

# Genome Sequence Projects

From “One gene – one hypothesis” ....

...To Global, “Systems” approaches

## Higher-order networks

Genomes  Proteomes  Interactomes, Kinomes, etc...

## Novel technologies

Transcriptomics      Proteomics      Metabolomics      etc.

# Πρωτεωμική τεχνολογία

Ανάλυση του συνόλου (ή υποσυνόλων) των εκφραζόμενων πρωτεϊνών ενός κυτταρικού τύπου ή ιστού:

- Επίπεδα έκφρασης
- Μετα-μεταφραστικές τροποποιήσεις
- Δημιουργία συμπλόκων
- Ενζυμική δραστηριότητα

Εφαρμογές στην βιοϊατρική έρευνα:

- Άμεσες: Συσχέτιση προφίλ έκφρασης πρωτεϊνών με την διάγνωση (ή πρόγνωση της εξέλιξης) μιας ασθένειας
- Έμμεσες: Βασική έρευνα, καλύτερη κατανόηση μηχανισμών, ταυτοποίηση νέων στόχων

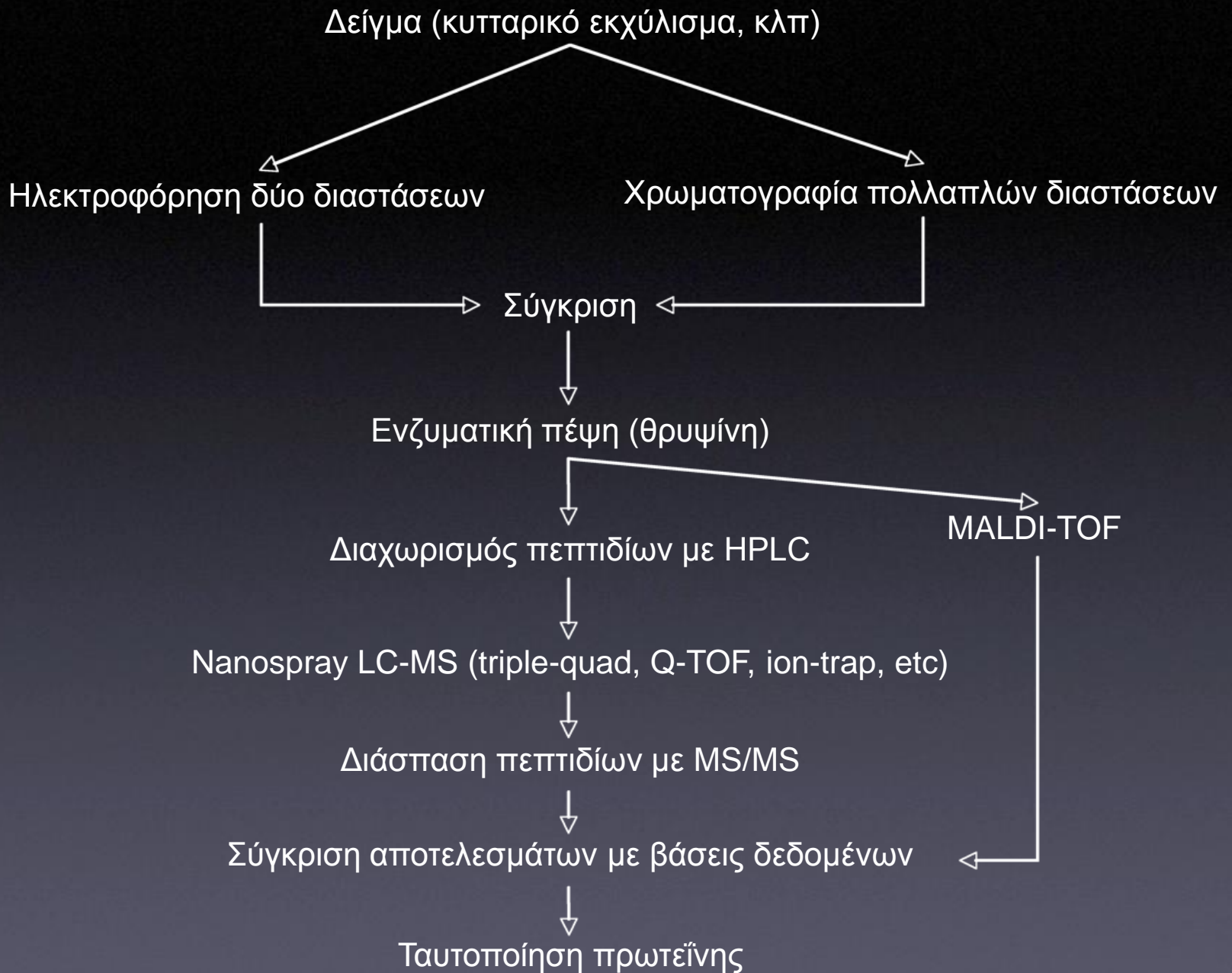
# Ανάλυση διαφορικής έκφρασης γονιδίων

## Μειονεκτήματα των μικροσυστοιχιών DNA:

- Δεν αναγνωρίζουν: Μετα-μεταφραστικές τροποποιήσεις  
Ενδοκυτταρικές μετατοπίσεις πρωτεϊνών  
Δημιουργία συμπλόκων
- Επίπεδα RNA  $\neq$  επίπεδα πρωτεΐνης

## Μειονεκτήματα της πρωτεωμικής τεχνολογίας:

- Χρονοβόρος διαδικασία / κόστος
- Χαμηλή ευαισθησία για πρωτεΐνες που εκφράζονται σε λίγα αντίτυπα
- Διαχωρισμός μικρού αριθμού πρωτεϊνών (1000-2000)

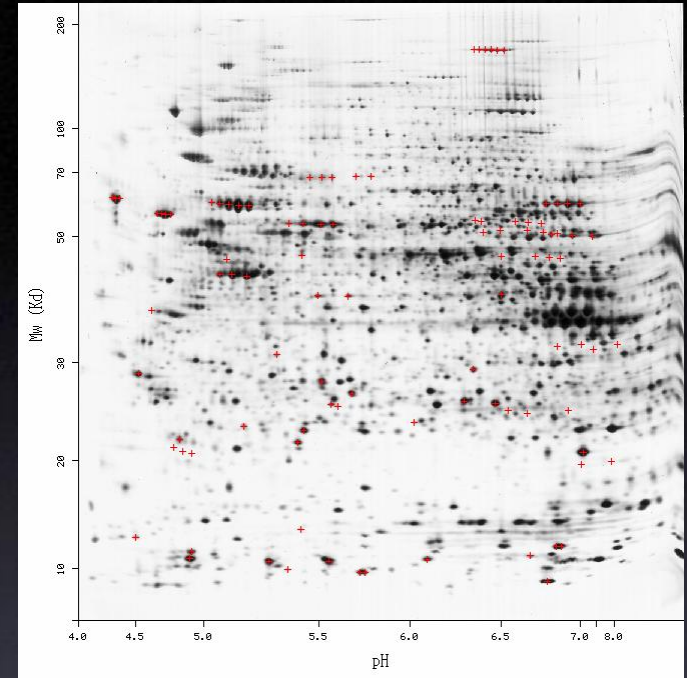




# Ηλεκτροφόρηση δύο διαστάσεων

1. Ισοηλεκτρική εστίαση  
(διαχωρισμός με βάση το ισοηλεκτρικό σημείο)
2. Ηλεκτροφόρηση πηκτώματος SDS - πολυακρυλαμιδίου  
(διαχωρισμός με βάση το μοριακό βάρος)

Υλικό: Μακρομοριακά σύμπλοκα  
Βιοψίες / ιστοί  
Κυτταρικές σειρές

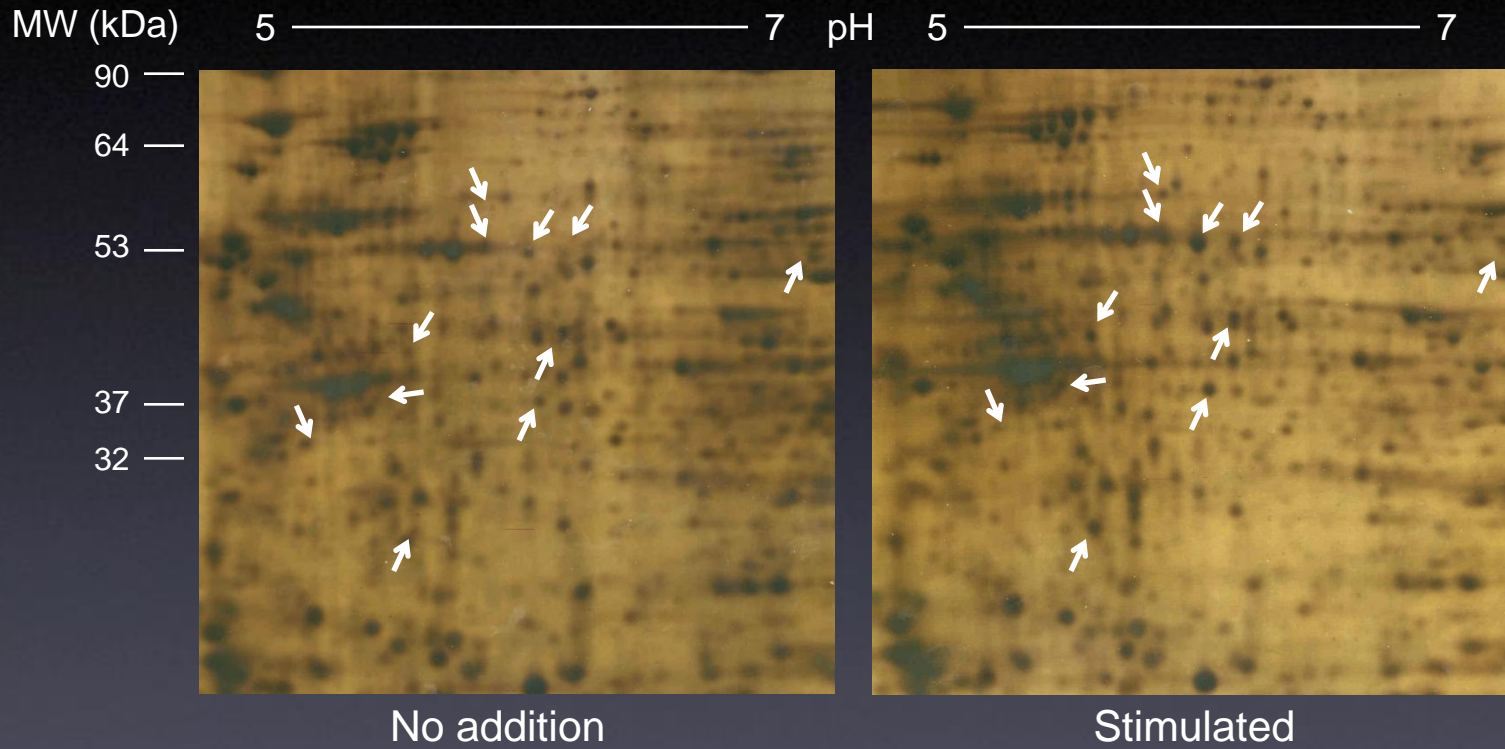


Σήμανση με φθορίζουσες ενώσεις ή χρώση με άργυρο

Ανάλυση των διαφορών στην έκφραση πρωτεϊνών όταν ένα κύτταρο μεταβαίνει από μία κατάσταση σε άλλη:

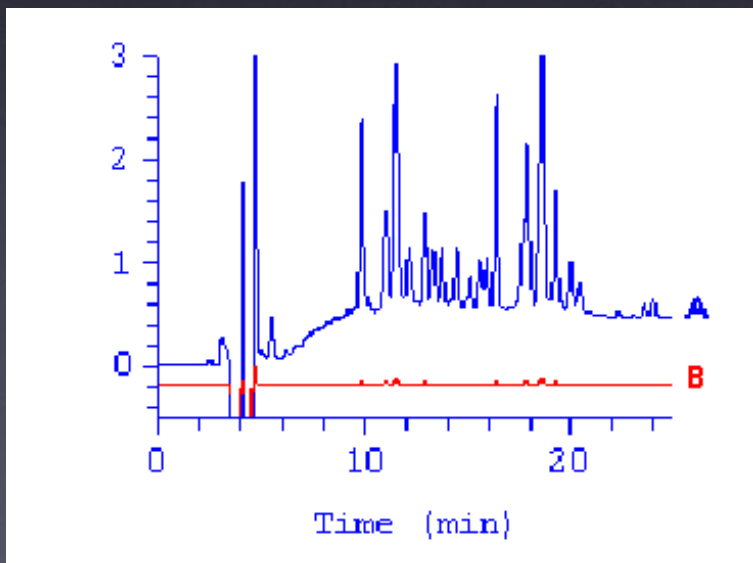
- Διέγερση με κυτοκίνες / αυξητικούς παράγοντες
- Κυτταρική διαφοροποίηση
- Απόπτωση
- Καρκινική μεταβολή

# Comparison of protein expression profiles using 2D-electrophoresis



# Nano - HPLC

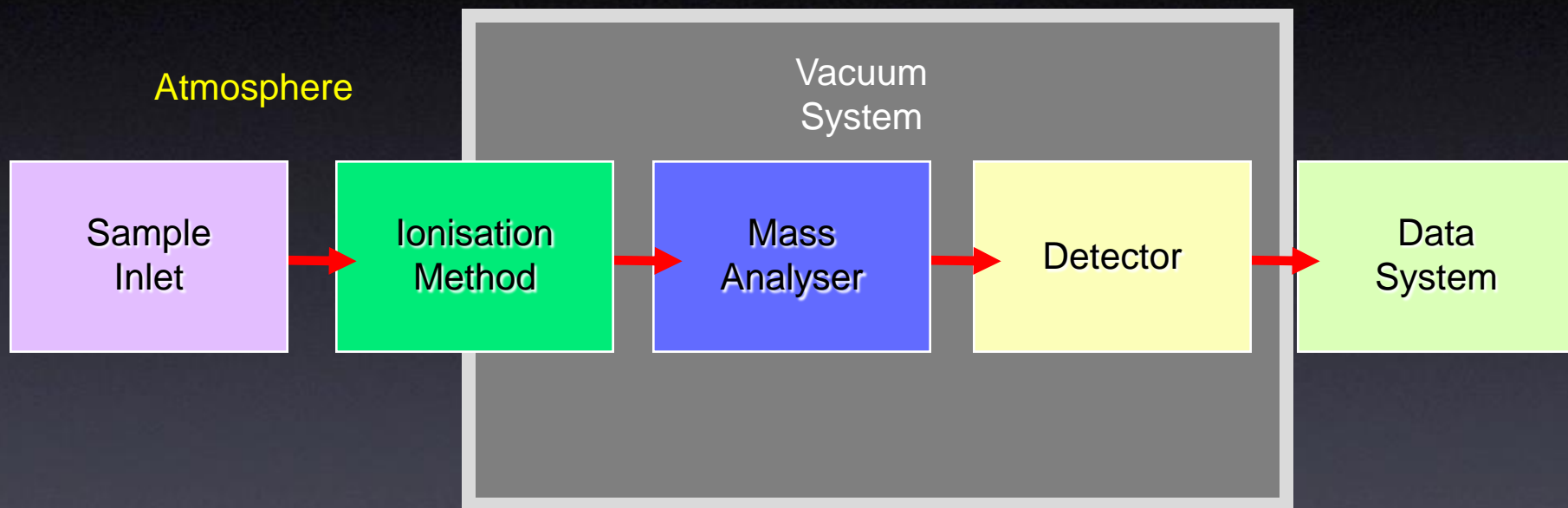
Υγρή Χρωματογραφία Υψηλής Πίεσης  
με τριχοειδείς στήλες (75  $\mu\text{m}$  x 15 cm)  
και ροές κάτω των 200 nl/min



