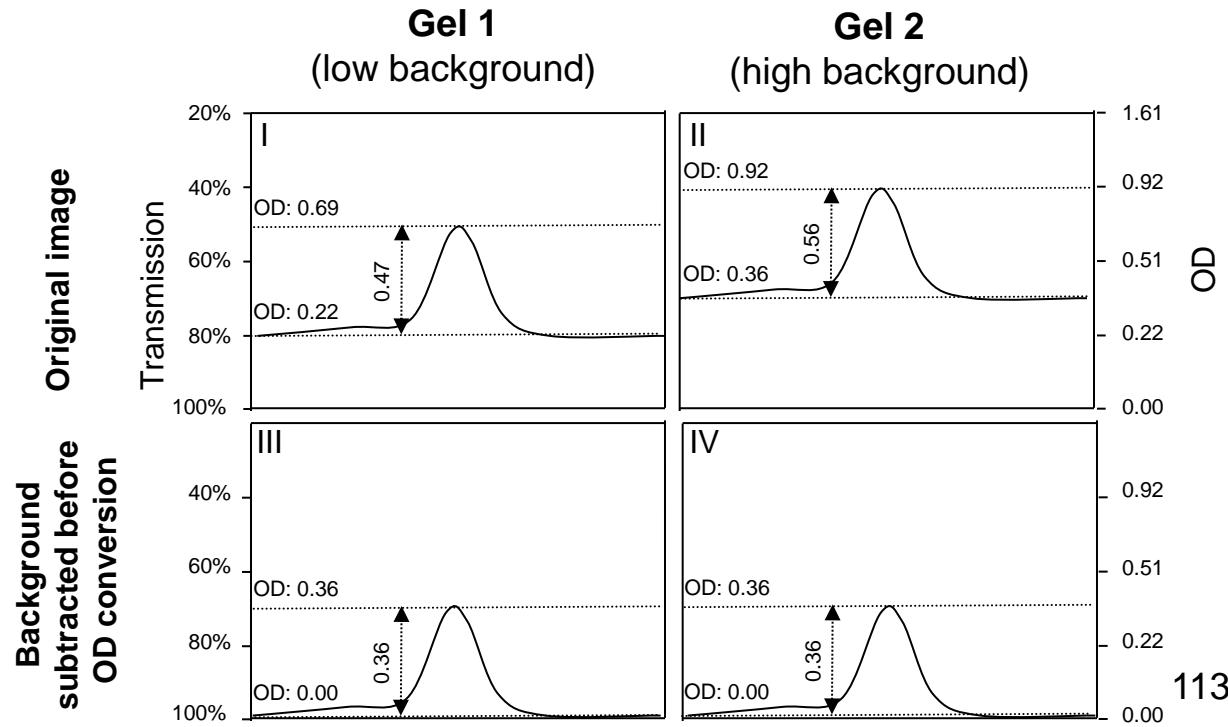
Contrairement à la transmission, la densité optique (DO) est proportionnelle à la concentration en protéine.

$$DO = -\log(I/I_0)$$

$$DO(A+B) = DO(A) + DO(B)$$



La conversion en DO doit être faite avant la soustraction du bruit de fond.



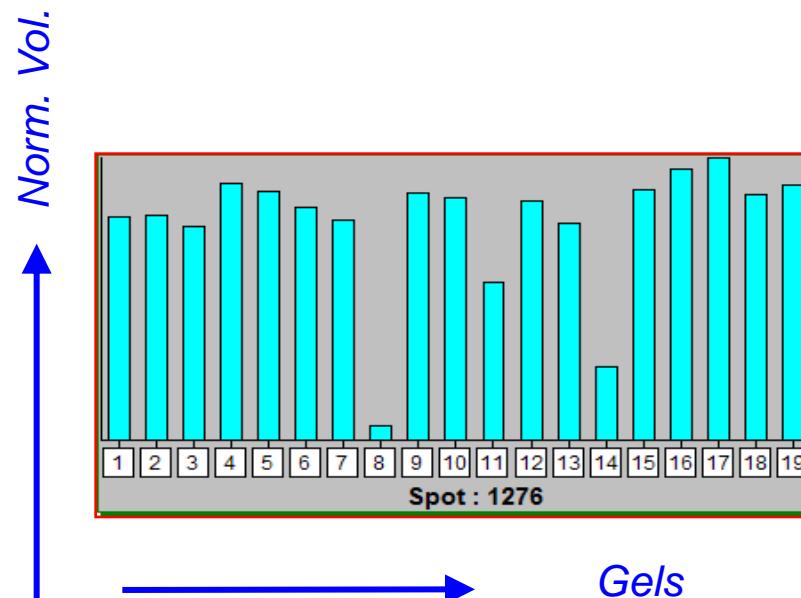
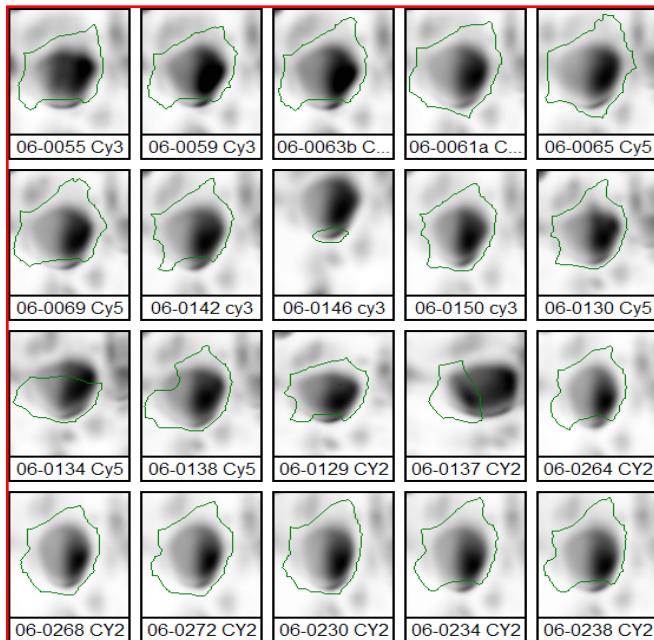
# Traitement d'images: détection/quantification

Etape complètement automatisée cependant l'utilisateur peut éditer (ajout/suppression, modification des contours) les spots.

Comporte une étape d'élimination du bruit de fond.

## Problème 1: Quantification incorrecte des spots

### Inconsistance de la détection

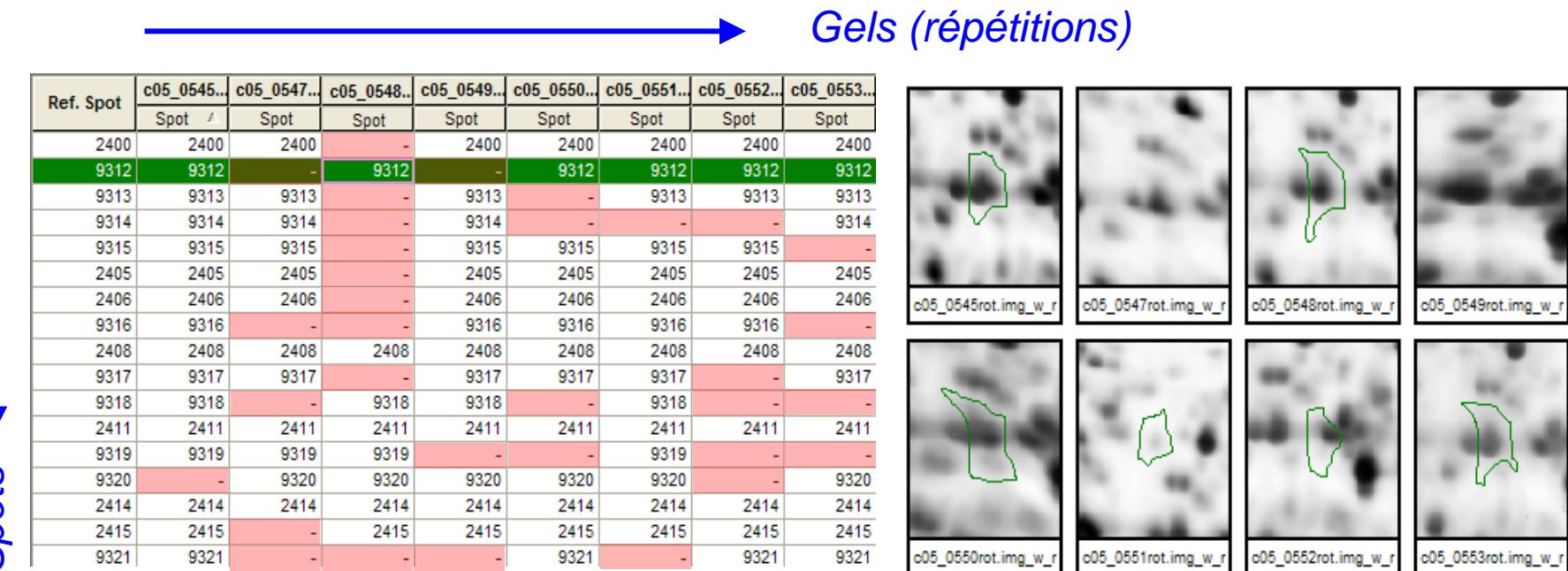


Cela tend à augmenter la variance du volume du spot, avec pour conséquence:

- Réduction de la puissance statistique
  - Accroissement du nombre de faux négatifs

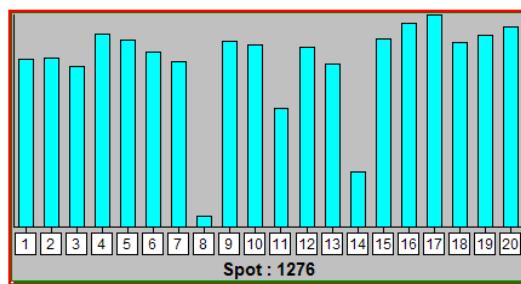
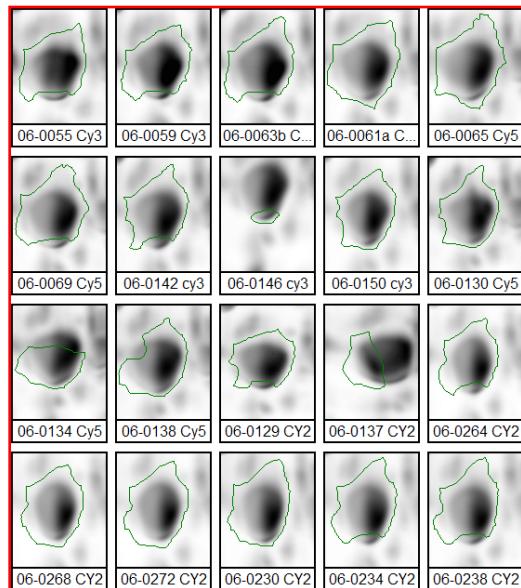
## Problème 2: Faible nombre de répétitions

Les répétitions sont nécessaires pour distinguer les vraies différences d'expression des fluctuations dues au hasard mais plus de répétitions signifie plus de données manquantes à cause des erreurs de détection et d'appariement.

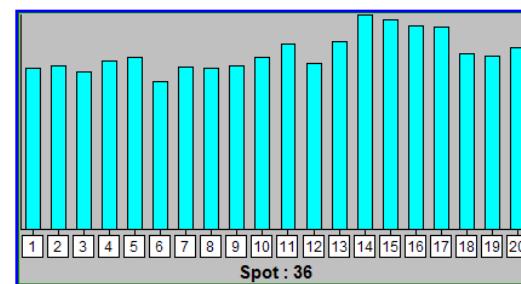
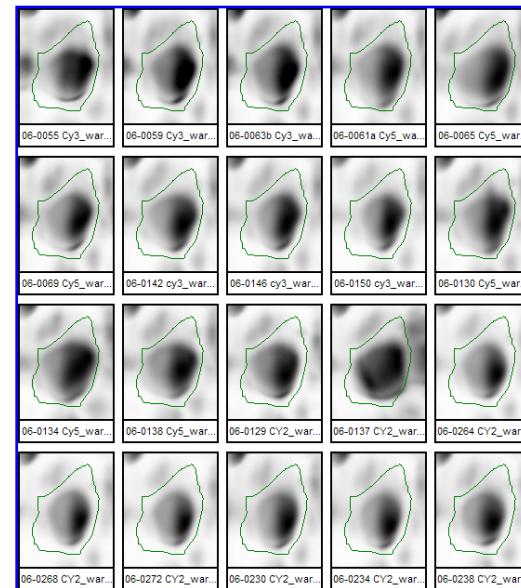