Σύστημα nano-LC με αυτόματο δειγματολήπτη

Μεγάλη βελτίωση στην ευαισθησία  
ανίχνευσης πεπτιδίων σε σύγκριση  
με απλή HPLC

# Components of a Mass Spectrometer





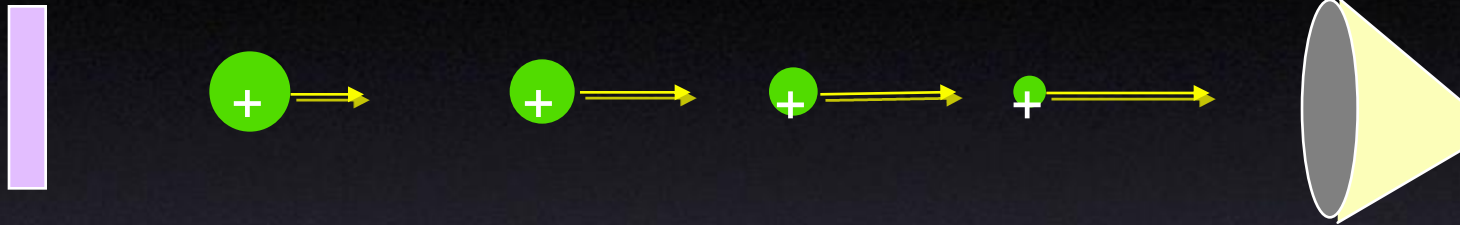
# MASS ANALYZERS

## Types of Mass Spectrometer

- Sector
- Time-of-flight
- Fourier Transform (ICR)
- Quadrupole
- Ion trap

Mass Spectrometers must be able to separate ions by their mass-to-charge ( $m/z$ ) ratios

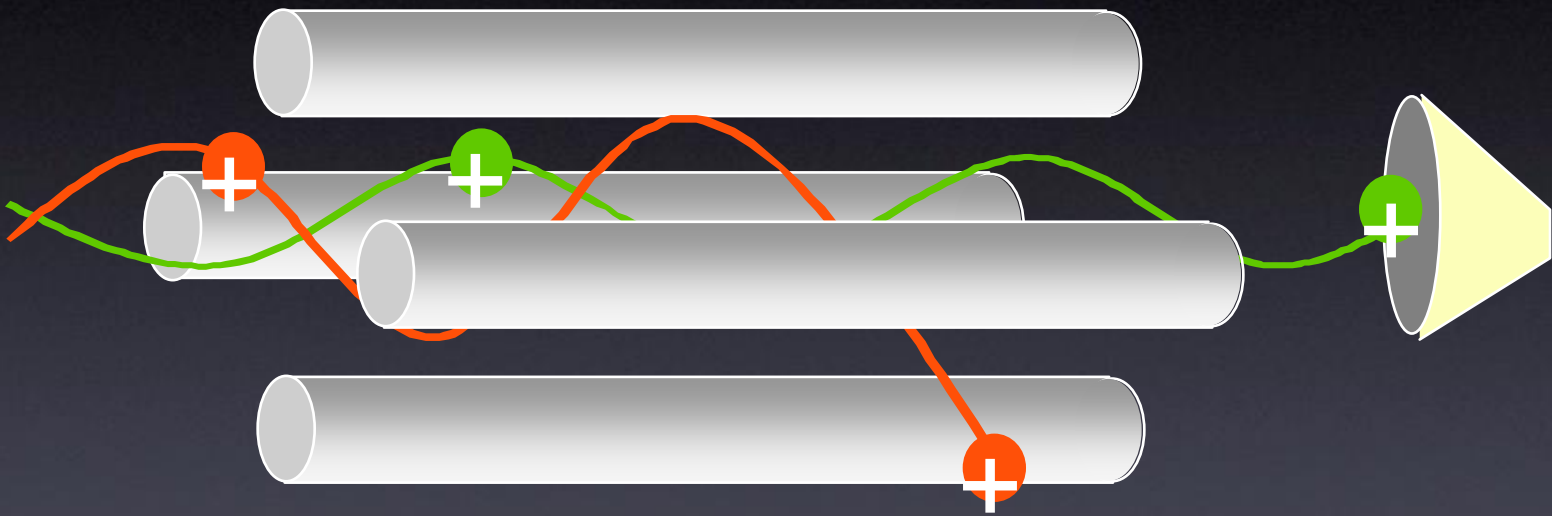
# Time-of-Flight Analyser



- ◉ All ions start together
- ◉ Smaller ions move faster
- ◉ Measure velocity over known distance (time of flight)



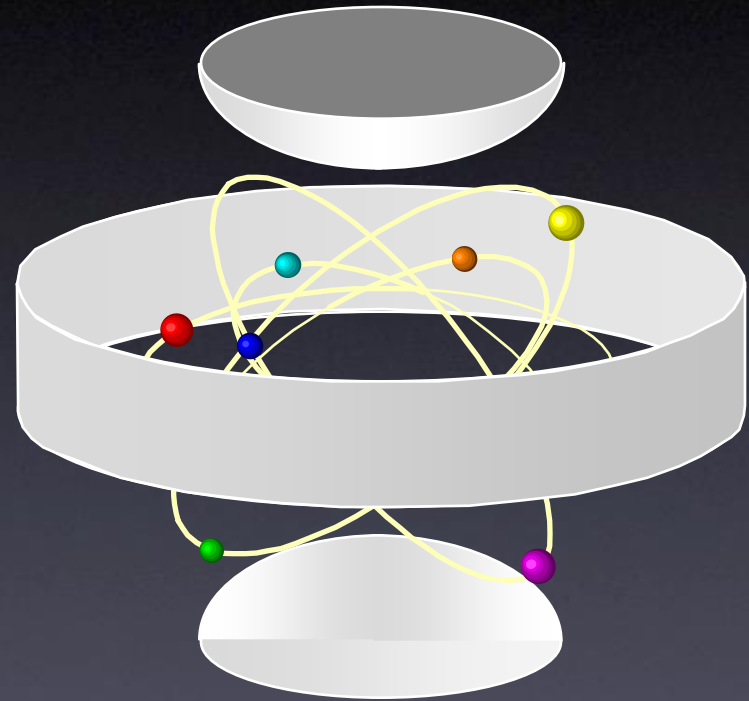
# Quadrupole Mass Analyser



- Ions scanned by varying the DC/Rf voltage across the quadrupoles

# Ion Trap

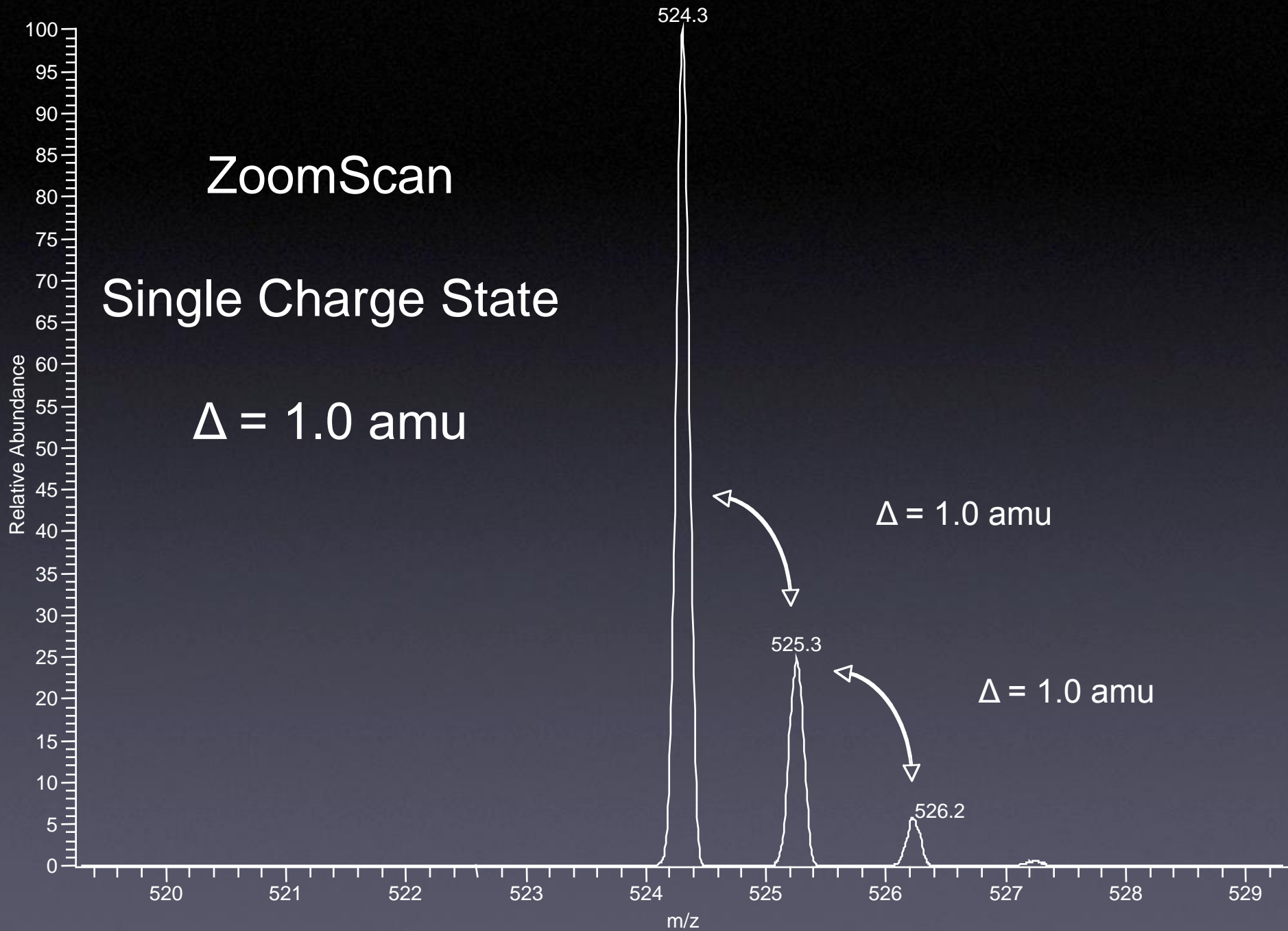
- Principle very similar to quadrupole
- Ions stored by RF & DC fields
- Scanning field can eject ions of specific  $m/z$

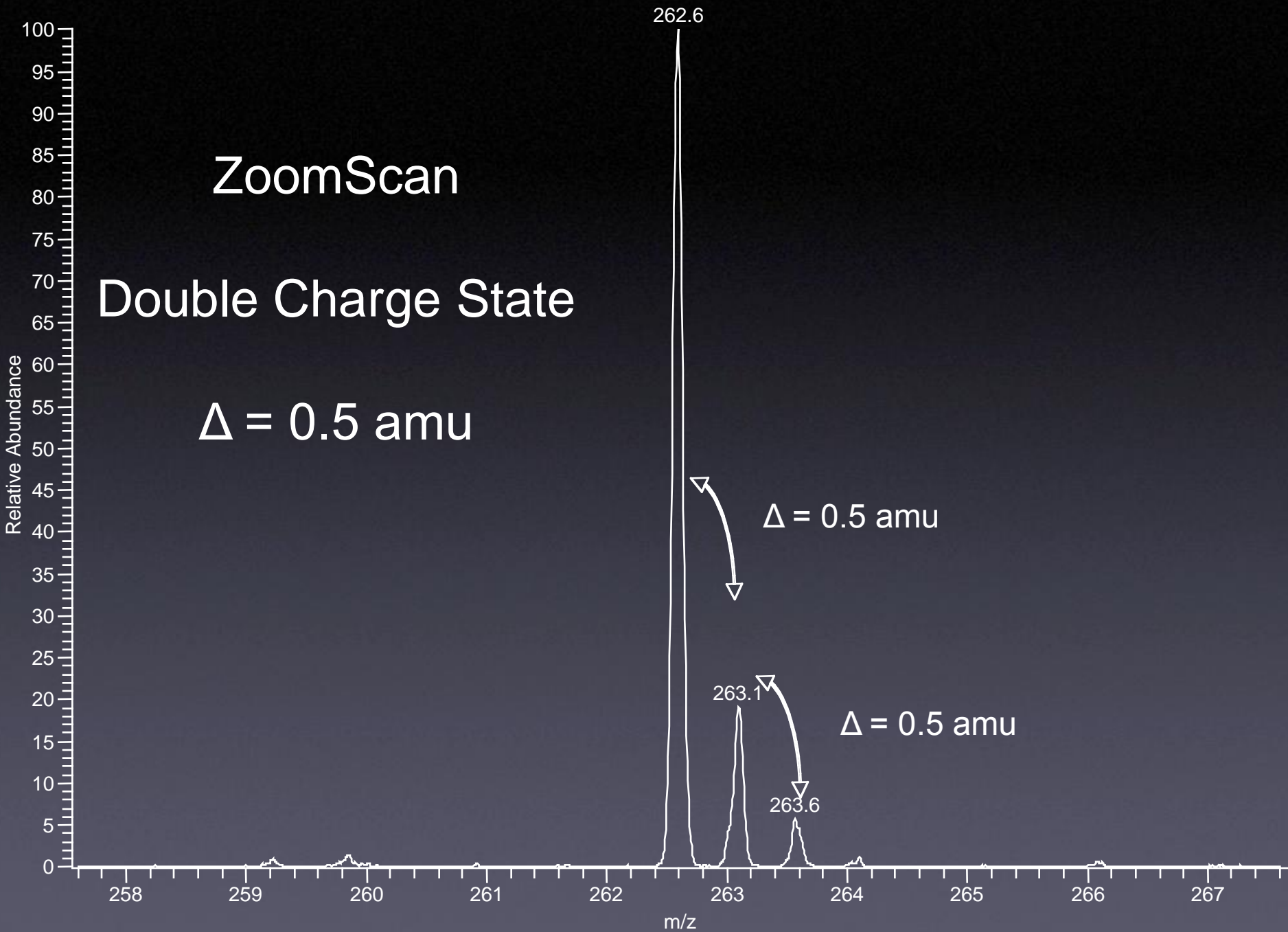




# Multiply Charged Ions

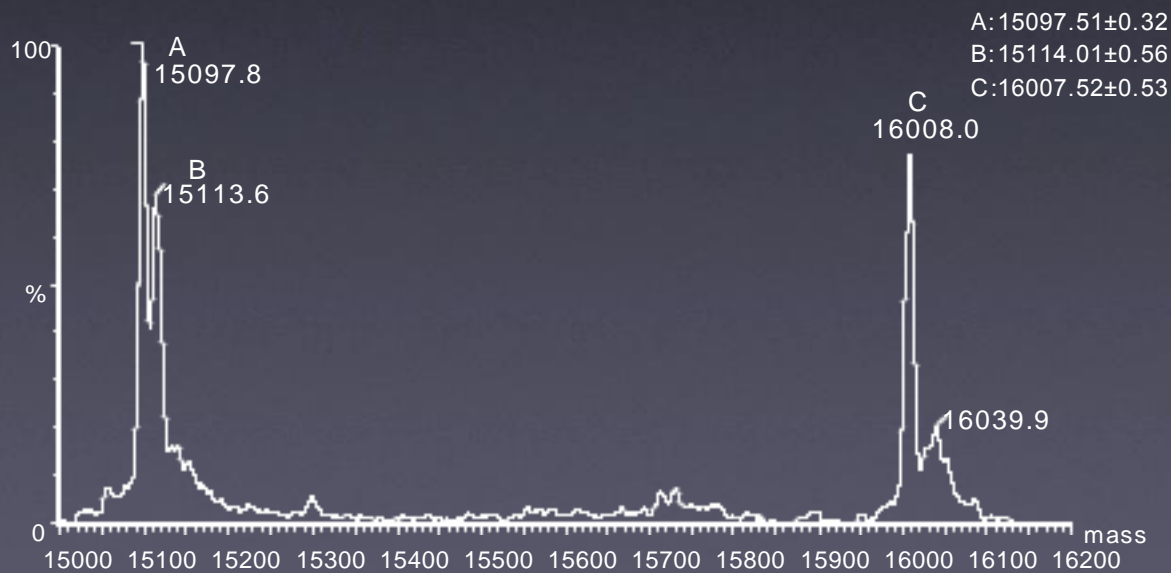
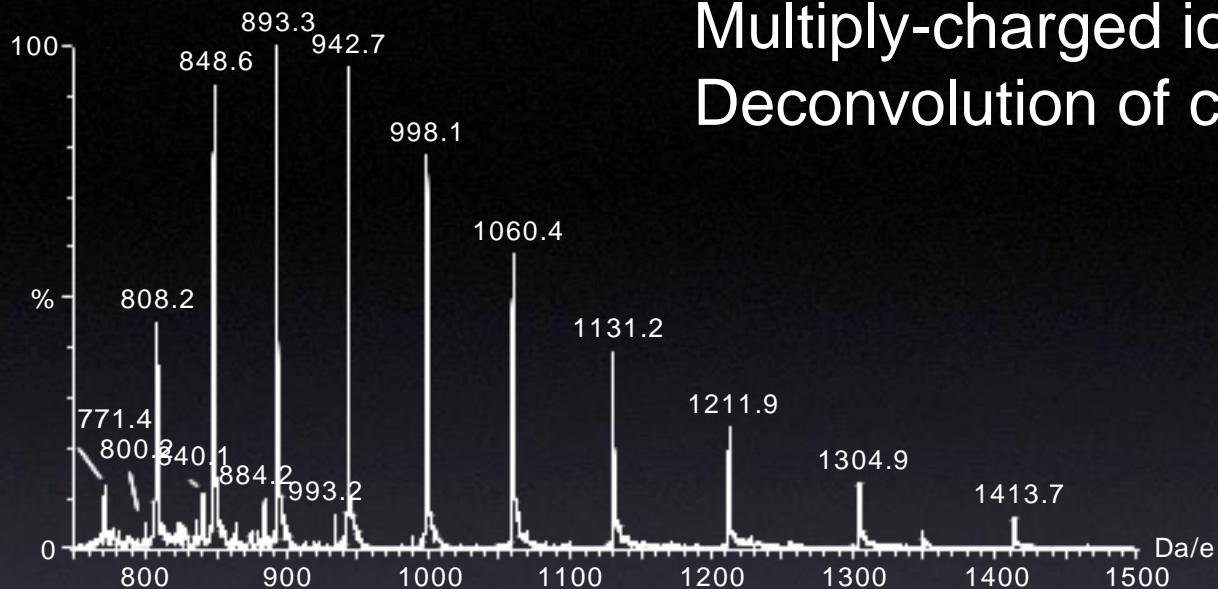
- Mass spectrometers operate on the basis of mass to charge ratio ( $m/z$ )
- Mass assignments are normally made assuming a single charge per ion (i.e.  $m/z = m$ )
  - Single charge                      - apparent mass =  $(M+H)/1$
  - Double charge                      - apparent mass =  $(M+2H)/2$
  - Triple charge                      - apparent mass =  $(M+3H)/3$
  - n charges                      - apparent mass =  $(M+n)/n$





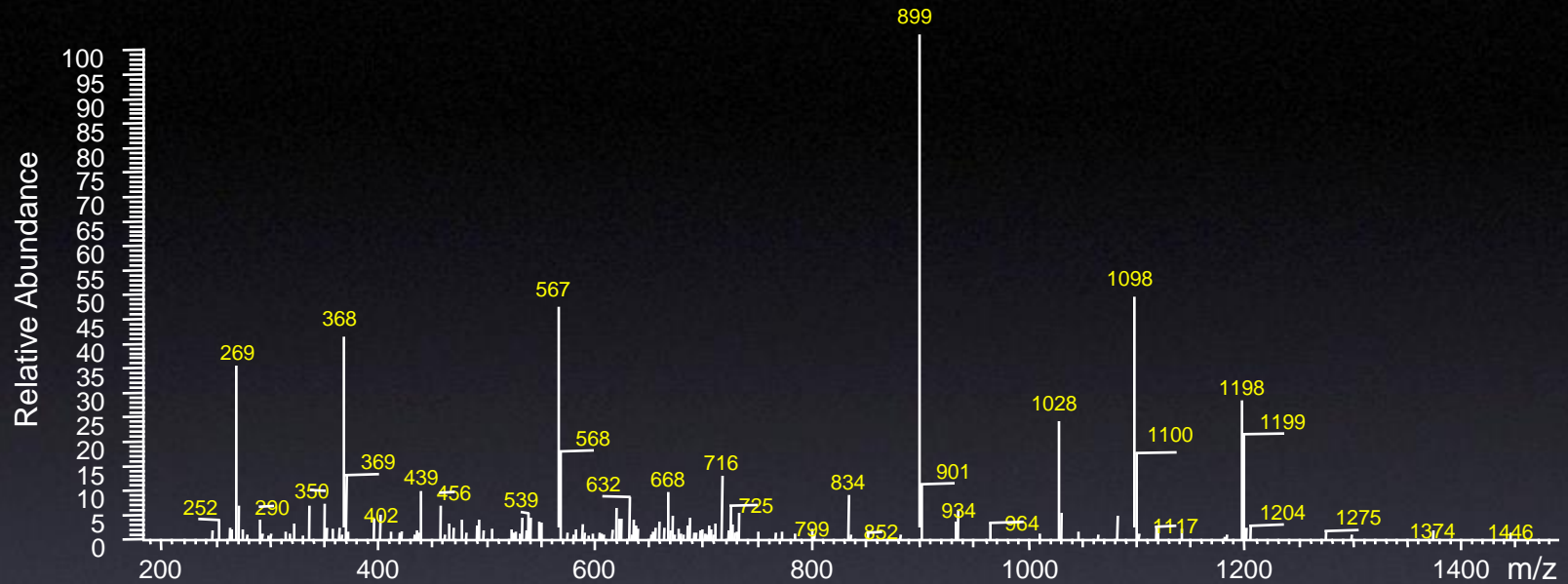
20pmol/ul horse heartmyoglobin

# Multiply-charged ions: Deconvolution of complex spectra



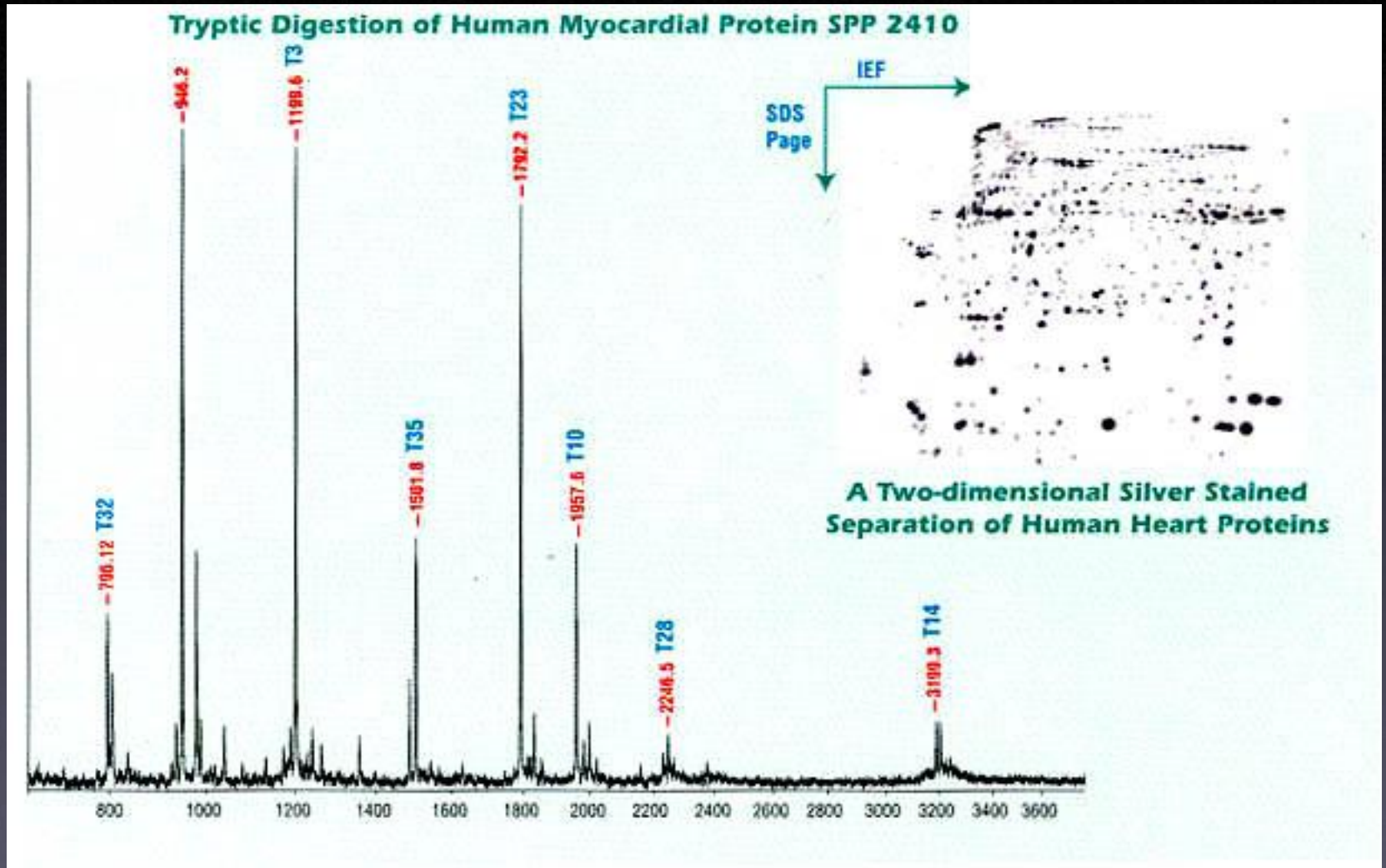


# Ανάλυση πρωτεϊνών με φασματογραφία μάζας

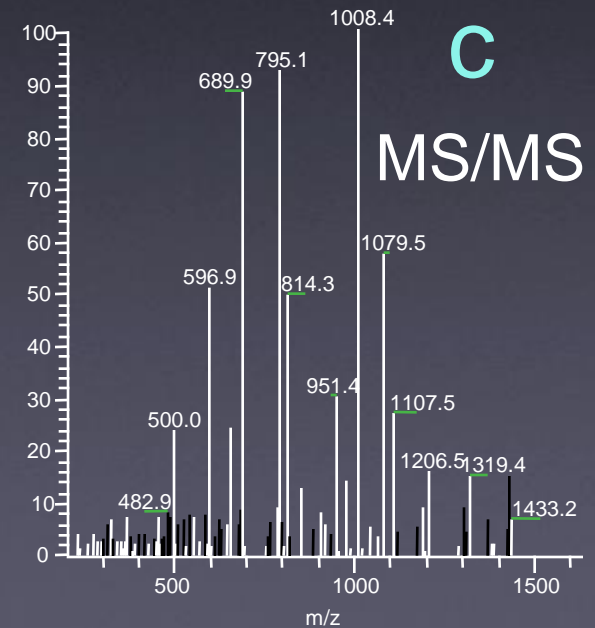
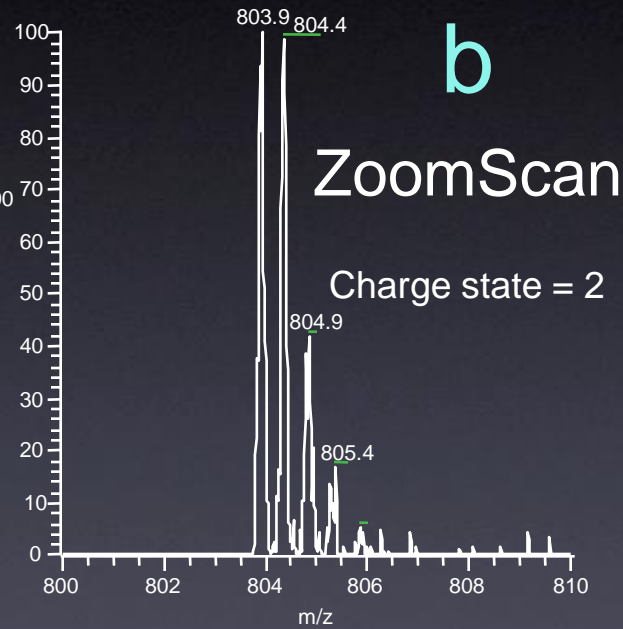
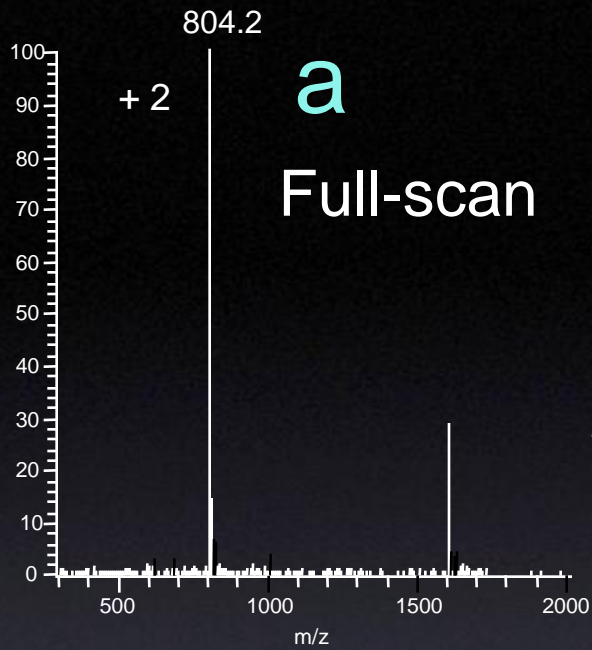


- Τα μοριακά βάρη των πεπτιδίων που προέρχονται από την ενζυμική διάσπαση μιας πρωτεΐνης αποτελούν το «δακτυλικό αποτύπωμα» της.
- Κάθε μία από τις κορυφές αυτές μπορεί να αναλυθεί περαιτέρω με διαδοχικό MS/MS για την ανάγνωση της αλληλουχίας των αμινοξέων.
- Η ταυτοποίηση μιας πρωτεΐνης γίνεται αναλύοντας τα στοιχεία αυτά σε τράπεζες δεδομένων.