# Traitement d'images: détection/quantification

Application de la même bordure de spot à tous les gels: SameSpot

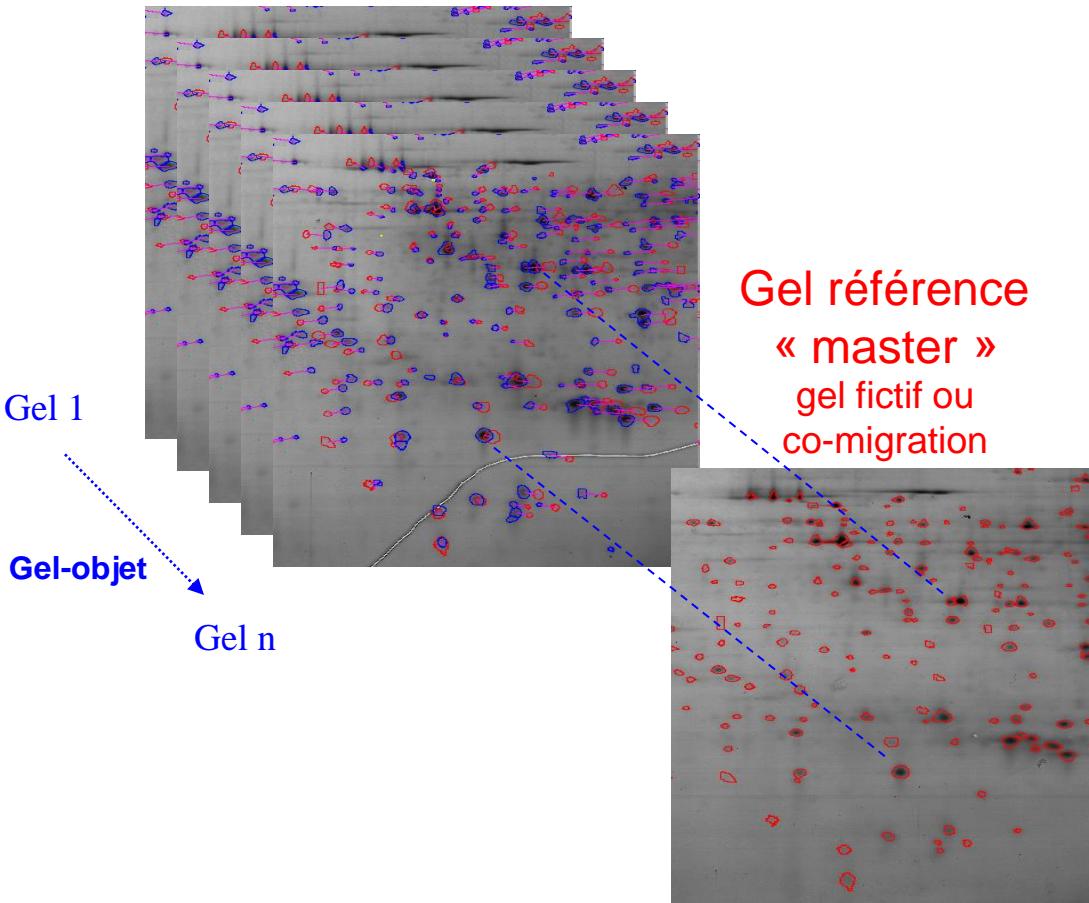


Analyse traditionnelle



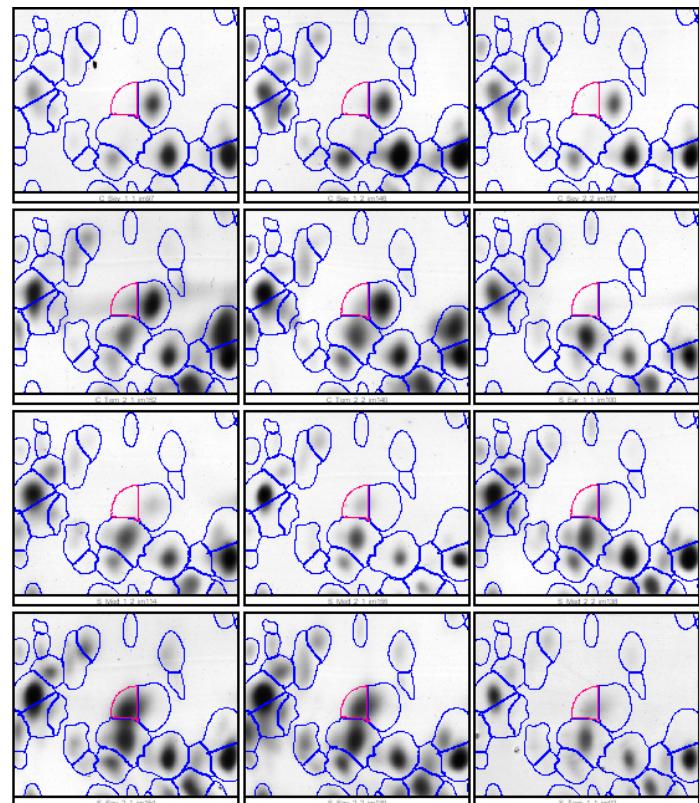
Analyse SameSpots

# Traitement d'images: appariement



appariement à réaliser plus ou moins de manière automatique.  
Aucun logiciel n'est meilleur que l'œil humain !  
Cependant, étant donné le nombre de spots considérés  
(2000 x 50), l'assistance informatique est indispensable.

Progenesis :  
tous les gels présentent la même  
« grille » rendant l'appariement inutile.



## Traitement d'images: normalisation

Corriger les variations de charge du gel (mauvais contrôle de la concentration de protéine déposée ou de la coloration).

=> Les volumes des spots sont normalisés, en général division par un facteur traduisant le niveau de coloration du gel.

Plusieurs méthodes de normalisation:

1. Prise en compte du volume de la **totalité des spots** détectés, appariés ou non.

Pb: Le rectangle dans lequel cette somme est faite doit être défini de manière identique dans tous les gels.  
Possibilité de déterminer une zone de calibration différente de la zone de détection par programmation hors logiciel 2D.

2. Prise en compte de tous les spots **présents dans tous les gels**.

Avantage: a priori seulement de vrais spots, mêmes spots pour tous gels.  
Pb: risquent d'être peu nombreux dans les analyses à grande échelle.

3. Normalisation par rapport à une liste de **spots connus d'avance**.

Pb: ne doivent pas être trop peu nombreux.

4. Normalisation par rapport à un **gel de référence**:

pour tout gel-objet, calcul de la médiane ou moyenne du rapport  $\text{vol}_{\text{ref}}/\text{vol}_{\text{obj}}$  pour tous les spots appariés dans les 2 gels.

Avantages:

- ne dépend pas de la précision de la délimitation de la zone d'intérêt.
- permet un calcul sur un grand nombre de spots.
- sélection possible des spots ni trop gros ni trop petits pour la calibration.
- ne prend pas en compte les spots à variation qualitative.

## Traitement d'images: exportation des données

Une fois normalisées, les quantités de protéines peuvent être exportées sous forme tabulée pour les analyses statistiques.

Master	Gel 1	.....	.....	.....	Gel n
Spot 1					
Spot n					

# Statistiques

appliquées à la 2-DE

# Statistiques: dispositif expérimental

## Les sources de variation

- Variations techniques
  - Variations liées à la migration 1D et 2D: dépendent de la série
  - Variations de coloration : dépendent de la série (AgNO<sub>3</sub>)
  - Variations liées à la préparation de l'échantillon: peuvent dépendre de la série aussi (solutions, tampons, centrifugations...)
- Variations biologiques
  - Variations entre individus subissant le même traitement
  - Variations entre individus subissant des traitements différents

La variation liée à la série est importante: en tenir compte dans le dispositif

Idéal: tous échantillons dans la même série.

Possible seulement si les séries sont grandes: pouvoir faire au minimum des séries de 12 gels

Si l'expérience impose plus de gels que de gels possibles dans une série

*Ex: 10 traitements A...J, 3 répliquats par traitement*

Série 1
A
B
C
D
E
F
G
H
I
J

Série 2
A
B
C
D
E
F
G
H
I
J

Série 3
A
B
C
D
E
F
G
H
I
J

Tous les traitements sont représentés dans chaque série

- S'il y a un effet série, il ne sera pas confondu avec l'effet traitement
- On pourra tenir compte de l'effet série lors de l'analyse statistique

# Statistiques: dispositif expérimental



Série 1

A	A	A
B	B	B
C	C	C
D	D	D

Série 2

E	E	E
F	F	F
G	G	G
H	H	H

Série 3

I	I	I
J	J	J

Effet série indissociable de l'effet traitement

si A, B, C et D sont différents des autres, est-ce dû à la série ou à ces traitements ?

## Dans la pratique:

- Accidents, gels ratés... souvent cette situation idéale n'est pas atteinte
- Nombre de traitements supérieur au nombre de gels dans une série.

Règle générale: maximum de traitements représentés dans chaque série

En cas de série supplémentaire contenant des gels ratés dans les précédentes, les accompagner d'autres gels.



Série 1

A	B	C
D	X	F
G	H	I
J		

Série 2

A	B	C
D	X	F
G	H	I
J		

Série 3

X	B	C
D	E	F
G	H	I
J		

Série 4

E	E	A

# Statistiques: répétitions nécessaires

## Tests statistiques: répétitions requises

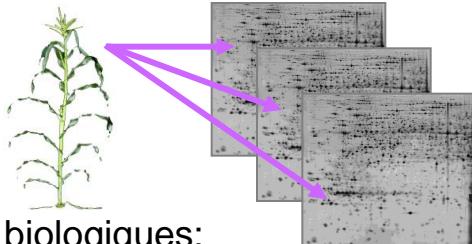
Répétitions: estimation de la variance résiduelle: la variation non contrôlée.

Effet traitement significatif :

variance liée au traitement > variance résiduelle.

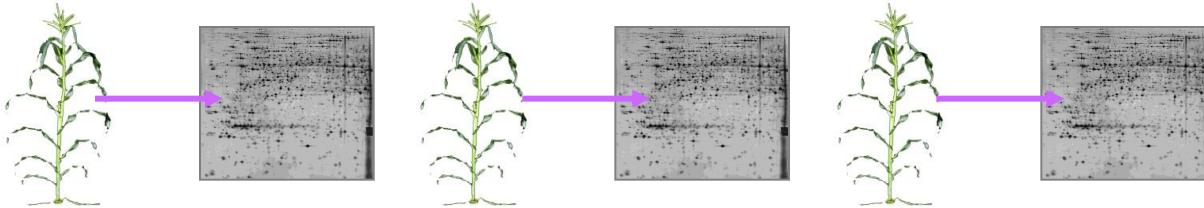
Répétitions techniques:

**gels provenant du même individu (voire de la même extraction)**



Répétitions biologiques:

**chaque gel provient d'un individu différent**



La variation entre répétitions biologiques contient à la fois la variation technique et la variation biologique: nécessairement plus grande que la variation technique.

Les tests statistiques sont plus significatifs avec les répétitions techniques mais les variations entre plantes sont dans la variance traitement et pas dans la résiduelle.

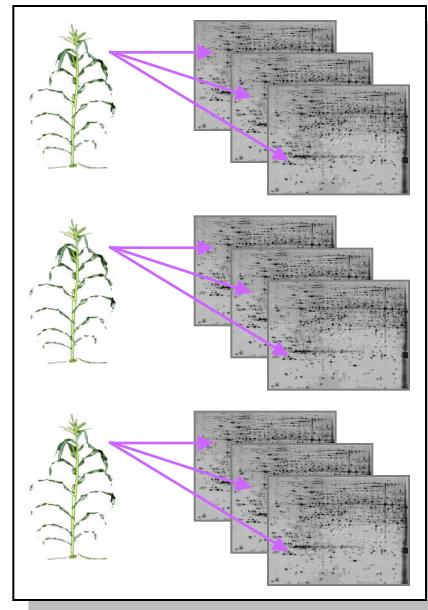
⇒ Sans répétitions biologique, à quoi est due la signification ?

Variations individuelles ?

Variations liées à l'extraction ?

Variation liée au traitement ?

=> Utiliser les répétitions biologiques: teste l'effet du traitement

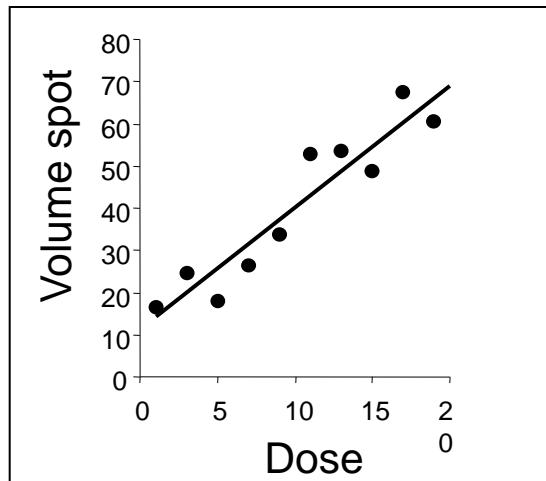


# Statistiques: répétitions nécessaires

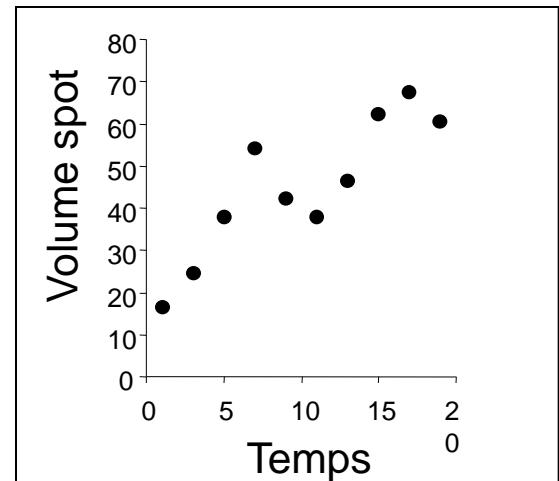
**Répétitions non nécessaires quand le facteur analysé est quantitatif**  
(corrélations, régressions)

Attention:

- Nombre de points
- Relation linéaire



Pas de répétition biologique  
Pas de linéarité  
=> Perte de résolution



**Réduction de la variation causée par les variations individuelles**

Chaque extraction faite à partir de plusieurs individus

Idéal: plusieurs répétitions biologiques, chacune composée d'un mélange de plusieurs plantes

**Nombre de répétitions**

Important : nombre de gels qui serviront à estimer la variance résiduelle.

Empiriquement:

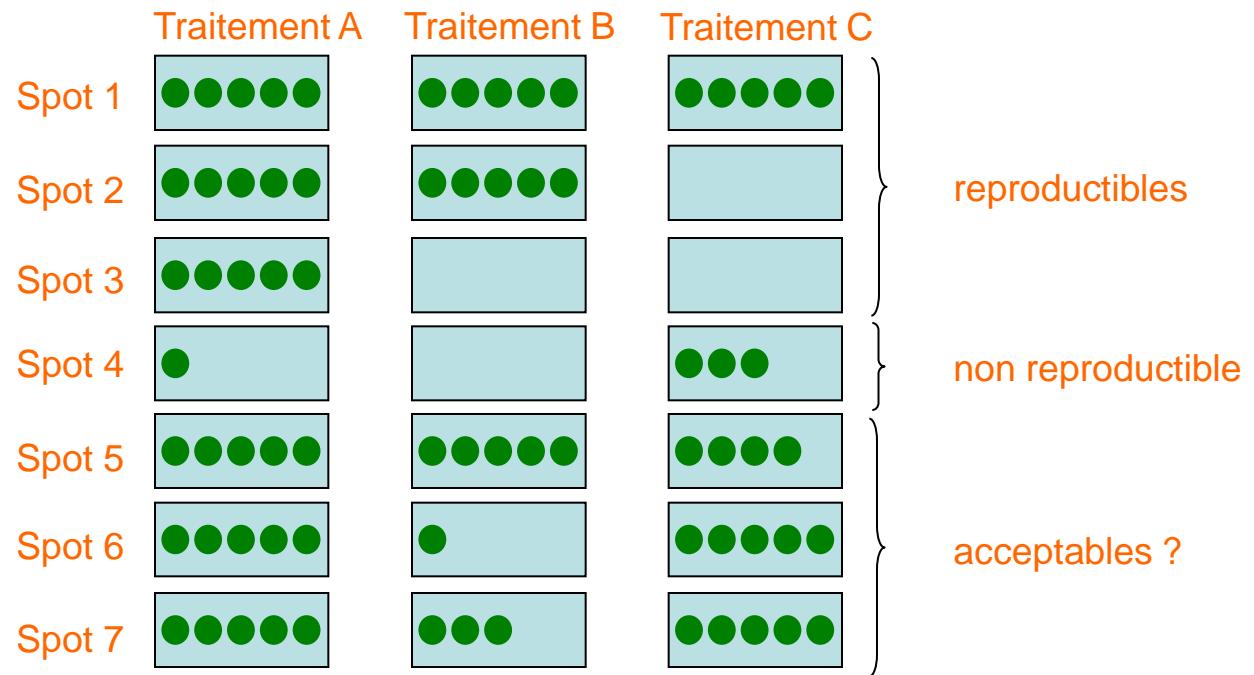
Pas moins de 3 répétitions/traitement

Pas moins de 5 répétitions/traitement s'il n'y a que 2 traitements

# Statistiques: spots reproductibles et variations quantitatives

Les spots reproductibles ne sont pas présents partout: peuvent être absents dans un ou plusieurs traitements.  
Combien de données manquantes tolérer ?