# Συνδιασμός ηλεκτροφόρησης δύο διαστάσεων και MALDI-TOF MS για την ταυτοποίηση πρωτεϊνών



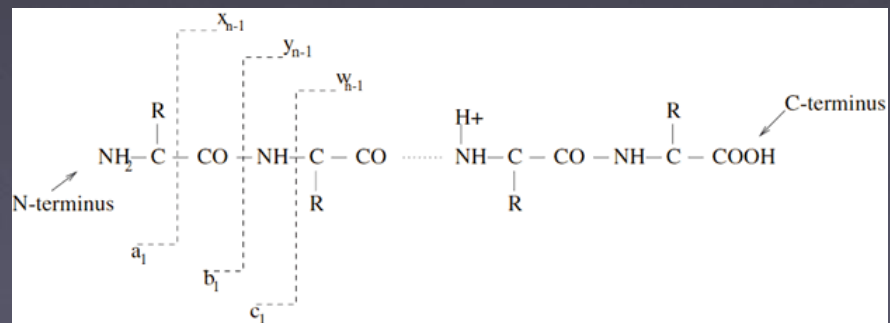
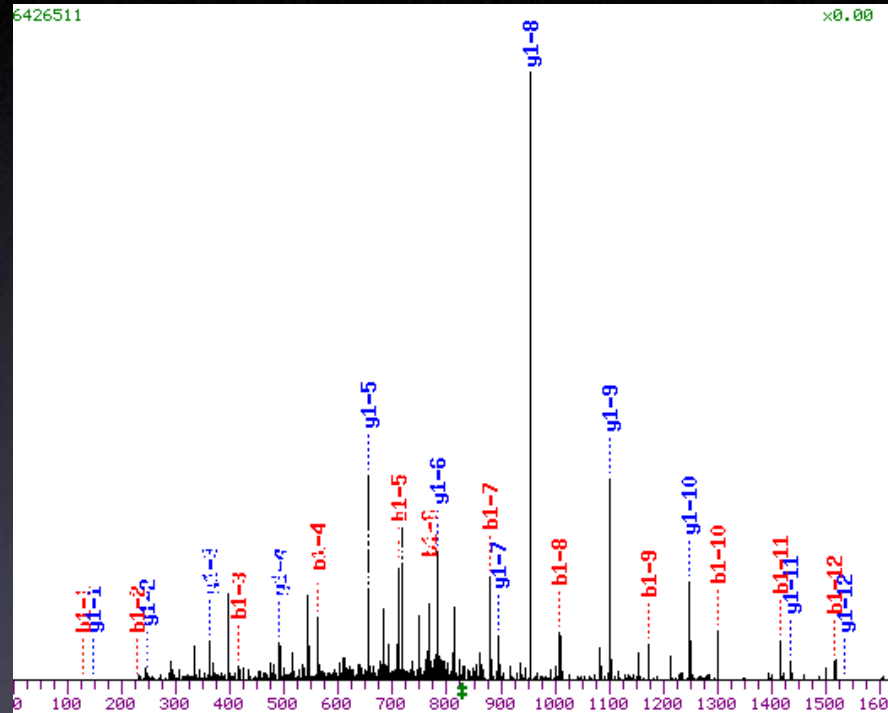
# “Triple-play” analysis of peptides



# Ταυτοποίηση πρωτεϊνών με την διαδικασία MS-MS

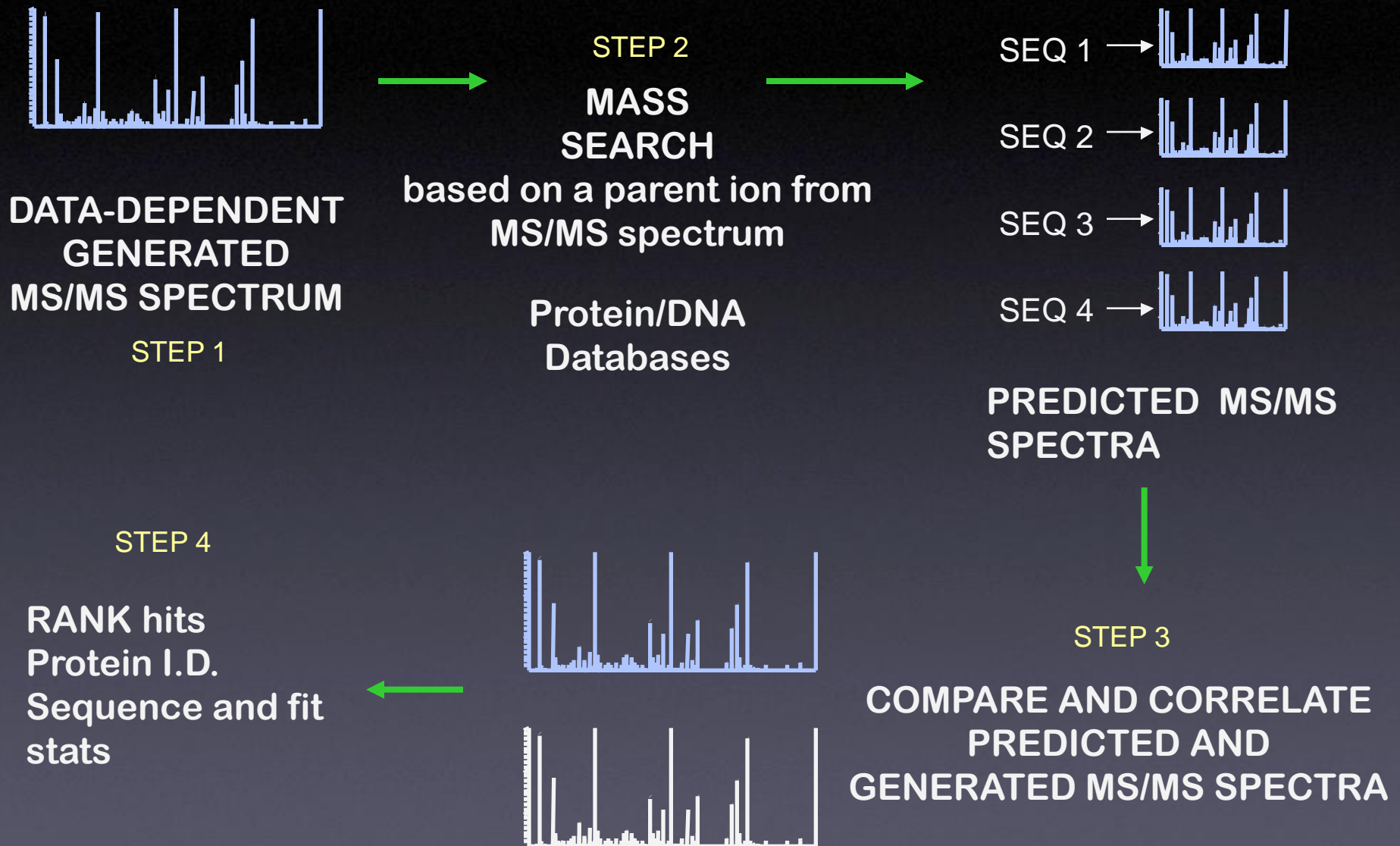
Ιόντα b και y που παράγονται από την διάσπαση πεπτιδίου

Seq #	b	y	#
E 1	130.1	-	13
V 2	229.2	1532.7	12
W 3	415.5	1433.6	11
F 4	562.6	1247.4	10
F 5	709.8	1100.2	9
G 6	766.9	953.0	8
L 7	880.0	896.0	7
Q 8	1008.2	782.8	6
Y 9	1171.3	654.7	5
Q 10	1299.5	491.5	4
D 11	1414.6	363.4	3
T 12	1515.7	248.3	2
K 13	-	147.2	1





# Searching databases with MS/MS data



# Ταυτοποίηση πρωτεΐνης σε βάση δεδομένων

database=D:\DataBase/sprot.fasta, accession=AOP2\_MOUSE

peptide(s)=VVDSLQLTGTK KGESVMVPTLSEEEAK VVFIFGPDKK VVDSLQLTGTKPVATPVDWK LPFPIIDDK  
PGGLLLGDEAPNFEANTTIGR LIALSIDSVEDHLAWSK FHDFLGDSWGILFSHPR DLAILLGMLDPVEK

Analyzing ...

>AOP2\_MOUSE (O08709) ANTIOXIDANT PROTEIN 2 (1-CYS PEROXIREDOXIN)

PGGLLLGDEA PNFEANTTIG RIRFHDFLGD SWGILFSHPR DFTPVCTTEL GRAAKLAPEF AKRNVKLIAL SIDSVEDHLA  
WSKDINAYNG ETPTEKLPFP IIDDKGRDLA ILLGMLDPVE KDDNNMPVTA RVVFIFGPDK KCLKLSILYPA TTGRNFDEIL  
RVVDSLQLTG TKPVATPVDW KKGESVMVVP TLSEEEAKQC FPKGVTKEK PSGKKYLRYT PQP

>average mass = 24721

position sequence (NCBI BLAST link)  
-----

162- 172 VVDSLQLTGTK  
182- 198 KGESVMVPTLSEEEAK  
132- 141 VVFIFGPDKK  
162- 181 VVDSLQLTGTKPVATPVDWK  
97- 105 LPFPIIDDK  
1- 21 PGGLLLGDEAPNFEANTTIGR  
67- 83 LIALSIDSVEDHLAWSK  
24- 40 FHDFLGDSWGILFSHPR  
108- 121 DLAILLGMLDPVEK

Protein Coverage: 125/223 = 56.1% by amino acid count, 13647/24721 = 55.2% by mass

Search SWISS-PROT with AOP2\_MOUSE via accession, descr./ID, or full text field.

# Proteins differentially expressed in Ras<sup>G12V</sup> transformed Caco-2 cells

## UPREGULATED

### 1. GLYCOMETABOLISM

Transketolase (P29401)

Phosphoglycerate mutase 1 (P18669)

Triosephosphate isomerase (P60174)

Malate dehydrogenase, cytoplasmic (P40925)

### 2. BINDING PROTEINS

Cellular retinoic acid-binding protein 2 (P29373)

Annexin A8 (P13928)

Fatty acid binding protein 3 (P05413)

Nucleosome assembly protein 1-like 1 (P55209)

Ran GTP-binding nuclear protein (P62826)

Similar to annexin A2, isoform 1

Poly-(rC)-binding protein 1 (Q15365)

Tumor protein, translationally controlled 1 (P13693)

### 3. OXIDOREDUCTASE ACTIVITY

Dimethylaniline monooxygenase 5 (N-oxide forming) (P49326)

### 4. CYTOSKELETAL

Cofilin 1 (P23258)

### 5. BIOTRANSFORMATION

Peptidyl-prolyl cis-trans isomerase A (P62937)

Phenylethanolamine N-methyltransferase (P11086)

### 6. LIPID METABOLISM

Acetyl-CoA acetyltransferase, cytosolic (Q9BWD1)

### 7. AMINOACID METABOLISM

Fumarylacetoacetase (P16930)

Glutathione synthase (P48637)

### 8. TRANSCRIPTION AND TRANSLATION

Nucleophosmin (P09748)

Nucleoside diphosphate kinase B (P22392)

Homeobox protein Hox-C12 (P31275)

60S acidic ribosomal protein P0 (P05388)

Tyrosyl-tRNA synthase, cytoplasmic (P54577)

40S ribosomal protein S12 (P25398)

### 9. PROTEOLYSIS

Ubiquitin conjugated enzyme E2N (P61088)

26S proteasome non-ATPase regulatory subunit 7 (P51665)

Splice isoform 1 of proteasome subunit alpha type 7 (O14818)

Proteasome subunit alpha type 2 (P25787)

Cytosolic non-specific dipeptidase (Q96KP4)

Calpain small subunit 1 (P04632)

### 10. CHAPERONS

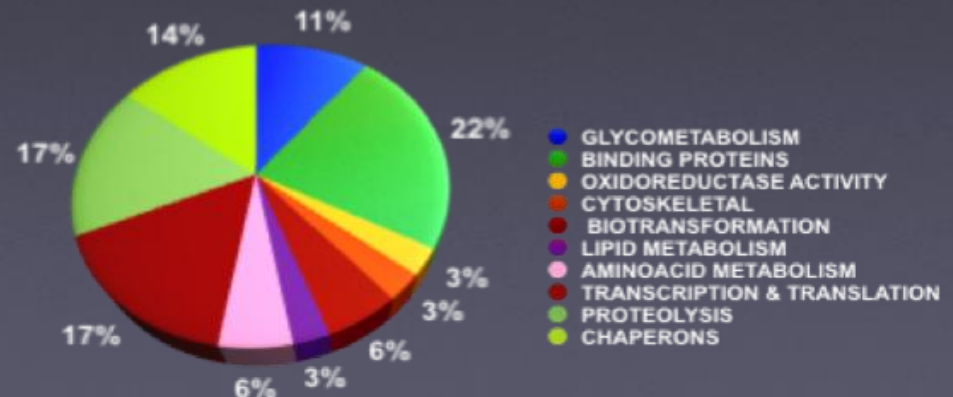
Heat shock protein beta 1 (P04792)

Nascent polypeptide-associated complex alpha subunit (Q13765)

T-complex protein 1, delta subunit (P50991)

Stress-induced phosphoprotein 1 (P31948)

Isoform 1 of proteasome activator complex subunit 3 (P61289-1)





# Ταυτοποίηση μετα-μεταφραστικών τροποποιήσεων πρωτεϊνών με Φασματογραφία Μάζας

Τροποποίηση	Διαφορά μάζας
Methylation	14.0157
Hydroxylation	15.9949
Oxidation of Met	15.9949
Formylation	27.9949
Acetylation	42.0106
Carboxylation of Asp and Glu	43.9898
Phosphorylation	79.9663
Sulphation	79.9568
Cysteinylation	119.0041
Pentoses (Ara, Rib, Xyl)	132.0423
Deoxyhexoses (Fuc, Rha)	146.0579
Hexosamines (GalN, GlcN)	161.0688
Hexoses (Fru, Gal, Glc, Man)	162.0528
Lipoic acid (amide bond to lysine)	188.0330
N-acetylhexosamines (GalNAc, GlcNAc)	203.0794
Farnesylation	204.1878
Myristoylation	210.1984
Biotinylation (amide bond to lysine)	226.0776
Pyridoxal phosphate (Schiff Base formed to lysine)	231.0297
Palmitoylation	238.2297
Stearoylation	266.2610
Geranylation	272.2504
N-acetylneuraminic acid	291.0954
Glutathionylation	305.0682
N-glycolylneuraminic acid (NeuGc)	307.0903
ADP-ribosylation (from NAD)	541.0611

Σημαντικός ρόλος στη λειτουργία των πρωτεϊνών και την κυτταρική ρύθμιση

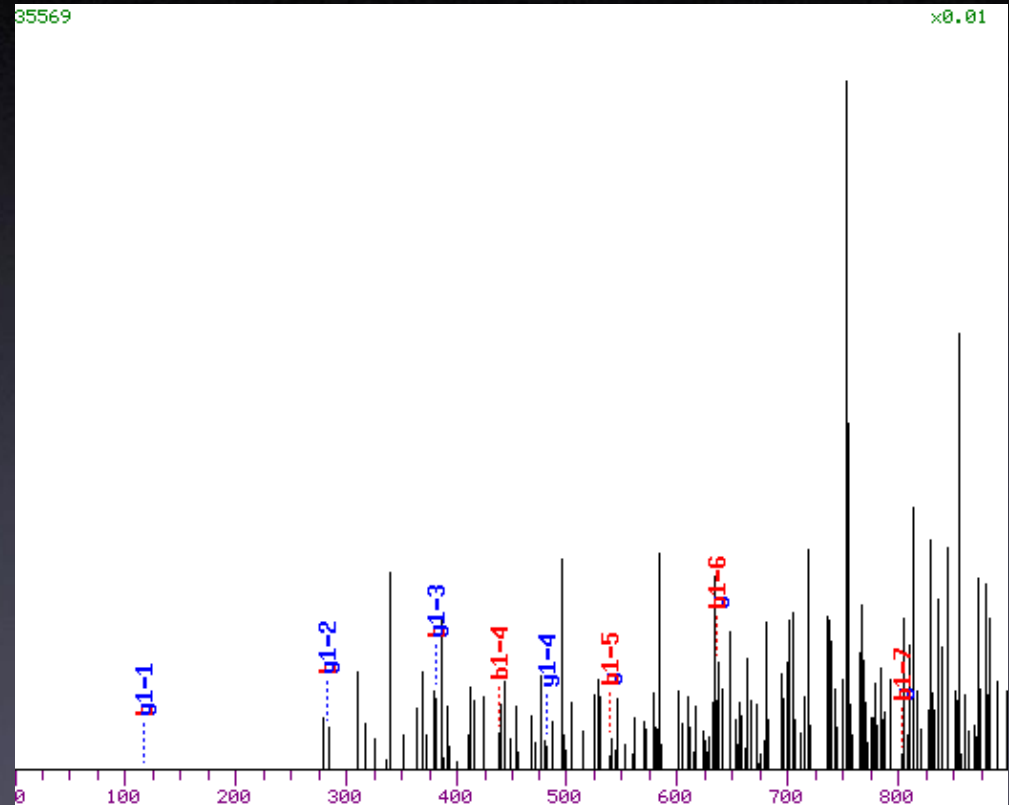


# Identification of phosphorylation sites on the M3/6 phosphatase by MS

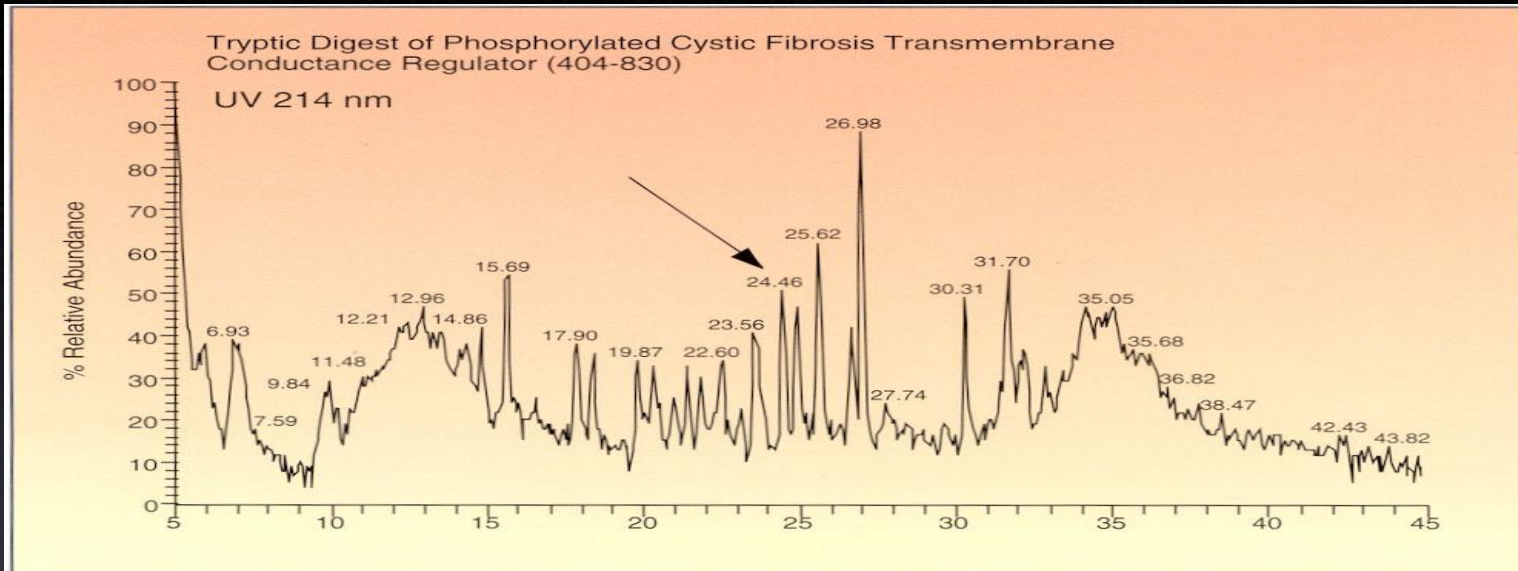
## Ser 515 and Ser 520

datafile=02100204.1245.1245.1.dta  
peptide=DS**PGTPSP**, precursor mass=915.5

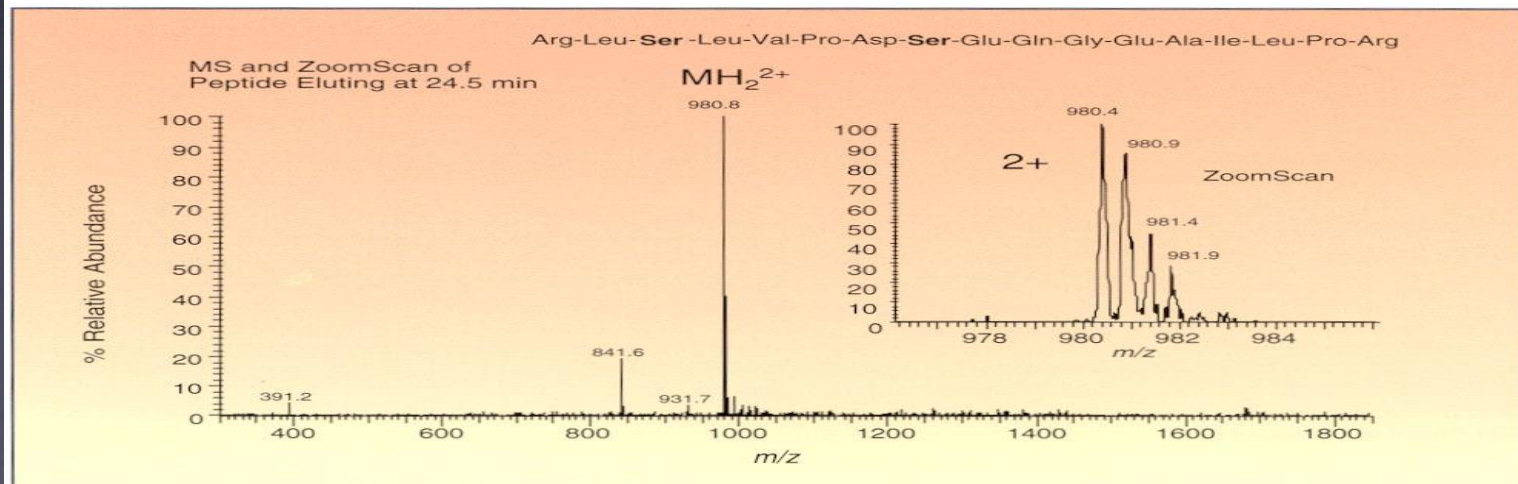
Seq	#	b	y	(+1)
---	--	-----	-----	--
D	1	116.1	-	8
<b>*S</b>	2	<b>283.2</b>	<b>802.7</b>	7
P	3	<b>380.3</b>	<b>635.6</b>	6
G	4	<b>437.3</b>	<b>538.5</b>	5
T	5	<b>538.4</b>	<b>481.4</b>	4
P	6	<b>635.6</b>	<b>380.3</b>	3
<b>*S</b>	7	<b>802.6</b>	<b>283.2</b>	2
P	8	-	116.1	1



# Ανίχνευση των θέσεων φωσφορυλίωσης πρωτεϊνών με χρήση ΦΜ 1

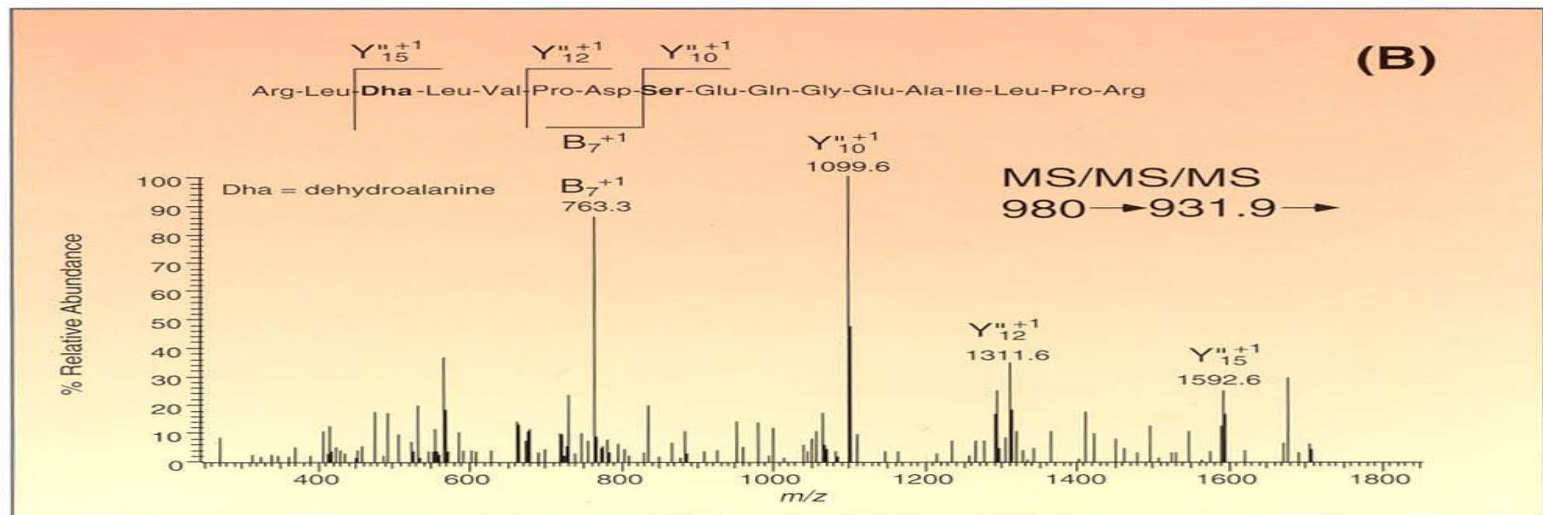
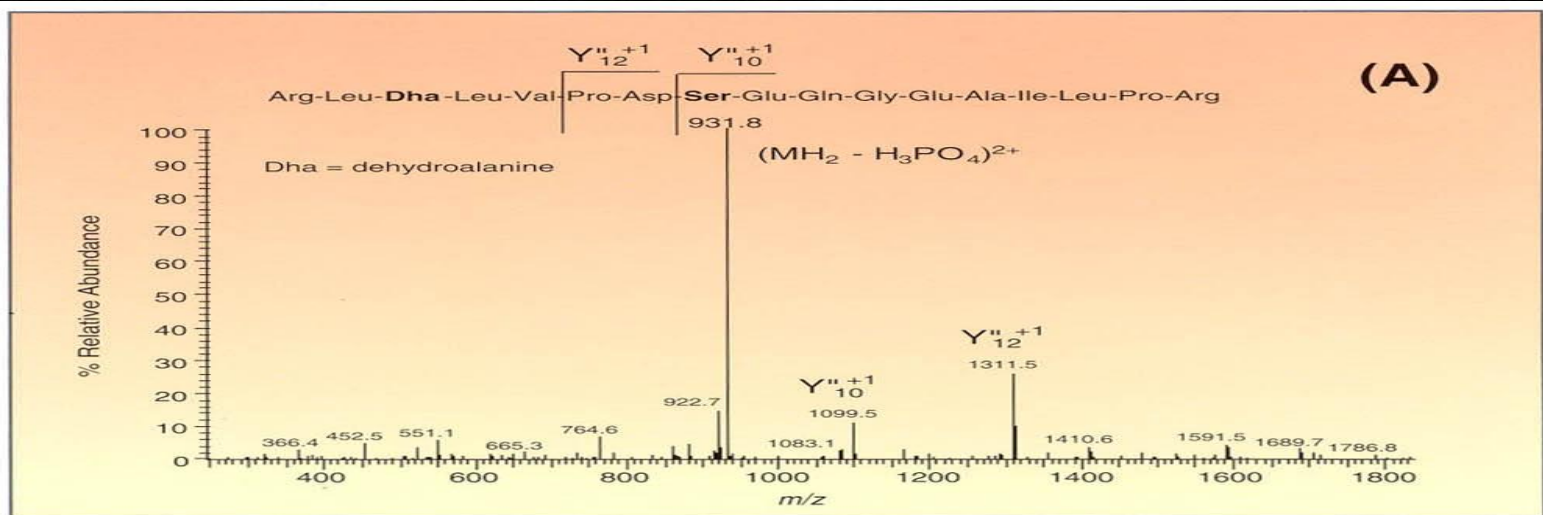