A partir de quand décider qu'un spot est présent ou absent dans un groupe?



Critères pouvant être utilisés

## 1) au niveau du groupe

« est présent dans un groupe de n gels tout spot présent au moins n-x fois » (x=1)

« est absent dans un groupe tout spot présent au plus y fois » (y=0)

## 2) au niveau de l'expérience

### Spots reproductibles

« Est reproductible dans l'expérience avec G groupes tout spot trouvé au moins G-z fois présent ou absent au niveau du groupe ».

### Spots qualitativement variables

« Est qualitativement variable dans l'expérience tout spot reproductible dans l'expérience et absent dans au moins un groupe ».

# Statistiques: spots reproductibles et variations quantitatives

## Différentes questions

- Y a-t-il des phénomènes généraux qui affectent simultanément de nombreuses protéines ?
- Ces phénomènes, s'ils existent, peuvent-ils être expliqués par des variables biologiques ?
- Peut-on mettre en relation la fonction des protéines identifiées avec les regroupements de spots apparemment co-régulés ?

*Matrice des corrélations*

*Analyse des composantes principales*

*Classification hiérarchique*

- Comment identifier les protéines qui répondent à un traitement donné

*Analyses de variance*

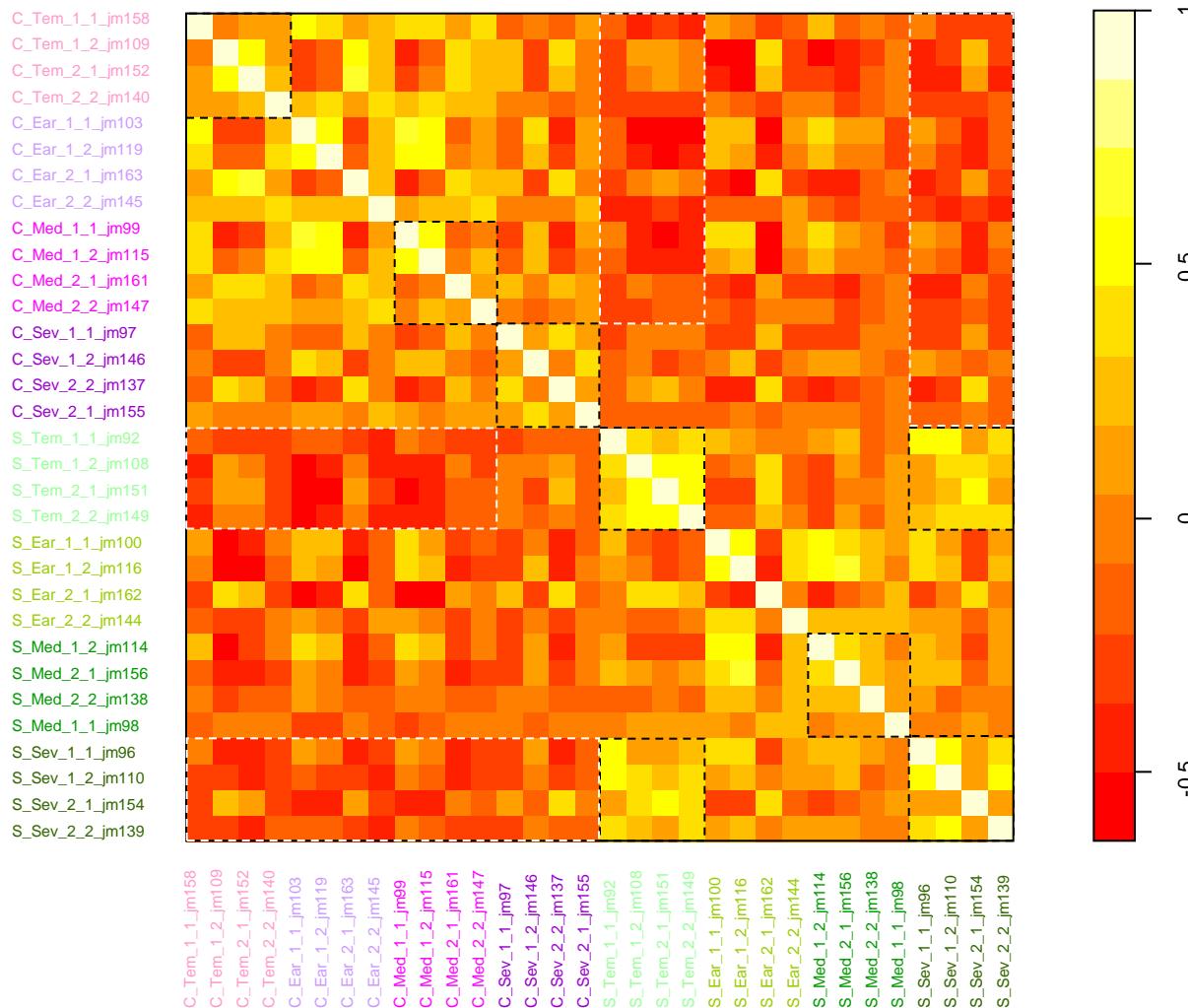
Les analyses multivariées (ACP, « clustering ») donnent des informations sur la structure des données en isolant des groupes de protéines présentant des profils d'expression similaires ou opposés.

Les analyses univariées (ANOVA) testent chaque protéine indépendamment les unes des autres. On perd donc les corrélations entre protéines mais cette analyse est bien plus robuste.

Analyses multivariées et univariées sont complémentaires.

# Statistiques: matrice de corrélation

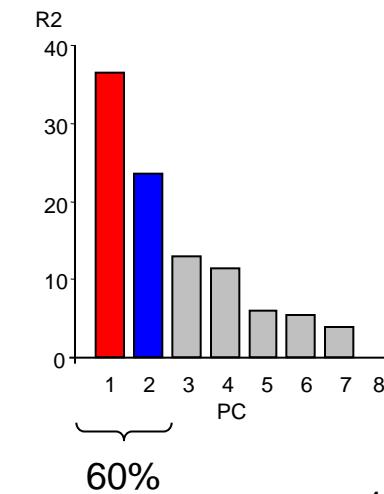
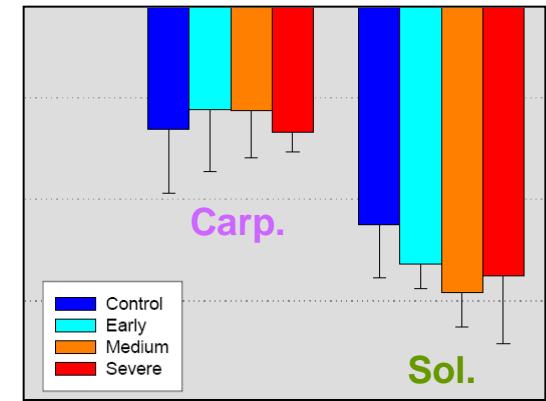
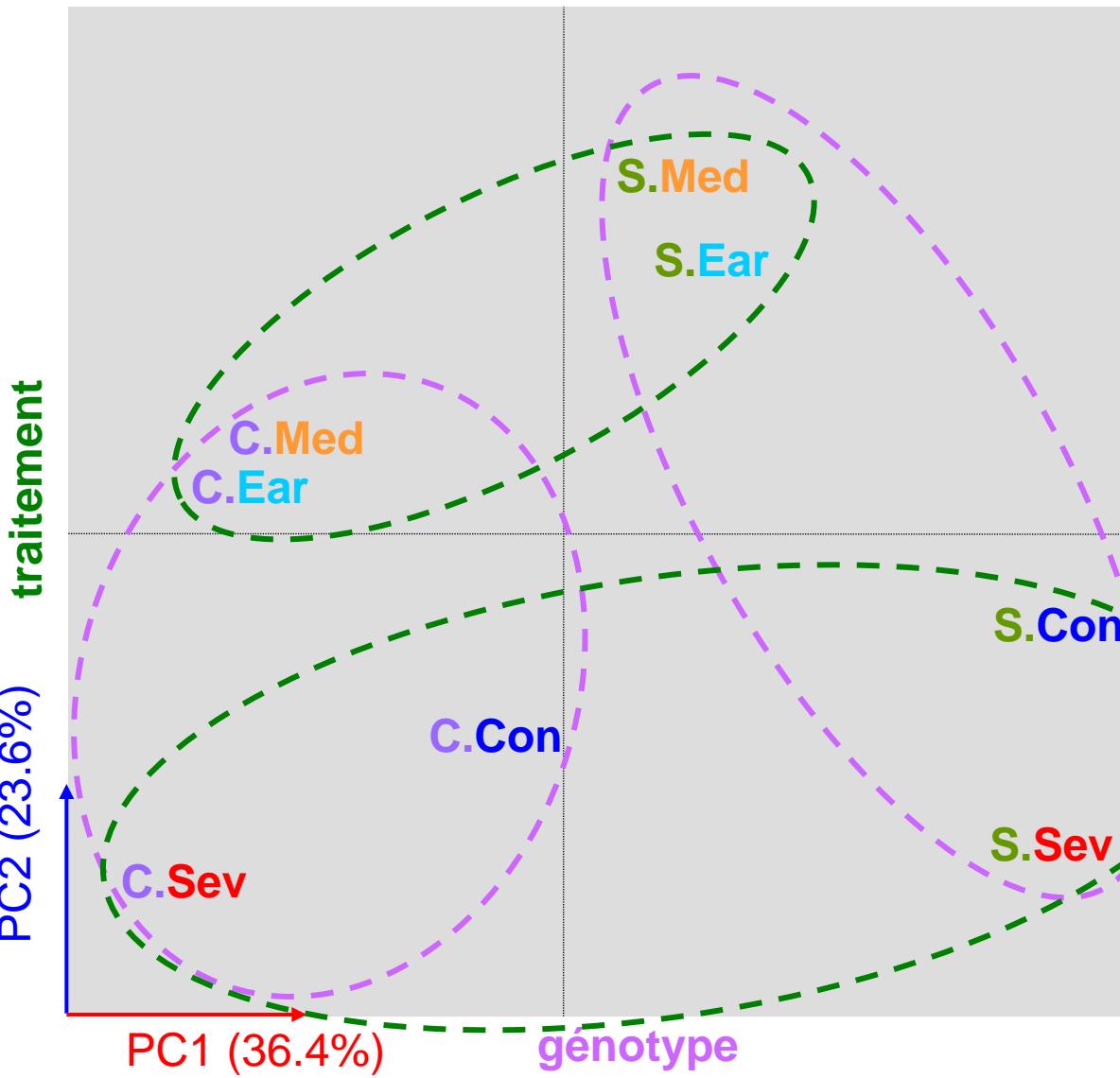
Permet de vérifier la qualité des données.



# Statistiques: Analyses en Composantes Principales (ACP)

Interprétation des CP.

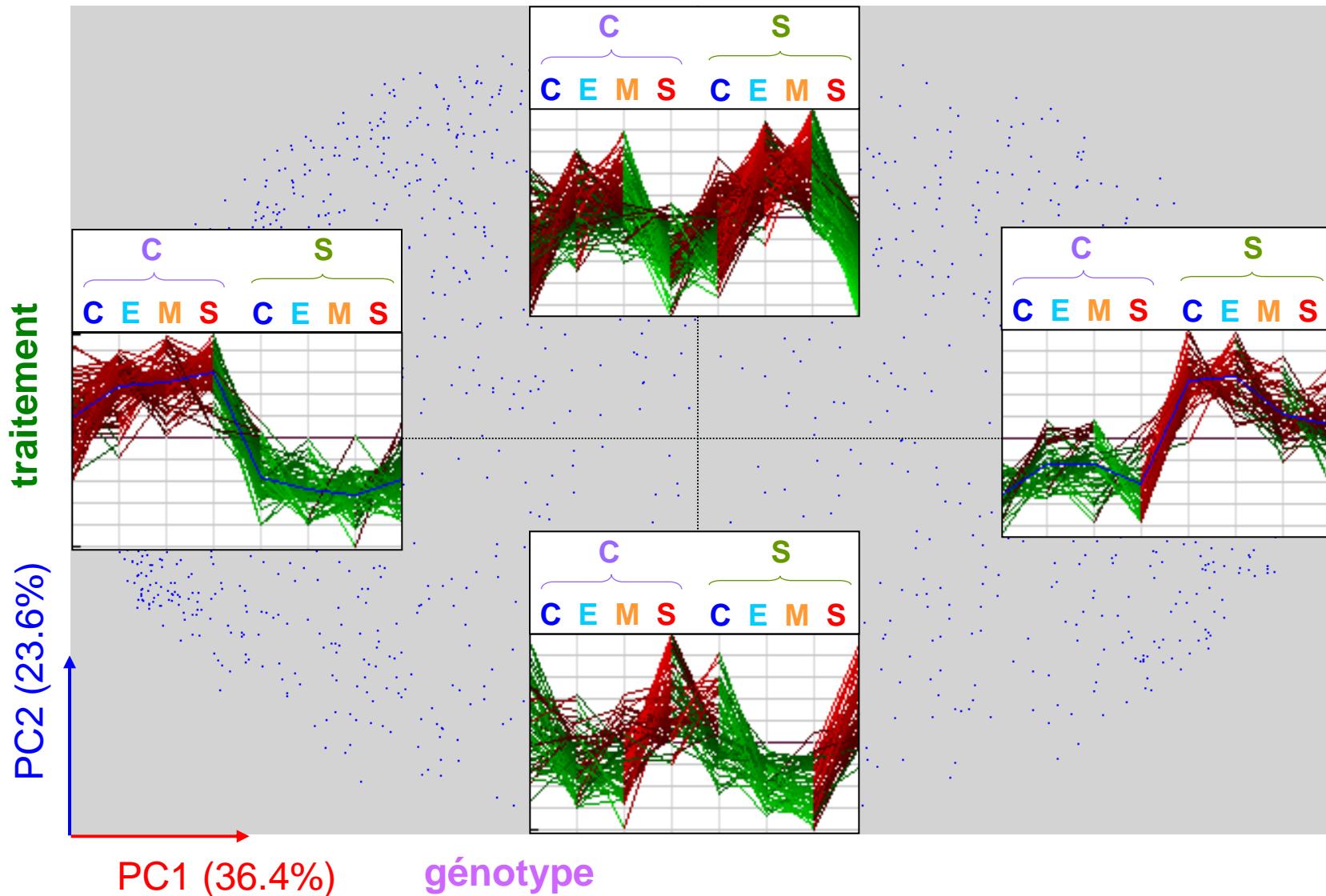
Représentation des variables (conditions expérimentales).



# Statistiques: Analyses en Composantes Principales (ACP)

Interprétation des CP.