*UV chromatogram of tryptic digest of phosphorylated CFTR.*



*Full scan MS and ZoomScan (inset) of a phosphorylated peptide from CFTR. Either of the two serines could be phosphorylated.*

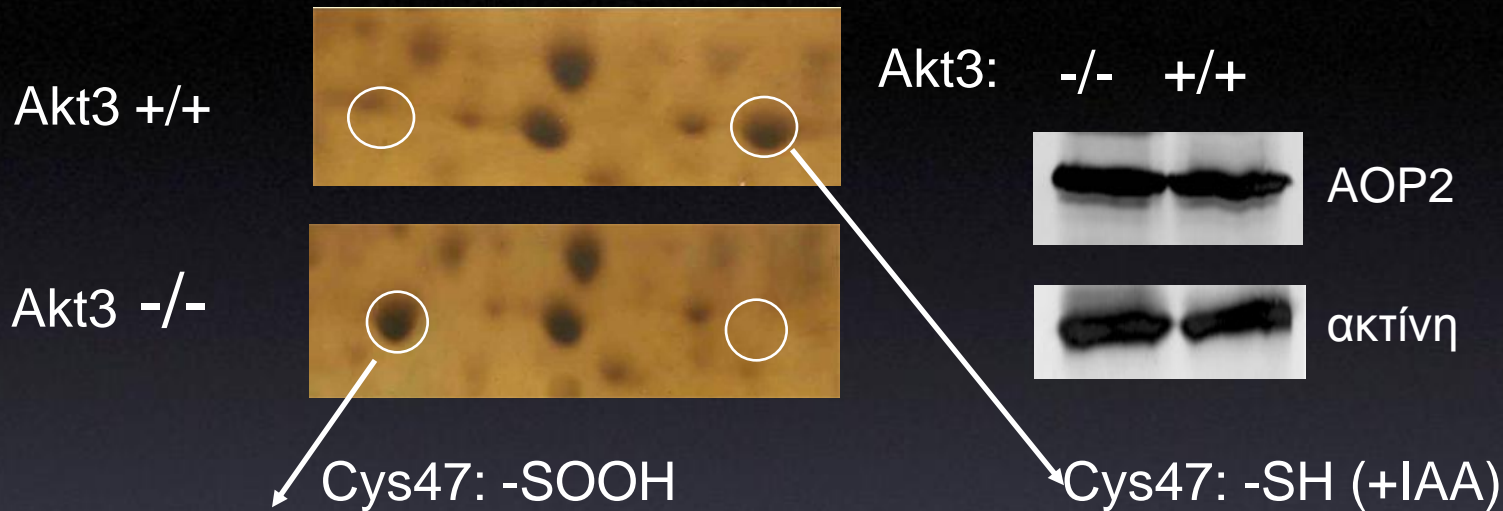
# Ανίχνευση των θέσεων φωσφορυλίωσης πρωτεϊνών με χρήση ΦΜ 2



The MS/MS spectrum of the phosphorylated serine has an ion at  $m/z$  931.8 which corresponds to loss of H<sub>3</sub>PO<sub>4</sub> from the parent ion. Presence of Y<sub>10</sub><sup>n+1</sup> and Y<sub>12</sub><sup>n+1</sup> in the MS/MS spectrum and Y<sub>10</sub><sup>n+1</sup>, Y<sub>12</sub><sup>n+1</sup>, Y<sub>15</sub><sup>n+1</sup>, and B<sub>7</sub><sup>+1</sup> in the MS/MS/MS spectrum confirms the phosphorylation site to be the serine at the third position and not the serine at position eight.



# Τροποποίηση της AOP2 σε δείγματα Akt3<sup>-/-</sup>

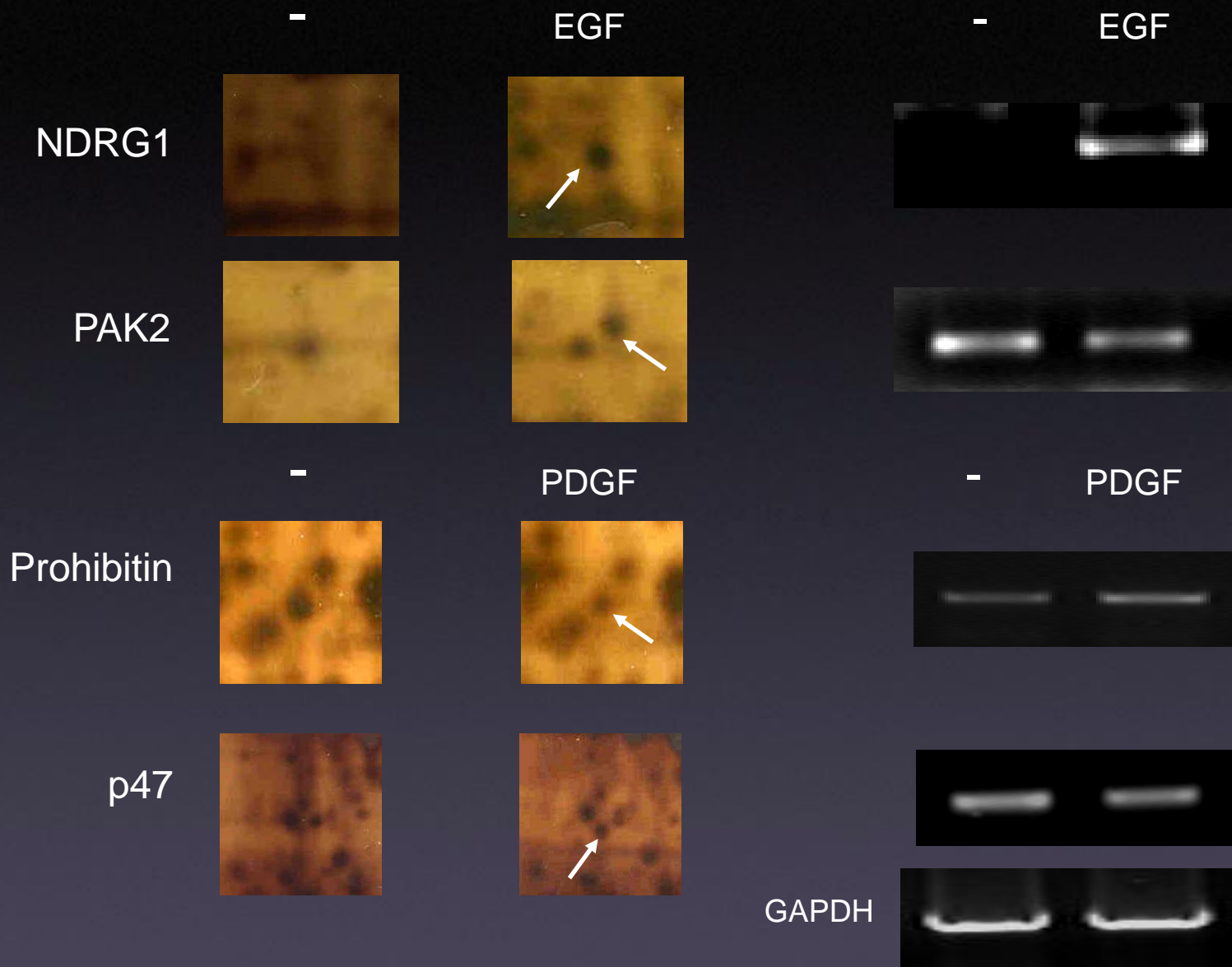


Seq #	b	y	(+1)	
D 1	116.1		-	12
F 2	263.3		1256.4	11
T 3	364.4		1109.2	10
P 4	461.5		1008.1	9
V 5		560.6	911.0	8
*C 6		695.8	811.9	7
T 7	796.9		676.7	6
T 8	898.0		575.6	5
E 9		1027.1	474.5	4
L 10	1140.2	345.4	3	
G 11		1197.3	232.2	2
R 12		-	175.2	1

Seq #	b	y	(+1)	
D 1	116.1		-	12
F 2	263.3		1281.4	11
T 3	364.4		1134.3	10
P 4	461.5		1033.2	9
V 5		560.6	936.0	8
C 6		720.8	836.9	7
T 7	821.9		676.7	6
T 8	923.0		575.6	5
E 9		1052.1	474.5	4
L 10	1165.3	345.4	3	
G 11		1222.3	232.2	2
R 12		-	175.2	1

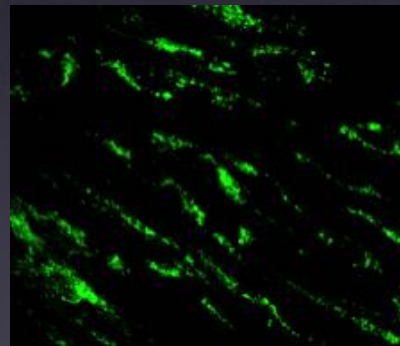
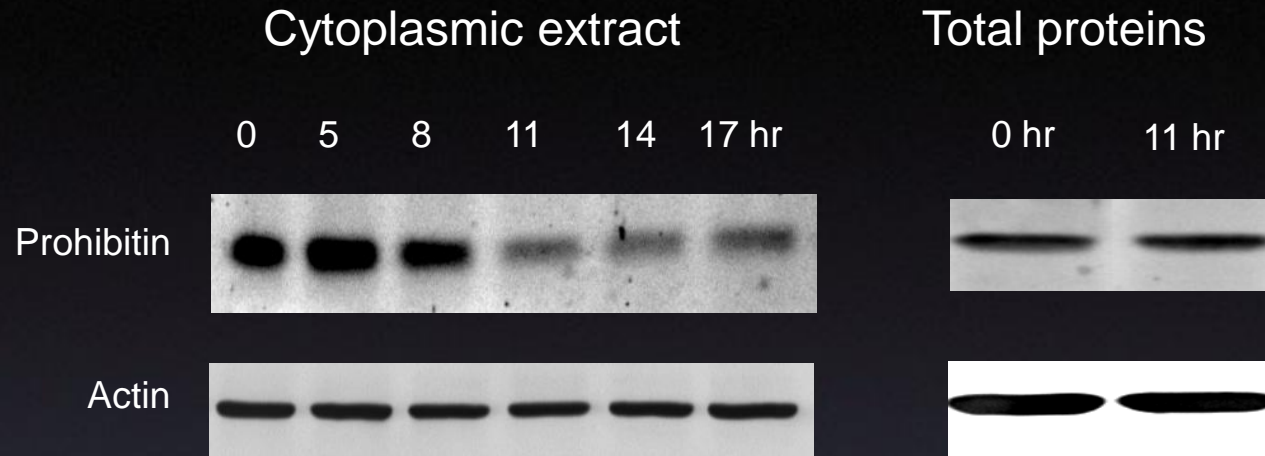
## 2D-electrophoresis of cytoplasmic extracts

## RT-PCR

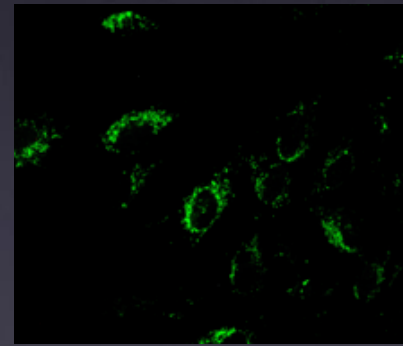




# PDGF treatment results in Prohibitin depletion from cytoplasmic extracts



- PDGF



+ PDGF

# Beyond 2D Electrophoresis....

## Isotopic labeling of peptides

Sample 1

Lysate

Trypsinisation in H<sub>2</sub>O

Sample 2

Lysate

Trypsinisation in H<sub>2</sub>O<sup>18</sup>

Combine

Multiple chromatography

Mass spectrometry



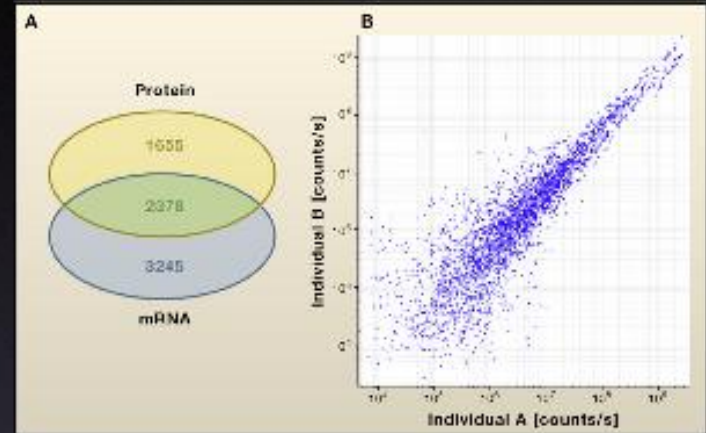
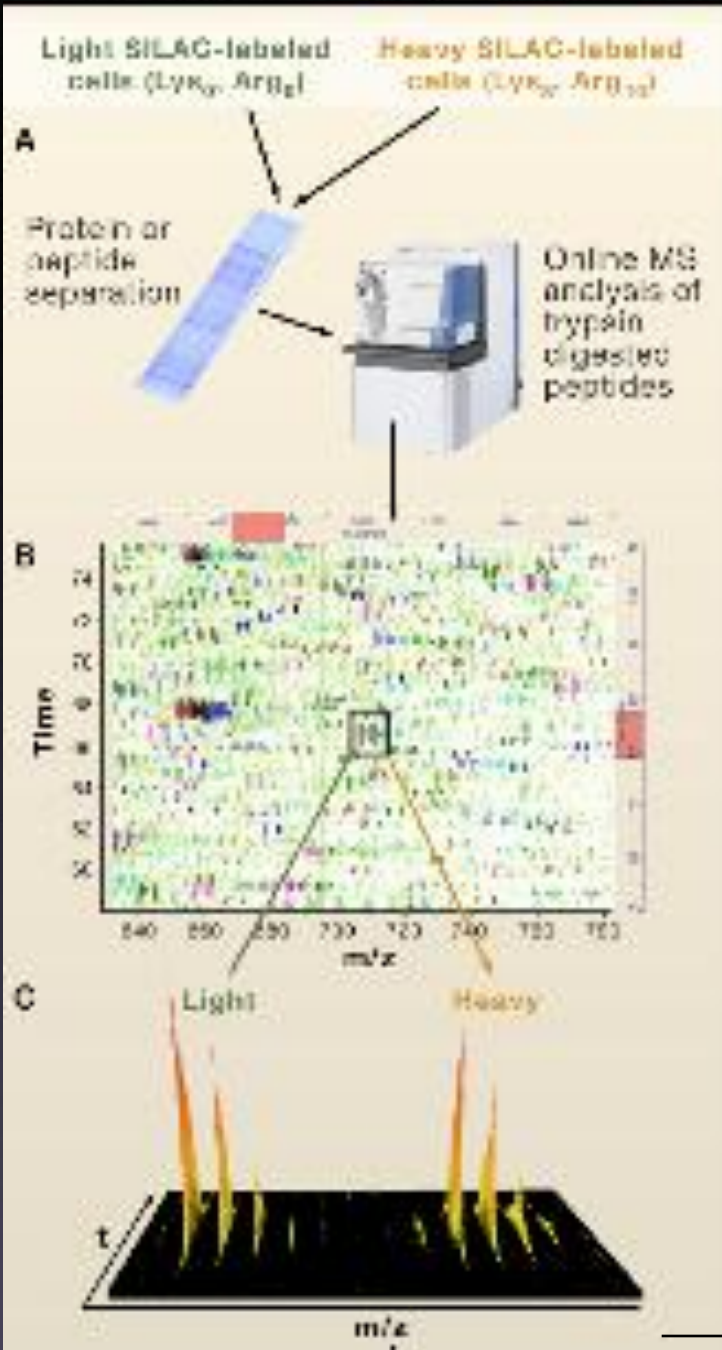
No change



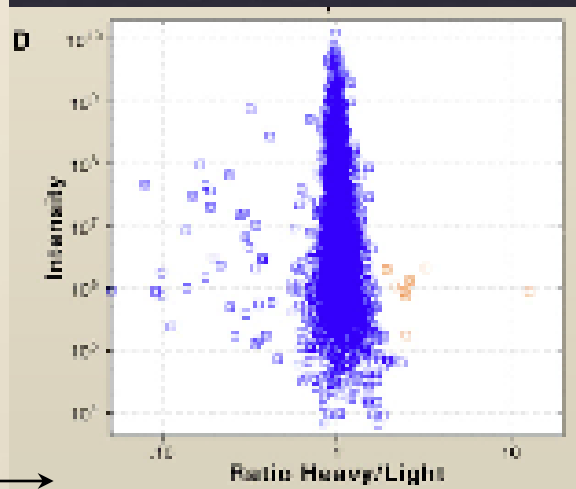
Quantitative change

# Ισοτοπική σήμανση με αμινοξέα (SILAC)

Ανίχνευση και ποσοτικοποίηση μεγάλου αριθμού πρωτεϊνών



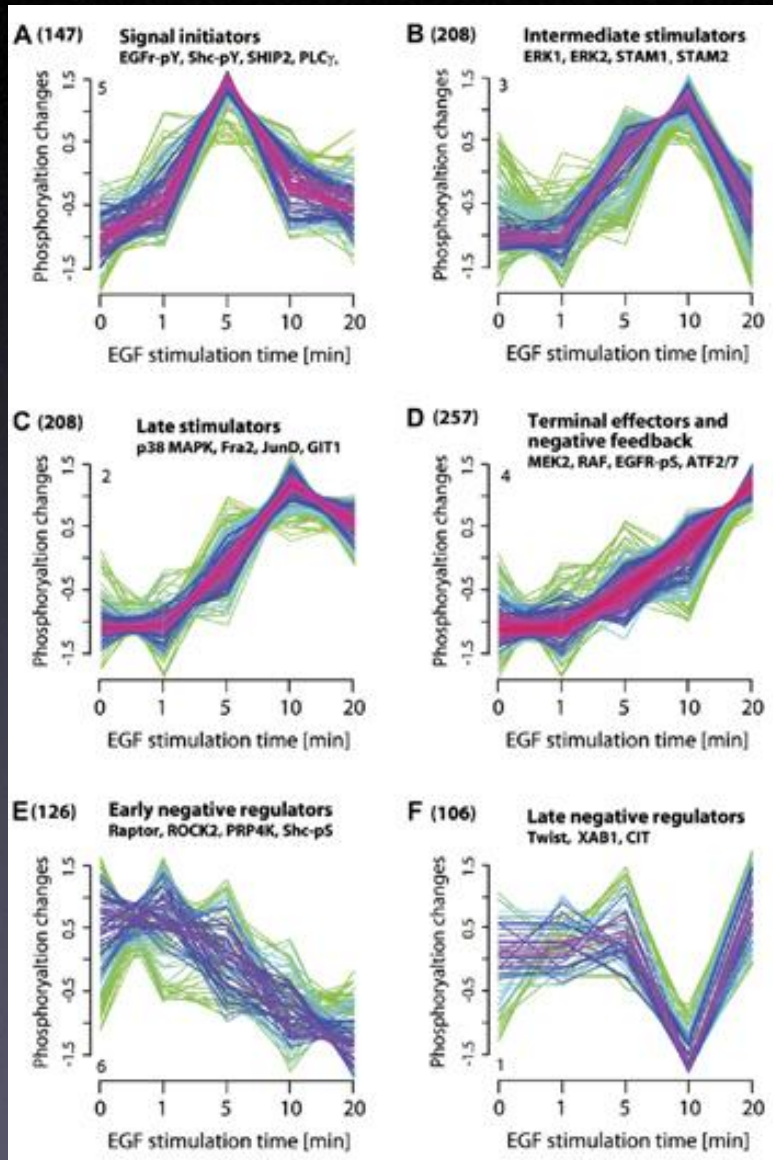
Σύγκριση με μικροσυστοιχείες Affymetrix



Is Proteomics the New Genomics?  
J. Cox and M. Mann<sup>1</sup>  
Cell, 130, 395 (2007)

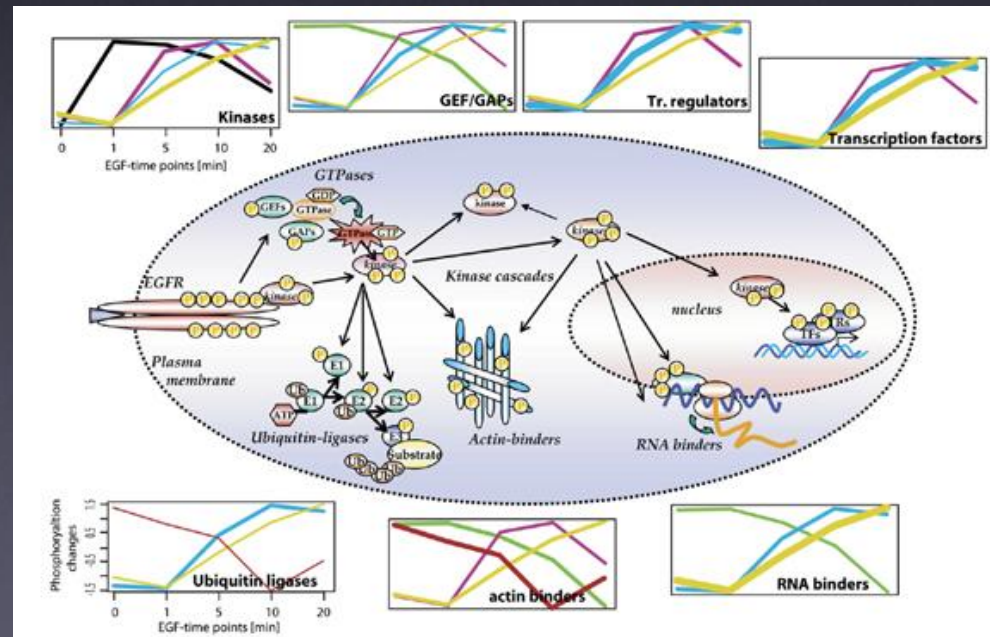


# Temporal proteomics for the analysis of global phosphorylation events



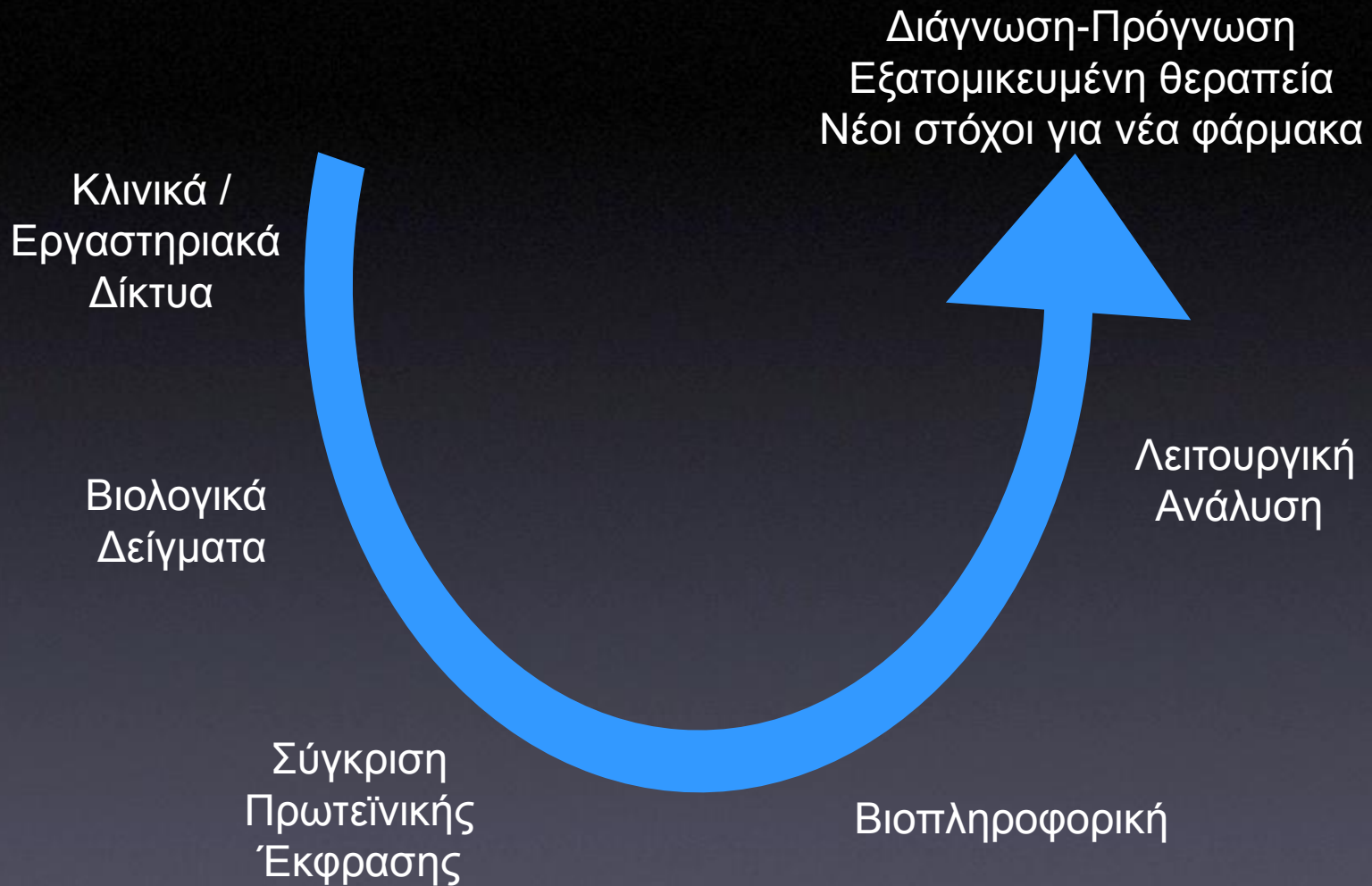
Mass spectrometric detection of 6,600 in vivo phosphorylation sites on 2,224 proteins

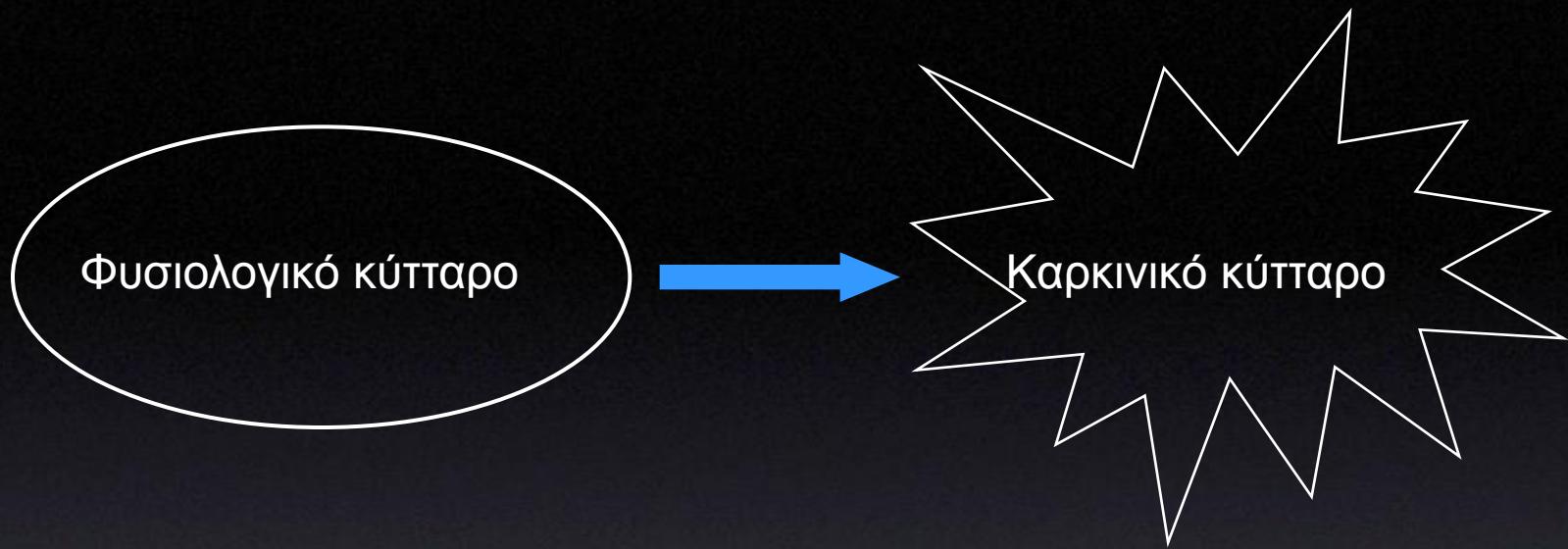
Olsen, J.V. et al. (2006) Cell 127, 635-648.