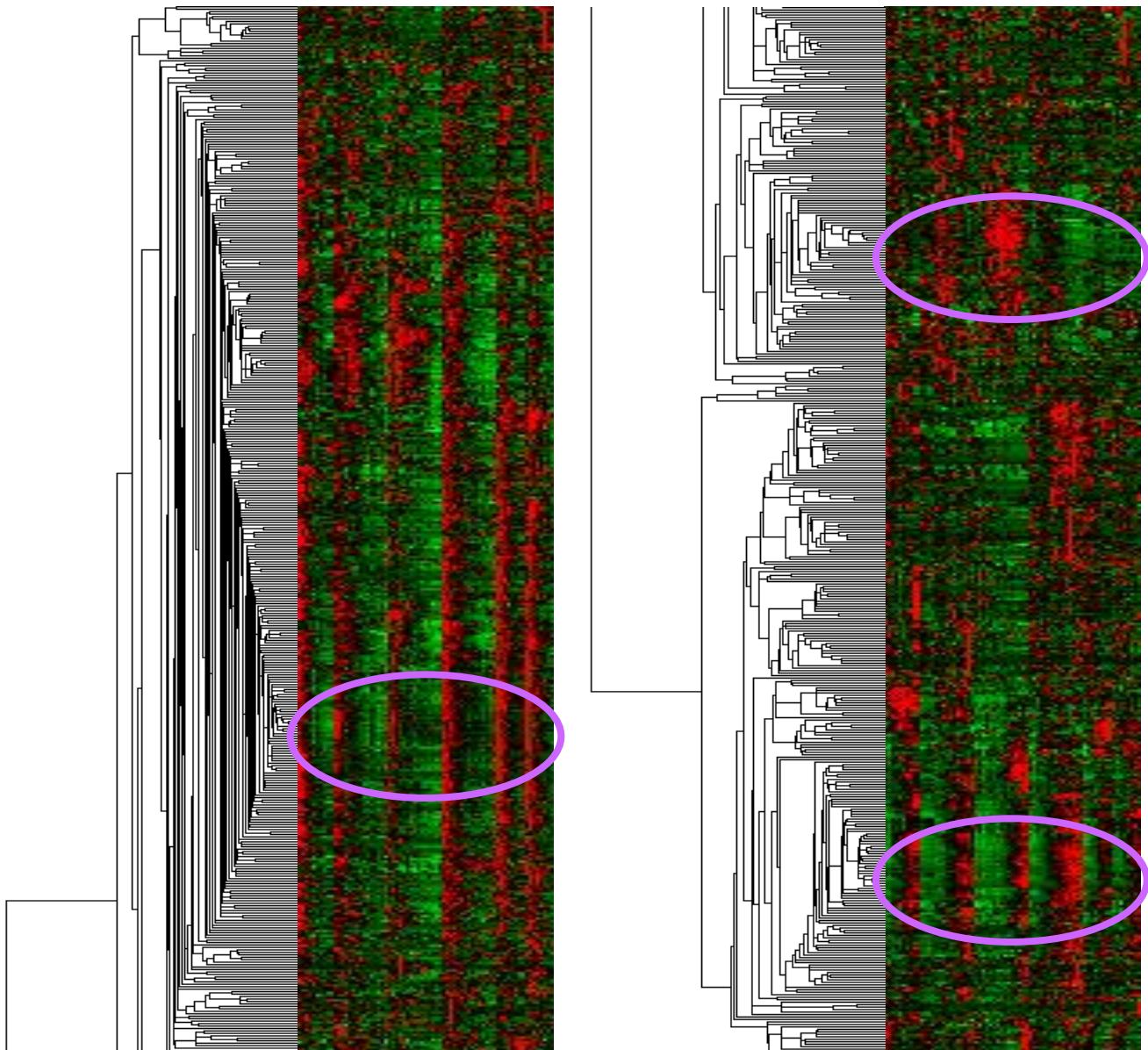
Représentation des individus (protéines).



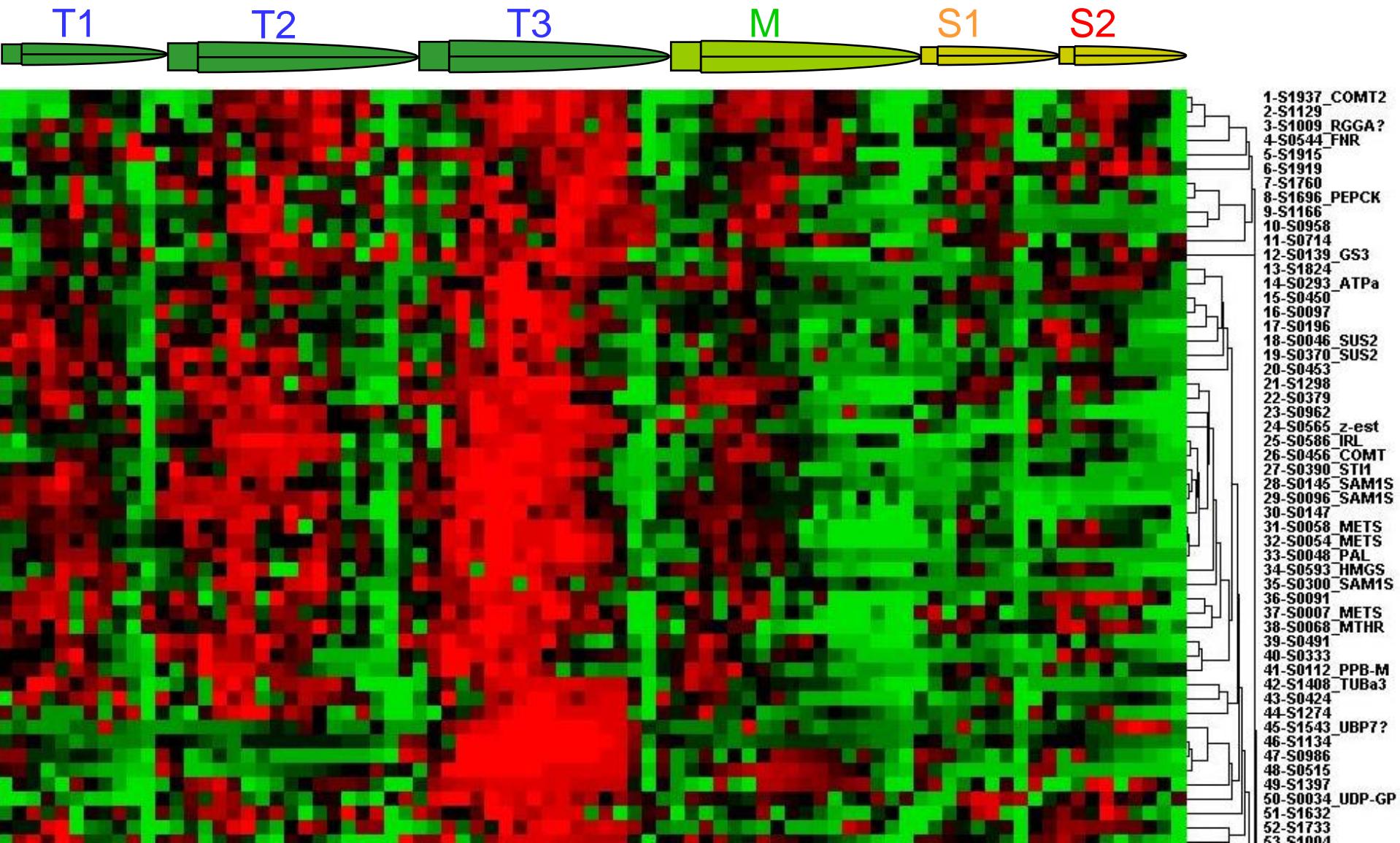
# Statistiques: Analyse en Classification Hiérarchisée (HCA, « clustering »)

- 850 protéines (*lignes*)
- 70 conditions :  
segment/stade/trtmt  
(*colonnes*)

Down-regulation  
Up-regulation



# Statistiques: Analyse en Classification Hiérarchisée (HCA, « clustering »)



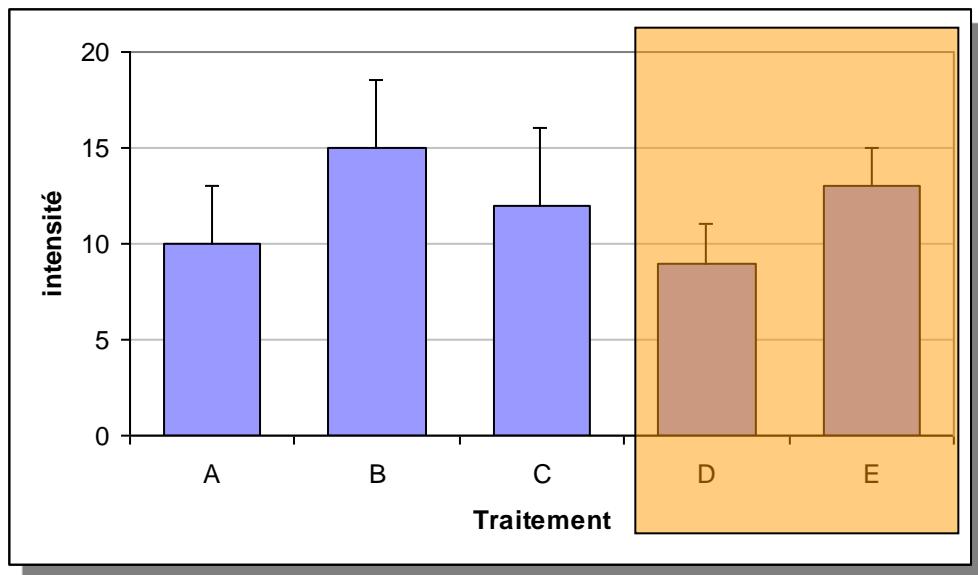
# Statistiques: analyse de variance (ANOVA)

L'analyse de variance est une comparaison de moyennes.

Cas le plus simple: analyse à un facteur (ex: le traitement)

L'ANOVA permet de dire s'il existe une différence significative entre au moins la plus petite et la plus grande moyenne.

Quand il y a plus de 2 traitements, il est nécessaire de faire d'abord une ANOVA avant toute comparaison individuelle. Même si un Student est significatif entre deux moyennes, la différence ne doit pas être considérée comme significative si l'ANOVA ne l'est pas.

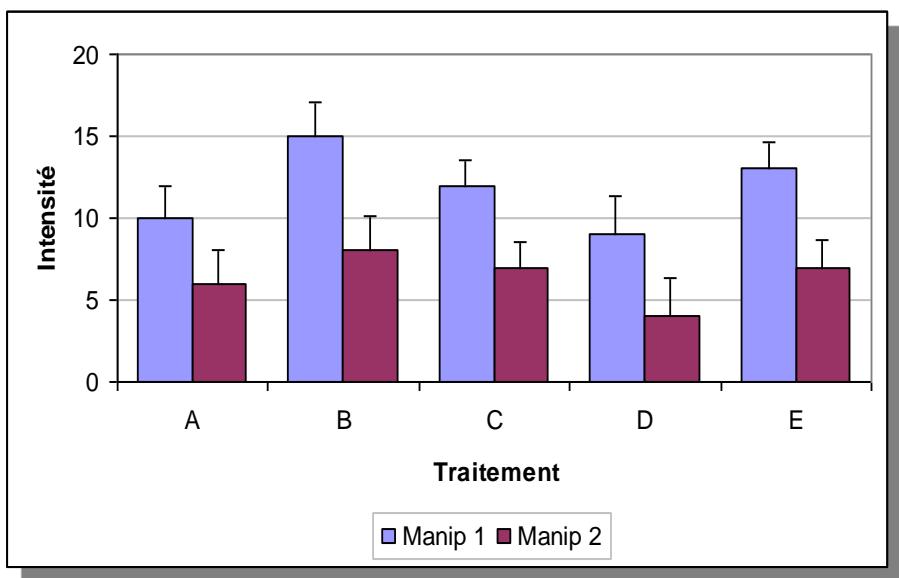


*Même si le t de Student est significatif entre les traitements D et E, l'ANOVA, en estimant la variance résiduelle sur l'ensemble des moyennes, peut indiquer qu'aucune différence n'est significative dans cet essai.*

# Statistiques: analyse de variance (ANOVA)

## L'analyse de variance à deux facteurs

Ex: analyse de l'effet du traitement dans deux expériences différentes.



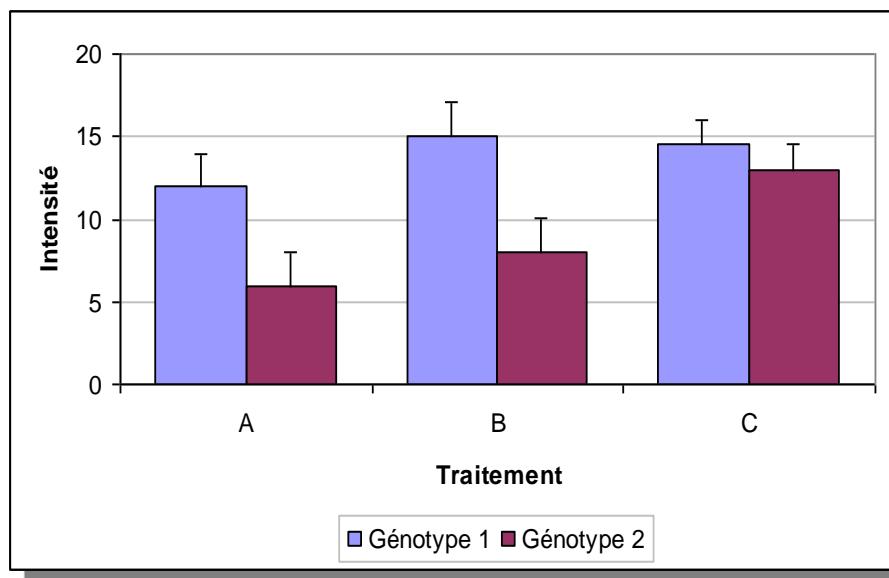
L'analyse à deux facteurs permet non seulement de savoir s'il existe un effet expérience, mais aussi de mettre en évidence l'effet traitement, qui serait noyé si on ne tenait pas compte de l'expérience.

## L'interaction entre plusieurs facteurs

Ex: analyse de l'effet du traitement dans deux génotypes

On choisit un modèle avec interaction:

$$\text{Intensité} = \text{Trait} + \text{Géno} + \text{Trait} * \text{Géno}$$



Les effets principaux (traitement, géotype) et l'interaction vont être testés par l'ANOVA. Une interaction significative signifiera que les deux génotypes ne réagissent pas de la même manière aux traitements

# Statistiques: analyse de variance (ANOVA)

## Très nombreux modèles possibles

- 1 facteur: effet du traitement
- 2 facteurs: effet du traitement + effet de la série
- 2 facteurs avec interaction: effet du traitement + effet de la date + interaction date/traitement
- Etc...

## Résultat

Effet significatif à 1% ( $P < 0,01$ )

=> Différence significative au moins entre la plus petite et la plus grande moyenne

## Comparaisons entre moyennes

Faire ensuite des comparaisons de moyennes suivant la question posée:

Comparaisons multiples (groupes), comparaisons à un témoin (Dunnet), contrastes...

## Attention

- L'ANOVA ne tient pas compte des variations qualitatives
- Ne pas remplacer les données manquantes
- Toujours vérifier les effectifs quand on a autorisé des données manquantes

## Comparaisons multiples

Risque alpha=1%  
Test sur 1000 spots } Environ 10 spots seront significatifs même si le traitement n'a aucun effet.

### Bonferroni:

Diviser le risque pris pour chaque spot par le nombre de spots analysés

⇒ Très sûr mais perte de beaucoup de vrais significatifs

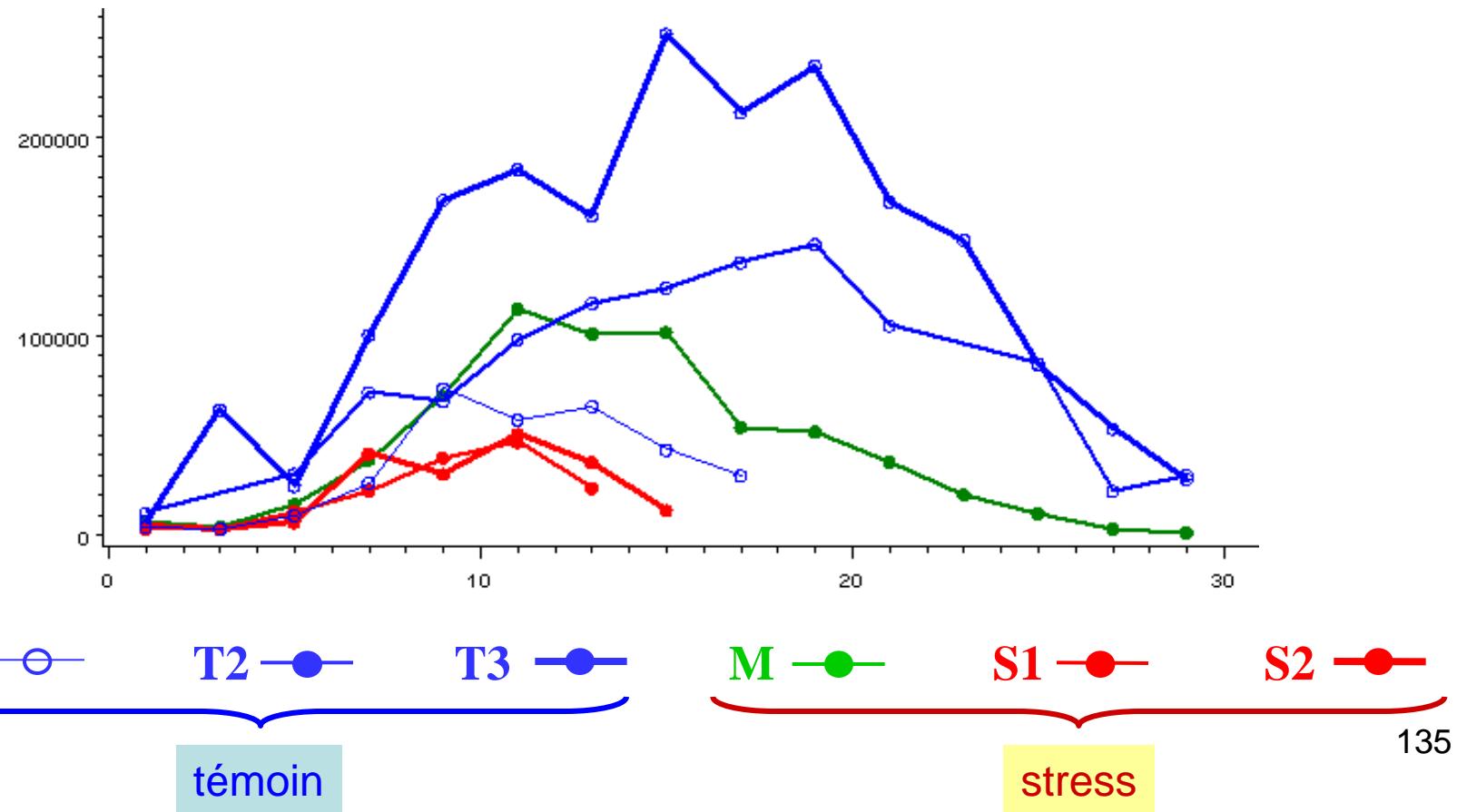
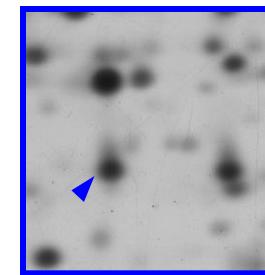
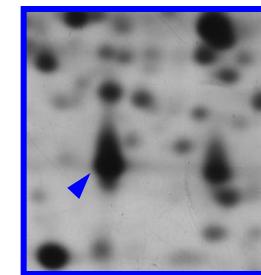
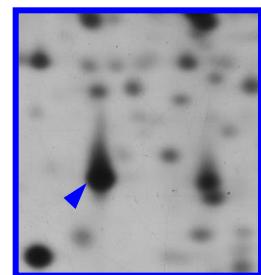
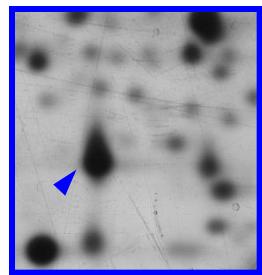
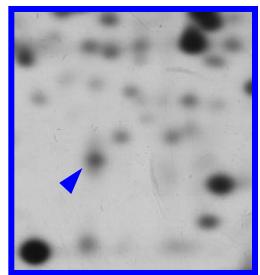
méthode FDR (False Discovery Rate, Benjamini et Hochberg, 1995):

Accepter que 1% des spots déclarés significatifs soient des faux positifs

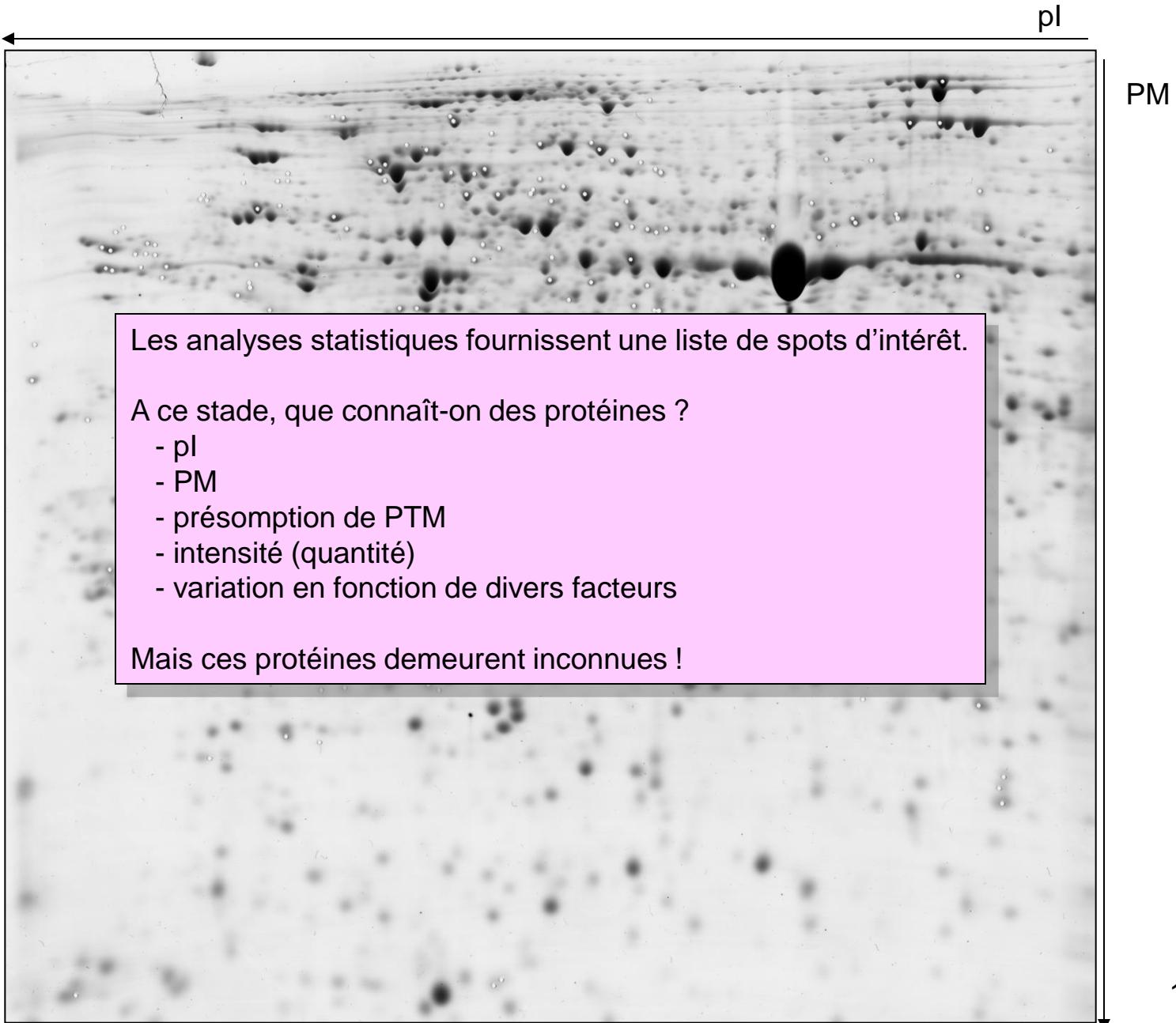
## Statistiques: vérification par observation des profils de protéines

Ex. spot 456.

Effet significatif pour traitement.

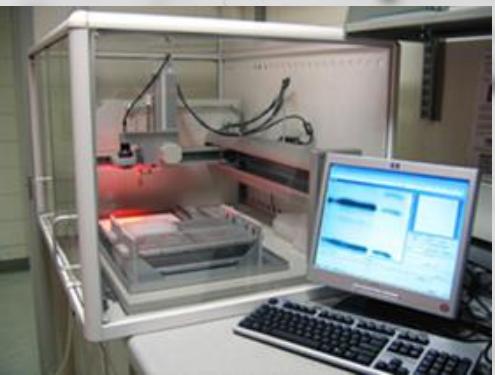


## Statistiques: liste de spots d'intérêt



# Identification des protéines

# Identification des protéines: excision



## Identification des protéines: excision

Les spots à exciser sont définis soit manuellement soit automatiquement par le logiciel d'analyse d'image.

L'excision ne doit pas être réalisée à sec (film d'eau recouvre le gel pour éviter qu'il ne sèche ce qui modifierait sa taille –donc la position des spots- et l'abîmerait).

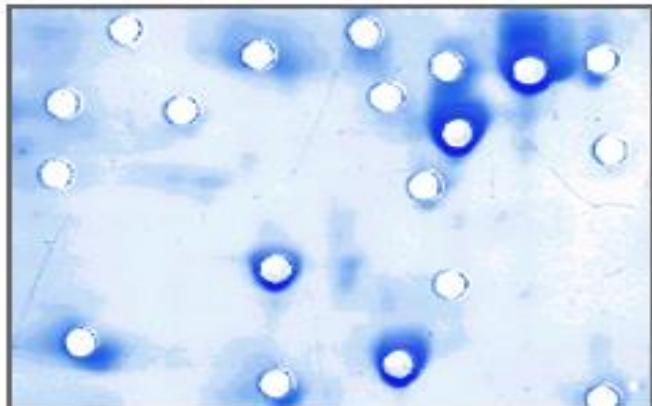
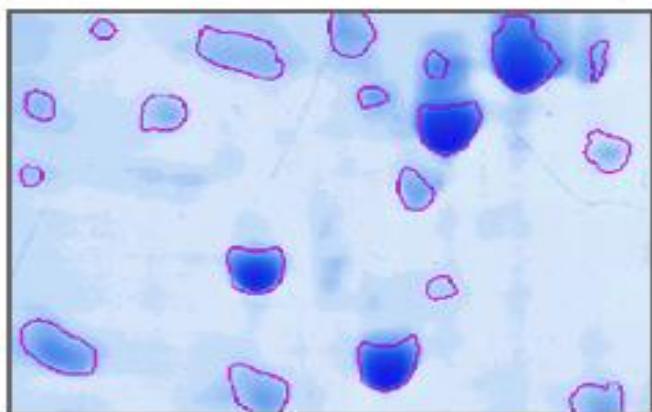
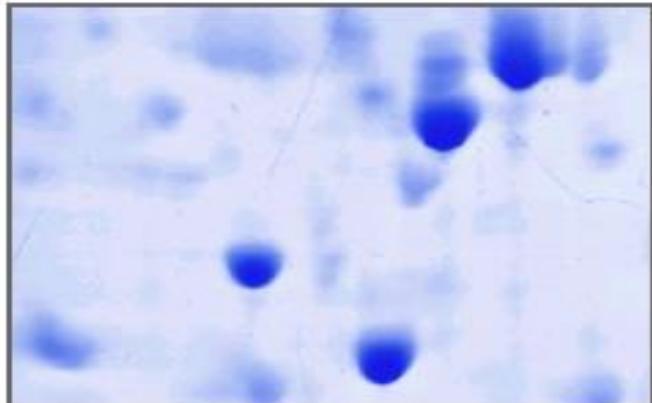
Il est important de ne pas contaminer les échantillons successivement excisés. Il faut donc soit changer de cône de pipette (excision manuelle) à chaque nouveau spot, soit soigneusement rincer les cônes ou l'aiguille (robot).

La taille du cône ou de l'aiguille peut être adaptée à celle des spots à exciser.

Il faut exciser au centre du spot, dans la zone la plus intense, afin de limiter les contaminations par les spots du voisinage (plus ou moins chevauchant) et de maximiser la quantité de molécules excisées.

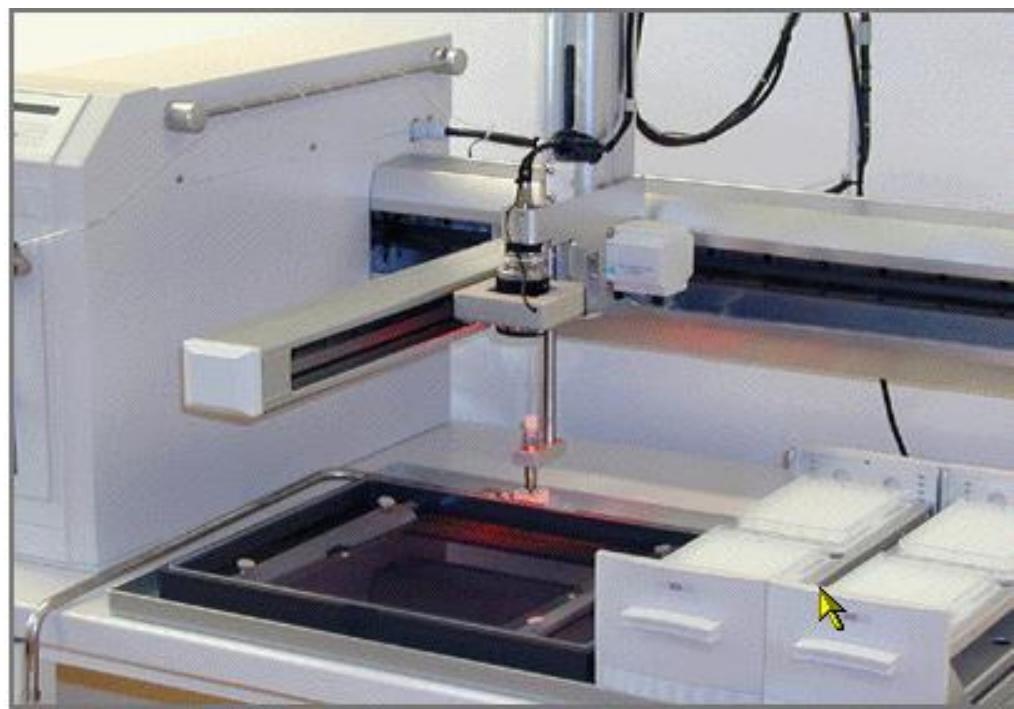
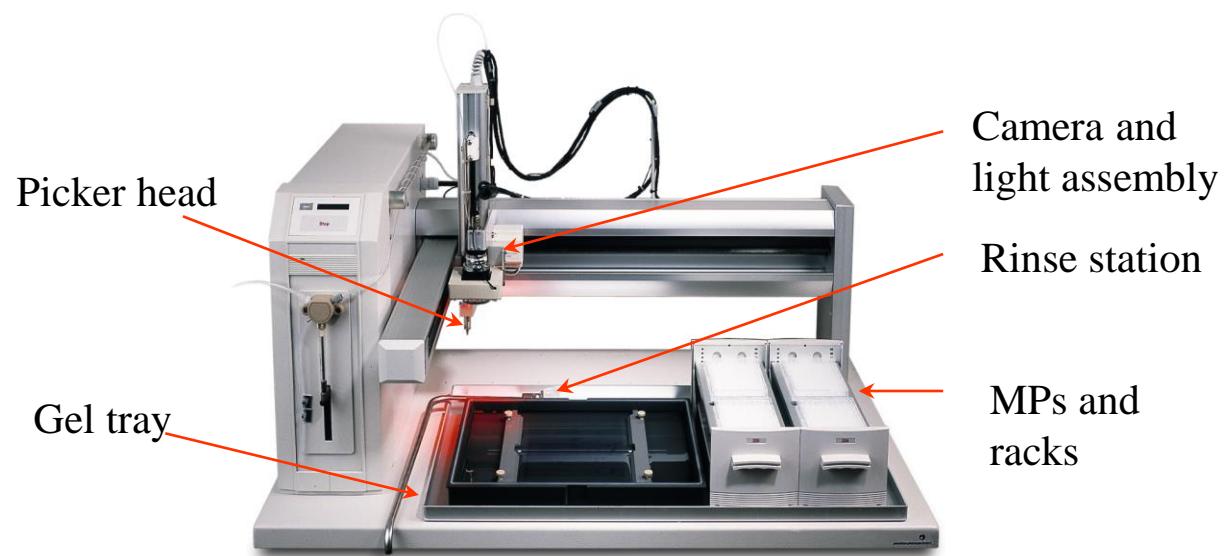
Il est possible d'exciser le même spot sur différents gels afin d'augmenter sa quantité, en particulier lorsque la coloration est très sensible ( $\text{AgNO}_3$ ) et que le nombre de molécules par spot est réduit par rapport à une coloration moins sensible (CBB).

Il est important de porter gants, masque et charlotte afin de limiter les contaminations par kératine.



# Identification des protéines: excision automatique

## Spot picker



# Identification des protéines: méthodes de clivage

## Clivages enzymatiques:

Trypsine → lysine K et arginine R

Endoprotéinase Arg-C → arginine R

Endoprotéinase Lys-C → lysine K

Endoprotéinase Glu-C → acides glutamique E et aspartique D

Endoprotéinase Asp-N → acide aspartique D et cystéine C

Chymotrypsine

Thermolysine

Elastase...

## Clivages chimiques:

CNBr → methionine M

Acide formique (70%) → proline P et acide aspartique D

CNBr + HFBA → méthionine M et tryptophane W

Hydroxylamine → glycine G et asparagine N

L'efficacité de la digestion dépend de la composition en AA de la protéine à identifier.

On peut ainsi essayer diverses méthodes de clivage si une protéine s'avère récalcitrante à la méthode utilisée en routine.

La digestion à la trypsine est la méthode la plus fréquemment employée.

# Identification des protéines: digestion à la trypsine

## 1 - Décoloration

En cas de coloration au CBB, décolorer les bandes sous agitation douce dans environ 150 µl d'Ammonium Bicarbonate ( AmBic 25 mM)/acétonitrile (ACN 50%) jusqu'à complète décoloration (1 à 4 h selon l'intensité de la coloration).

En cas de coloration au AgNO<sub>3</sub>, décolorer les bandes sous agitation douce dans environ 150 µl de Ferricyanide de potassium (2% w/v) et du Thiosulfate de sodium (3,2% w/v) jusqu'à complète décoloration. Rincer dans 150 µl d'eau environ jusqu'à complète décoloration en renouvelant si nécessaire.



## 2 – Déshydratation avec ACN

Enlever la solution précédente puis ajouter 150 µl d'ACN.

Agiter doucement pendant 10 min jusqu'à ce que le gel devienne blanchâtre.

## 3 - Séchage

Enlever l'ACN et sécher 5 min au speedvac ou 20 min à température ambiante.

## 4 – Digestion à la trypsine (coupe lysine K et arginine R)

Reprendre 20 µg de trypsine dans 200 µl d'HCl 1 mM et répartir en 4 tubes de 50 µl.

Noter la date et stocker à -20°C pendant 2 semaines maximum.

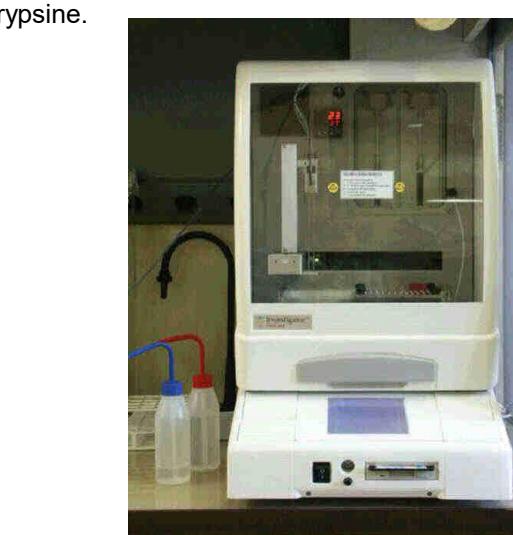
Avant utilisation, diluer la trypsine ainsi préparée dans 450 µl d'AmBic 50 mM froid.

Recouvrir le gel de trypsine à 10 ng/µl (10 à 25 µl selon les cas) et laisser gonfler pendant 10 min à 4°C.

Si nécessaire, centrifuger 30 sec pour faire descendre les morceaux du gel au fond du tube puis enlever l'excès de trypsine.

Si nécessaire, ajouter 5 µl bicarbonate d'ammonium 50 mM.

Incuber une nuit à 37°C.



## 5 - Récupération Des Peptides Hydrophiles

Ajouter 10 µl d'AmBic 50 mM (ou plus si nécessaire) au mélange de digestion et incuber 10 min sous agitation.

Si nécessaire, centrifuger 30 sec.

Récupérer le surnageant dans un tube ou une microplaquette propre.

## 6 - Extraction Des Peptides Hydrophobes

Ajouter 10 µl de tampon d'extraction (ou plus si nécessaire) au culot et incuber 10 min sous agitation.

Récupérer le surnageant et le réunir avec le précédent.

Ajouter 10 µl de tampon d'extraction (ou plus si nécessaire) au culot et incuber 10 min sous agitation.

Récupérer le surnageant et le réunir avec le précédent.

Ajouter 10 µl de tampon d'extraction (ou plus si nécessaire) au culot et incuber 10 min sous agitation.

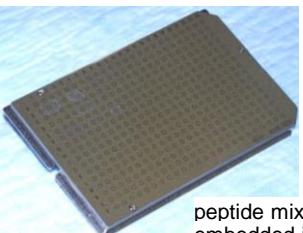
Récupérer le surnageant et le réunir avec le précédent.

## 7 – Séchage et acidification

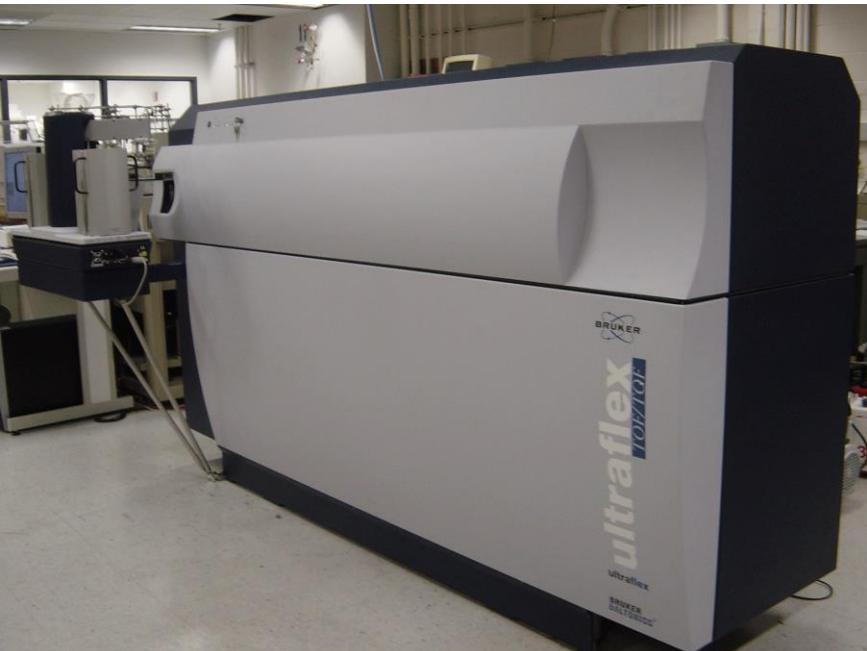
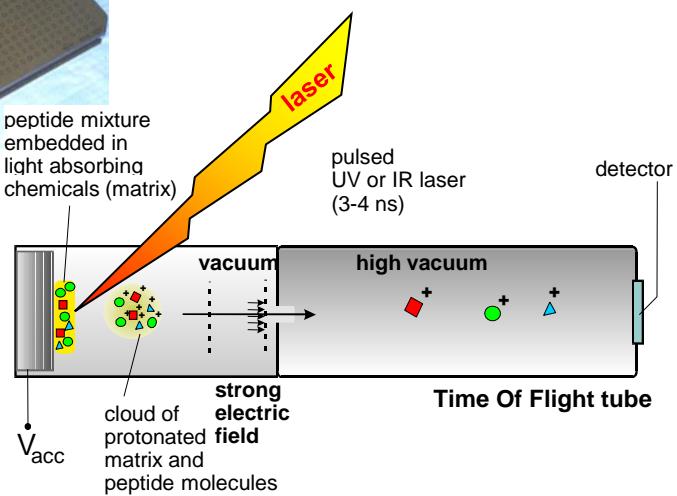
Réduire le volume par évaporation au speed-vac à température ambiante jusqu'à obtenir environ 25 µl puis ajouter 0,06 µl d'HCOOH par µl. En cas de volume inférieur, compléter avec de l'eau + 0,1% HCOOH.

# Identification des protéines: analyse par spectrométrie de masse (MS)

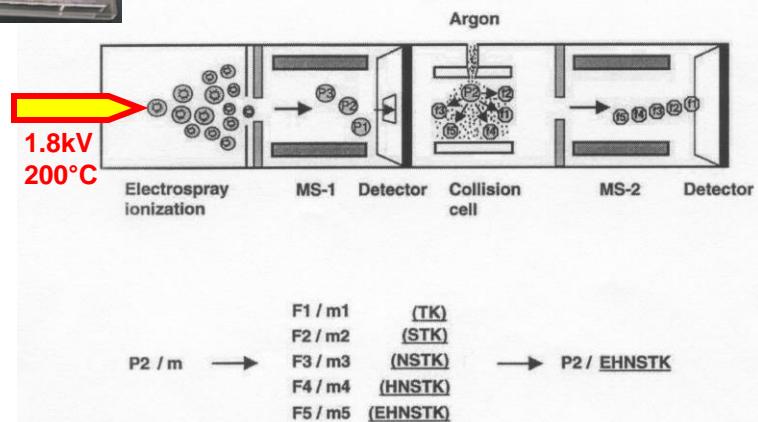
Les deux outils les plus fréquemment employés:



MALDI-TOF/TOF



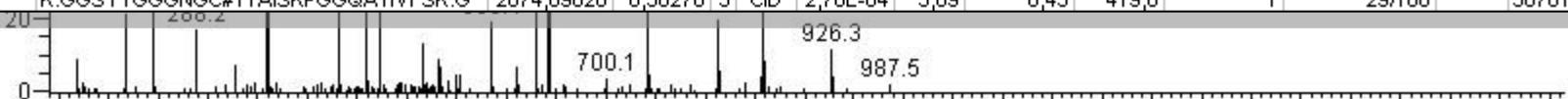
nLC-ESI-MS/MS



# Identification des protéines: interrogation des banques de données

RT: 0.00 - 65.08

Reference	Scan(s)	Peptide	P (pro)			Score Coverage		MW Accession Peptide (Hits)			% Area		
			MH+	DeltaM	z	Type	P (pep)	XC	DeltaCn	Sp	RSp	Ions Count	Peak Area
1 MGG_06197   Magnaporthe grisea spermidine synthase (335 nt)							4,47E-11	20,23	8,10	37850,1	335	2 (2 0 0 0 0)	6.61
1070 K.VLVIGGGDGGVLR.E			1212,42403	-0,08121	2	CID	3,24E-06	3,84	0,56	1079,1	1	19/24	766968384
1259 K.YQDVLFQSTDYGK.V			1677,83534	-0,34091	2	CID	4,47E-11	4,66	0,66	1535,1	1	22/26	908015744
2 MGG_13068   Magnaporthe grisea hypothetical protein (475 nt)							2,00E-09	40,23	10,50	53218,7	475	4 (4 0 0 0 0)	5.32
620 K.ADVVVVGLPKGEDEKL			1456,62321	-0,24670	2	CID	1,14E-05	3,65	0,48	1032,9	1	18/26	336023552
843 R.GFYETQSEDESQQGK.V			1718,75635	-0,95967	2	CID	2,00E-09	4,51	0,68	1680,5	1	22/28	366098176
1095 K.TSELVIFK.G			937,11480	-0,05225	2	CID	7,58E-04	2,93	0,42	776,8	1	12/14	171428704
1334 R.LPAVAPELYEELK.T			1472,70738	-0,43291	2	CID	1,38E-07	2,92	0,54	1107,7	1	21/24	473770272
3 MGG_04944   Magnaporthe grisea hypothetical protein (268 nt)							1,24E-08	20,19	9,70	27309,8	268	2 (2 0 0 0 0)	1.39
492 R.VITAGGVGEAC#SK.E			1249,38972	-0,44509	2	CID	1,24E-08	3,76	0,62	1836,5	1	20/24	125391128
614 K.GLGNNTGATTSQVYR.V			1324,42497	-0,52111	2	CID	1,02E-06	3,07	0,55	631,9	1	14/24	227421712
4 MGG_01085   Magnaporthe grisea hypothetical protein (249 nt)							1,97E-08	30,23	9,30	26539,3	249	12 (12 0 0 0 0)	81.03
571 K.VM*FGDEEEQR.I			1256,32002	-0,25063	2	CID	1,35E-05	3,18	0,55	795,4	1	16/18	1974796800
594 - 607 K.VM*FGDEEEQR.I			1256,32002	-1,39224	2	CID	6,25E-04	2,60	0,53	585,6	1	15/18	1710190720
631 K.VM*FGDEEEQR.I			1256,32002	0,20249	2	CID	2,82E-06	3,08	0,55	795,9	1	16/18	ND
663 K.VM*FGDEEEQR.I			1256,32002	0,33775	2	CID	2,71E-04	2,47	0,57	727,4	1	15/18	ND
738 K.VMFGDEEEQR.I			1240,32510	-0,33213	2	CID	1,34E-06	2,98	0,57	1013,5	1	16/18	1633217664
762 - 763 K.VMFGDEEEQR.I			1240,32510	-1,34568	2	CID	2,03E-06	2,62	0,37	933,0	1	16/18	1594789504
791 K.VMFGDEEEQR.I			1240,32510	-0,38047	2	CID	1,10E-06	2,74	0,47	1001,2	1	17/18	ND
970 R.DLQTSASAAEAYIR.A			1425,52607	-0,50942	2	CID	4,26E-06	3,58	0,57	1414,6	1	17/24	6627045376
979 R.DLQTSASAAEAYIR.A			1425,52607	-0,84858	1	CID	8,08E-04	2,26	0,40	198,9	1	12/24	229671312
994 R.DLQTSASAAEAYIR.A			1425,52607	-0,24233	2	CID	7,44E-06	4,28	0,55	2058,8	1	20/24	6753491456
1018 R.DLQTSASAAEAYIR.A			1425,52607	-0,35780	2	CID	1,97E-08	4,57	0,57	1902,9	1	19/24	ND
1042 R.DLQTSASAAEAYIR.A			1425,52607	-0,76674	2	CID	3,86E-08	3,21	0,63	1575,7	1	19/24	ND
5 MGG_00081   Magnaporthe grisea hypothetical protein (190 nt)							1,57E-05	20,18	14,30	19653,8	190	2 (2 0 0 0 0)	2.59
675 R.AGGGGSDDC#LGISAR.V			1393,43588	0,57833	2	CID	1,57E-05	3,61	0,48	1537,3	1	18/28	242274368
1083 R.GPFTIDADGNLR.V			1276,38031	0,67443	2	CID	7,47E-04	2,83	0,29	864,8	1	17/22	414653920
6 MGG_00505   Magnaporthe grisea hypothetical protein (417 nt)							9,56E-05	20,19	6,70	43597,6	417	2 (2 0 0 0 0)	3.05
786 K.PGGQATIVFSK.G			1105,26879	-0,64814	2	CID	9,56E-05	3,51	0,53	1442,9	1	17/20	265039264
938 K.GGSYTGGGNGC#TTAISKPGQQATIVFSK.G			2674,89820	0,36276	3	CID	2,70E-04	3,89	0,45	419,8	1	29/108	507614592



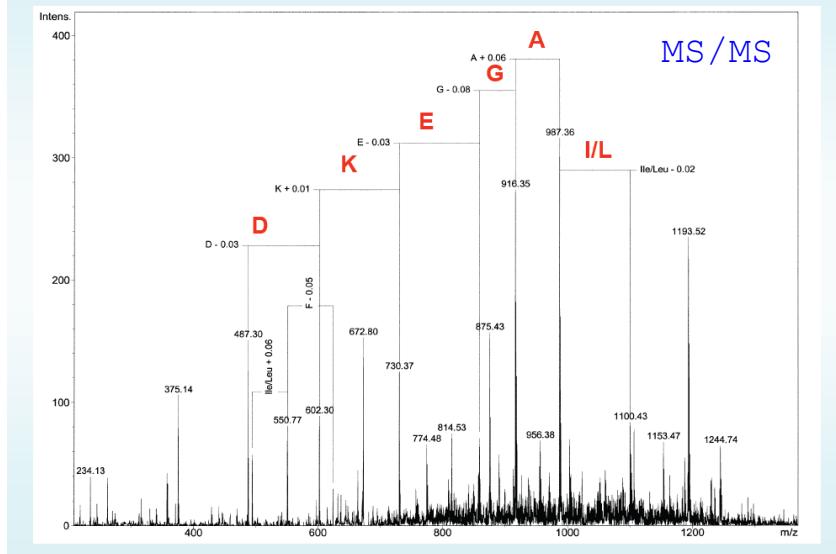
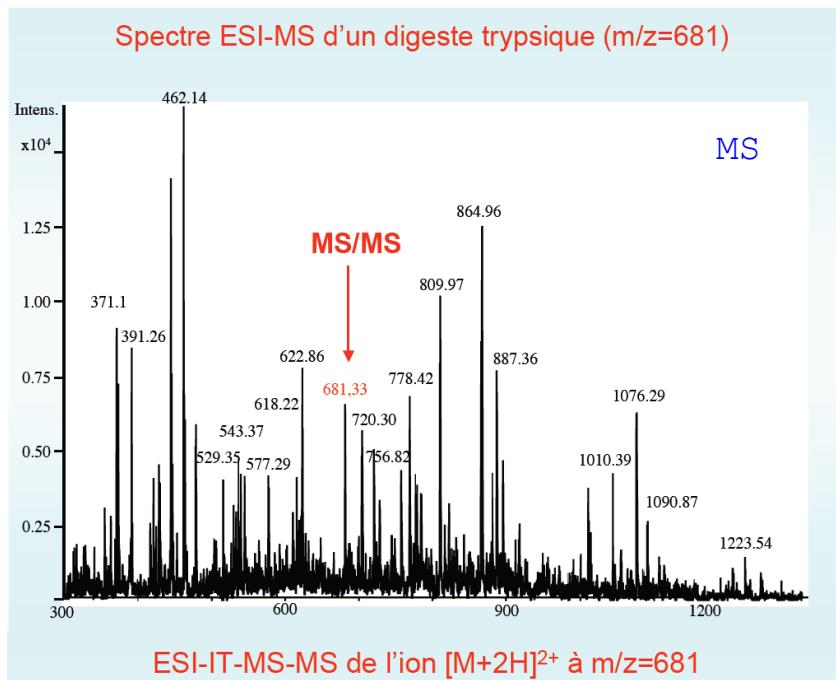
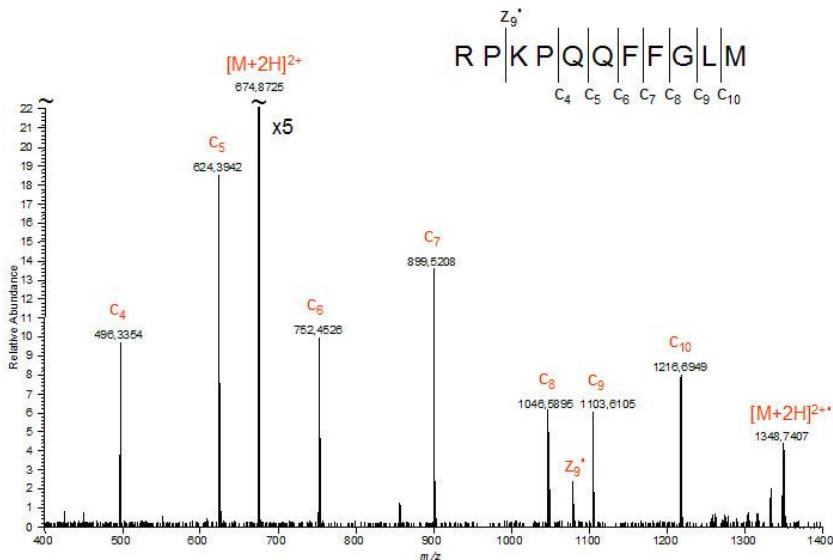
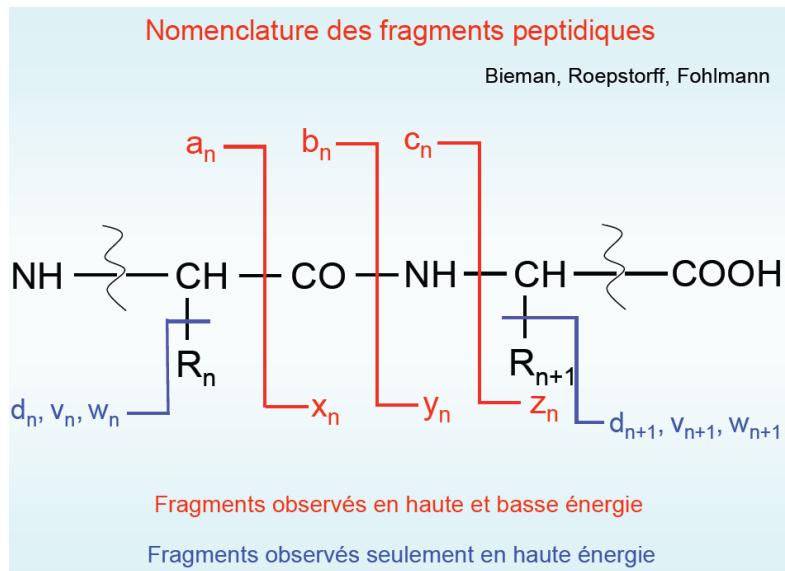
Identification tributaire des niveaux d'exhaustivité des banques de séquences et de la qualité d'annotation des séquences.

Absence de « hit »      ← mauvais spectre (faible spot, pb digestion, protéine incompatible avec MS)  
                                 ← séquence non existante

Que faire si la séquence protéique est absente des banques de données?

## Identification des protéines: séquençage *de novo*

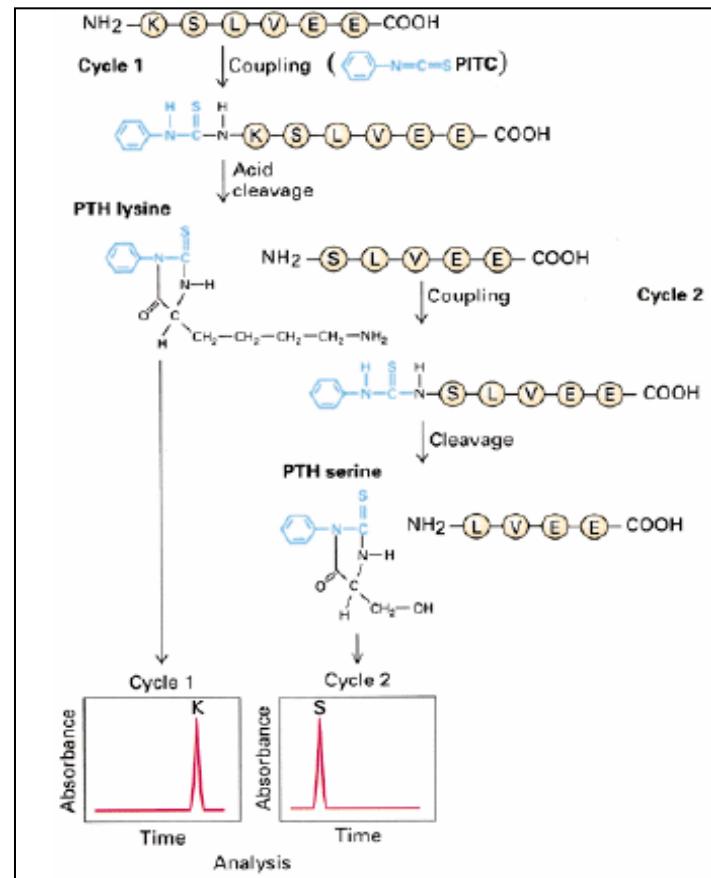
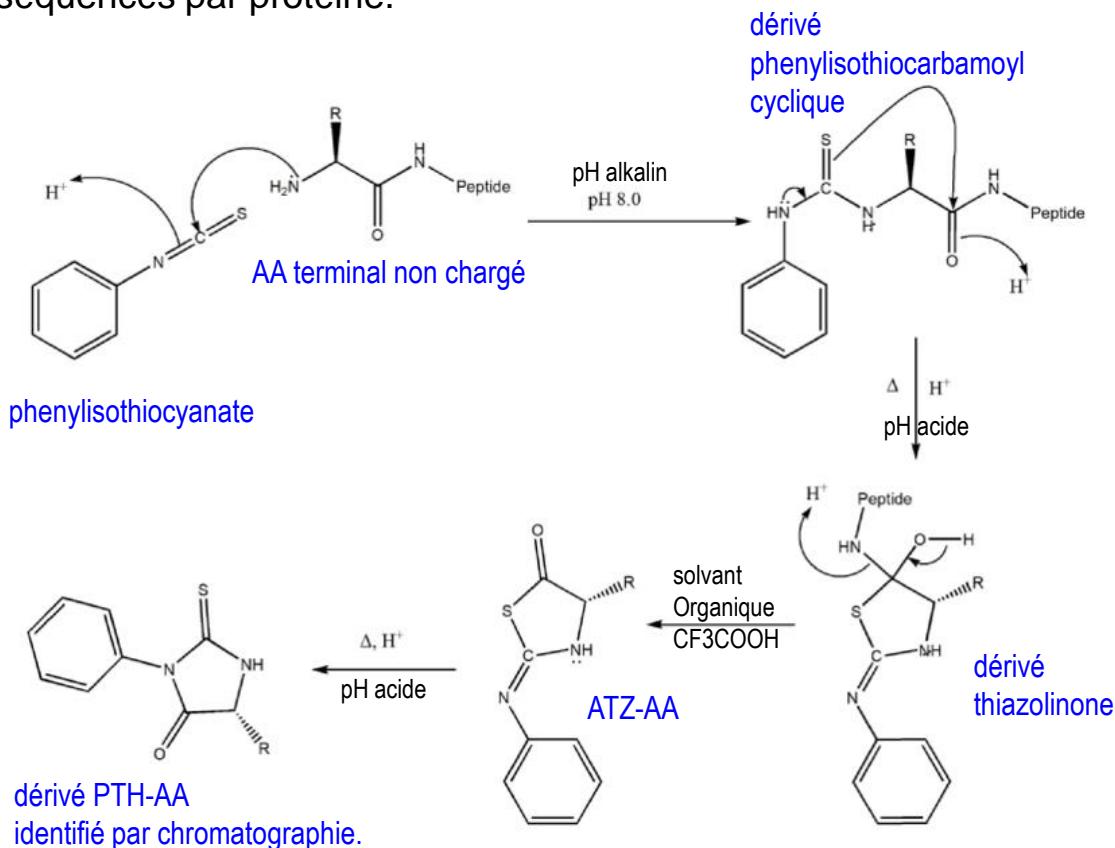
Interprétation manuelle des spectres de masses.  
Nécessite beaucoup de temps et une grande expertise.



# Identification des protéines: dégradation d'Edman

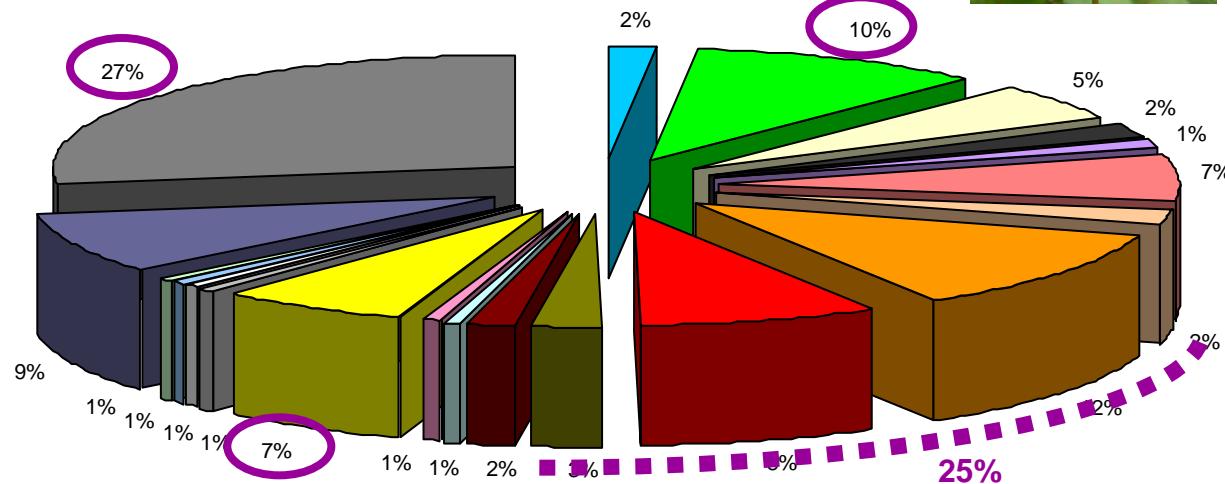
Edman P., 1949 « A method for the determination of amino acid sequence in peptides. »  
Archives of Biochemistry 22(3):475.

Principe: Résidu aminé N-terminal étiqueté puis clivé du reste de la protéine sans affecter les autres liaisons peptidiques. Processus itéré séquentiellement. En routine, 2 à 3 peptides comprenant 10-15 AA sont séquencés par protéine.

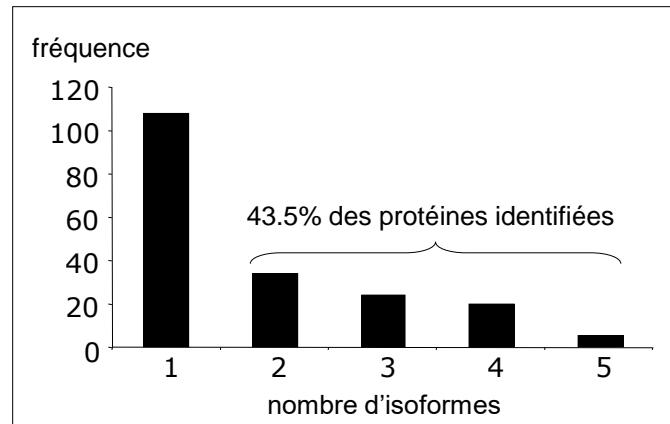
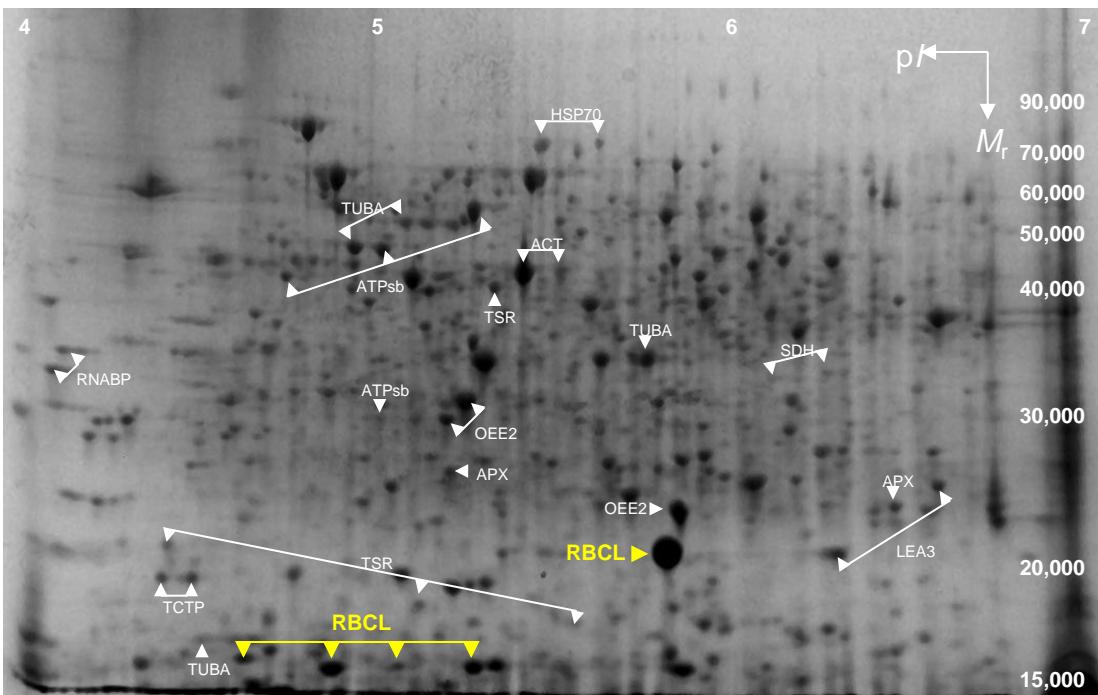


Limite: parce que ce procédé s'amorce à l'extrémité N-terminale d'une protéine, il échouera si l'AA N-terminal a été chimiquement modifié, ou bien caché au sein de la protéine et alors non accessible. Les ponts disulfures sont déterminés par une autre procédure.

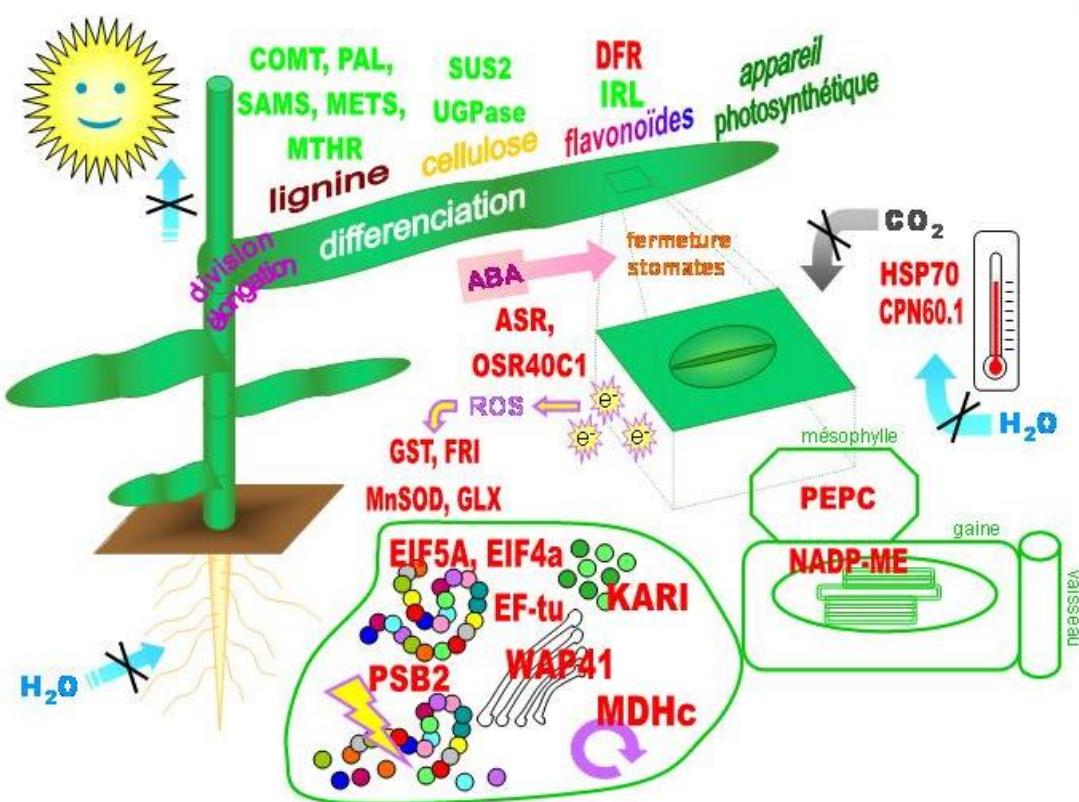
# Identification des protéines: signification biologique des résultats



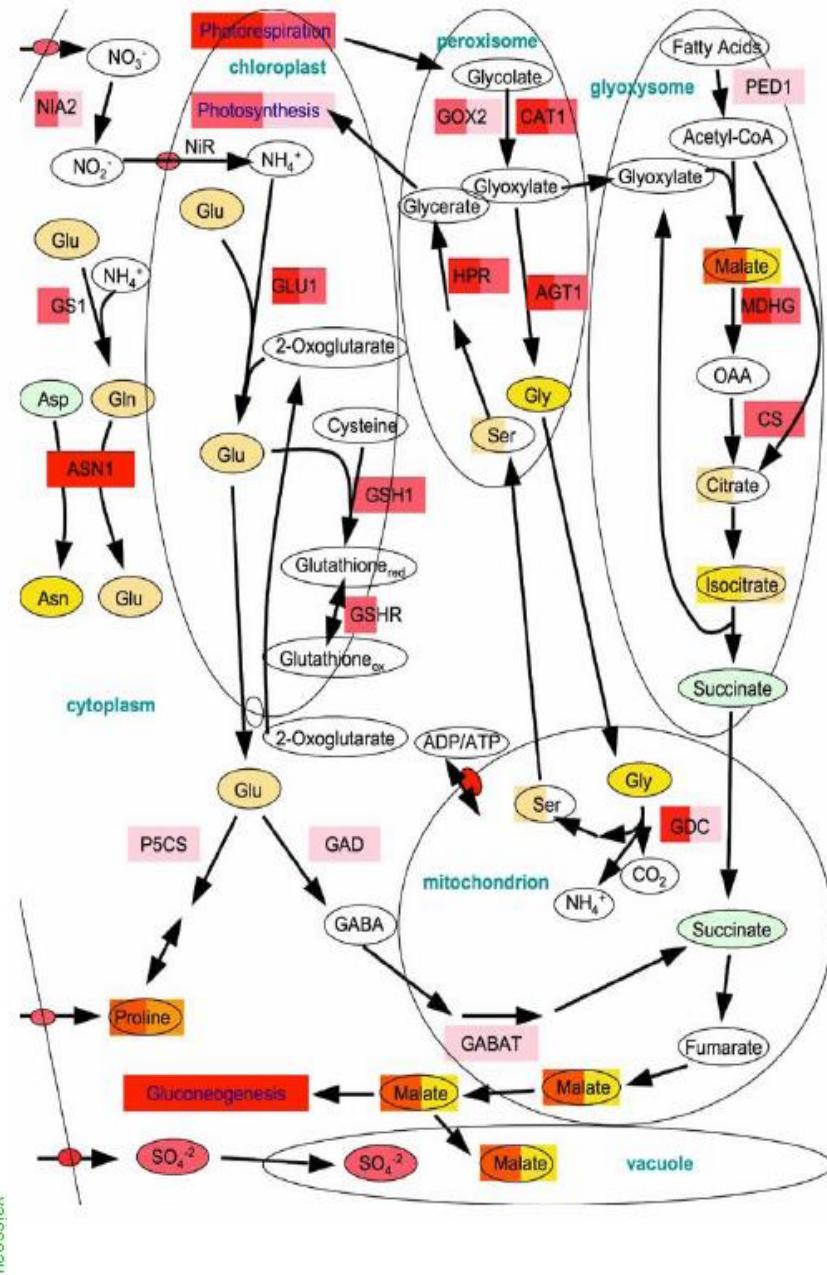
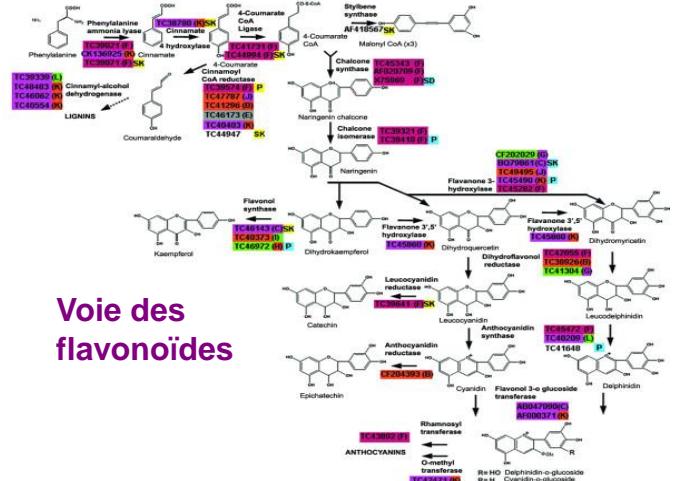
- █ glycolysis and gluconeogenesis
- █ photosynthesis
- █ electron transport/membrane-associated energy
- █ C-compound and carbohydrate metabolism
- █ amino-acid metabolism
- █ cellular biogenesis
- █ transcription
- █ protein synthesis
- █ protein destination
- █ protein degradation
- █ secondary metabolism
- █ cellular communication/signal transduction
- █ storage protein
- █ stress
- █ cellular transport and transport mechanisms
- █ lipid, fatty-acid and isoprenoid metabolism
- █ tricarboxylic-acid pathway
- █ spermidine metabolism
- █ unclassified protein
- █ unknown protein



# Identification des protéines: signification biologique des résultats



## Voie des flavonoïdes



Merci de votre attention