# Πρωτεωμική και ασθένειες





Πώς μετατρέπεται ένα φυσιολογικό κύτταρο σε καρκινικό;  
Ποιες διαφορές μπορούν να αξιοποιηθούν για  
διάγνωση / πρόγνωση / θεραπεία;

- ↓
- Σύγκριση:
- Γονιδιακής έκφρασης
  - Πρωτεϊνικής έκφρασης
  - Λειτουργίας πρωτεϊνών

## Εφαρμογές τις πρωτεωμικής τεχνολογίας στην διάγνωση

Παράδειγμα: Έγκαιρη διάγνωση καρκίνου ωοθηκών

Petricoin et al. Lancet Feb.16, 2002

- Καταγραφή των πρωτεϊνών του ορού σε 50 υγιείς γυναίκες και 50 ασθενείς
- Ανάπτυξη νέων εργαλείων βιοπληροφορικής για την ταυτοποίηση ενός προφίλ έκφρασης που καθορίζει με ακρίβεια την παρουσία καρκίνου
- Εφαρμογή σε 116 άγνωστα δείγματα:
  - ✓ Ταυτοποίηση όλων των καρκινικών δειγμάτων (50)
  - ✓ Ταυτοποίηση 63/66 μη καρκινικών δειγμάτων

Ευαισθησία: 100%

Εξειδίκευση: 95%

Θετικός δείκτης πρόγνωσης: 94%

## Proteomics and cancer: Running before we can walk?

### Serum Proteomics Scrutinized

SELDI-TOF still struggles to prove its worth as a clinical diagnostic tool

Methods for diagnosing early-stage cancer are in notoriously short supply. Although much ballyhooed work from National Cancer Institute/FDA researchers Emanuel Petricoin, Lance Liotta, and colleagues promised a revolution thanks to clinical proteomic screening, initial findings fell under scrutiny. Critics maintained that the process they used, surface-enhance laser desorption/ionization-time-of-flight (SELDI-TOF) mass spectrometry, lacked the sensitivity needed to detect low-abundance proteins. Sample preparation and instrument variation introduced noise into the experiment to make the results valid.

Nature 430, 611 (5 August 2004) | doi:10.1038/430611c; Published online 4 August 2004

### OvaCheck: breakthrough or also ran?

*The test for ovarian cancer uses proteomic pattern testing. It looked good in The Lancet two years ago. The results remain unreplicated*

### OvaCheck: doubts voiced soon after publication

#### Problems:

- Complexity of serum proteome
- Huge range of protein concentrations
- Heterogeneity between individuals
- Irreproducibility across time and between laboratories
- Non-specific changes (hormonal status, stress, inflammation, tissue necrosis, etc.)
- Lack of standardized collection and storage
- Lack of validation reagents
- Difficulties in bioinformatic analysis



## 'Omic diagnostics trip up on way to clinic

Malorye A. Branca

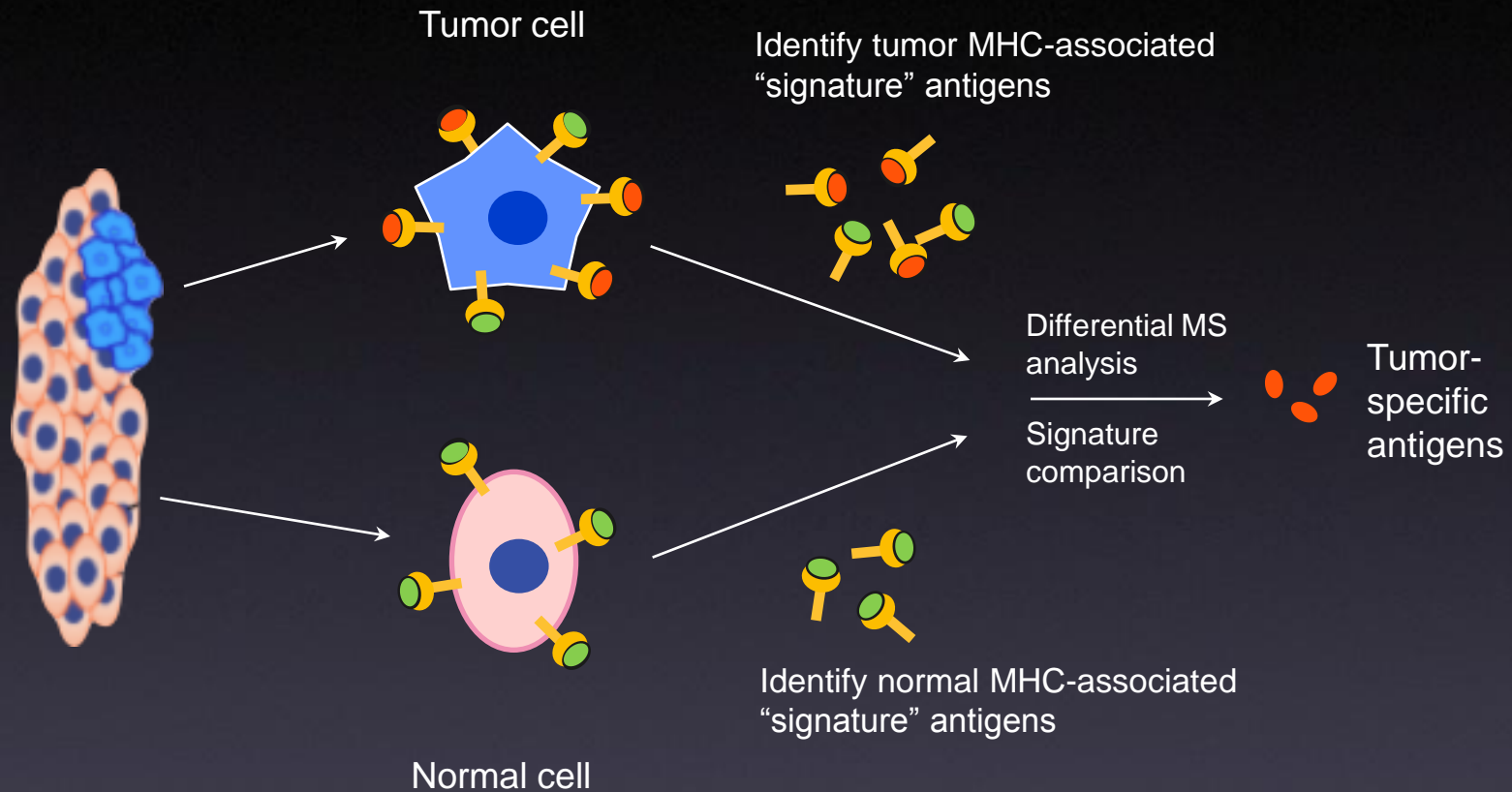
**Table 1 New and emerging 'omic tests**

Maker	Product	Approval status	Characteristics
Agendia	MammaPrint	Pending at FDA	A 70-gene expression signature, measured on an Agilent DNA microarray, predicts prognosis for certain breast cancer patients.
Correlogic Systems	OvaCheck	Pending at FDA	Protein expression pattern analysis. More than 150,000 starting data points are reduced to 5–10 for final evaluation
Exagen (Albuquerque, NM)	Breast cancer prognosis test	Submission to FDA for PMA anticipated by end of 2005	Traditional FISH test, derived from gene expression as well as other data.
Genomic Health	Oncotype Dx	Not an FDA-regulated product. Launched as laboratory-developed test in California, January 2004	A 21-gene expression assay done using RT-PCR for prognosis in breast cancer.
Roche	Leukemia array	Launch anticipated 2006–2007	A 300-gene expression microarray test.
Tm Bioscience	Ashkenazi Jewish Panel	Launched as analyte-specific reagent in June 2005, FDA-approved test anticipated in 2006	Tests 61 alleles in 8 genes, related to neurological diseases; uses same platform as cystic fibrosis test.

MB

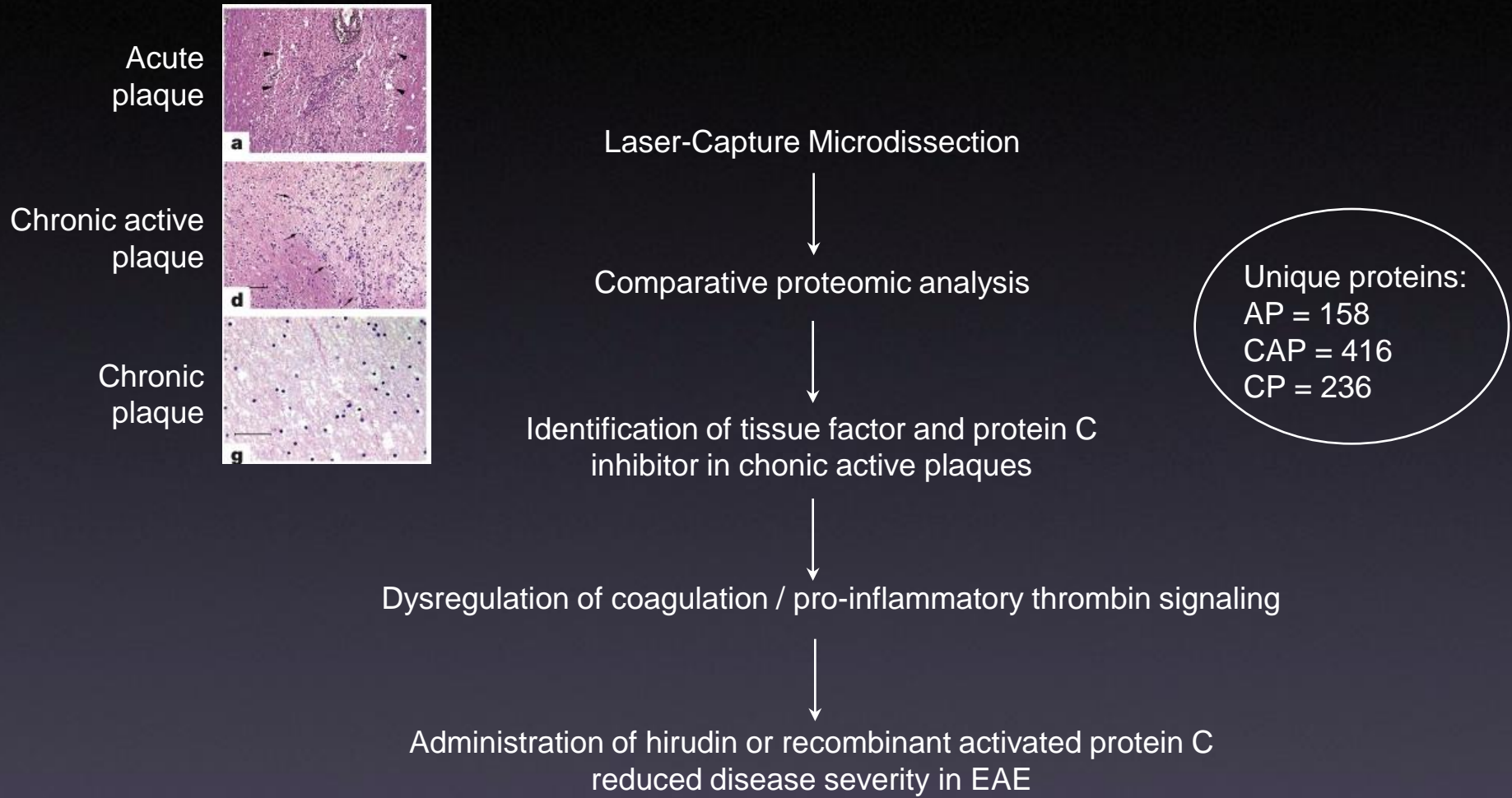
# Immunoproteome analysis

Identification of antigens for active immunotherapy

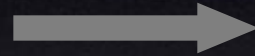


# Proteomic analysis of Multiple Sclerosis lesions

Identification of novel therapeutic targets



# Proteomic approaches in biomedical research



Mechanisms  
Biomarkers  
Disease Diagnosis & Prognosis  
Personalized treatment

Challenges: Improve plasma proteome analysis  
Achieve better proteome coverage  
Improve specimen quality  
Independent validation studies  
Combine with genomic and metabolomic technologies  
Improve bioinformatic tools



# Μικροσυστοιχίες πρωτεϊνών / πεπτιδίων

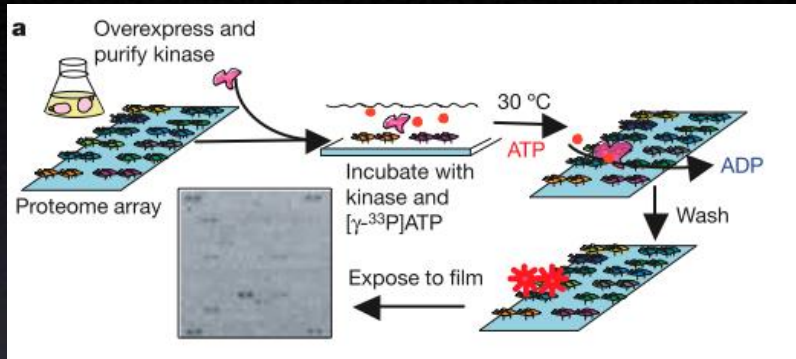
Συνδυάζουν τα πλεονεκτήματα των μικροσυστοιχιών  
(μεγάλη ταχύτητα και όγκος αναλύσεων, ευαισθησία, κλπ)  
με την ικανότητα ανάλυσης πρωτεϊνών

Συνολική ανάλυση σε κυτταρικά εκχυλίσματα του σχηματισμού συμπλόκων  
ή της ενζυμικής δραστηριότητας

## Παραδείγματα:

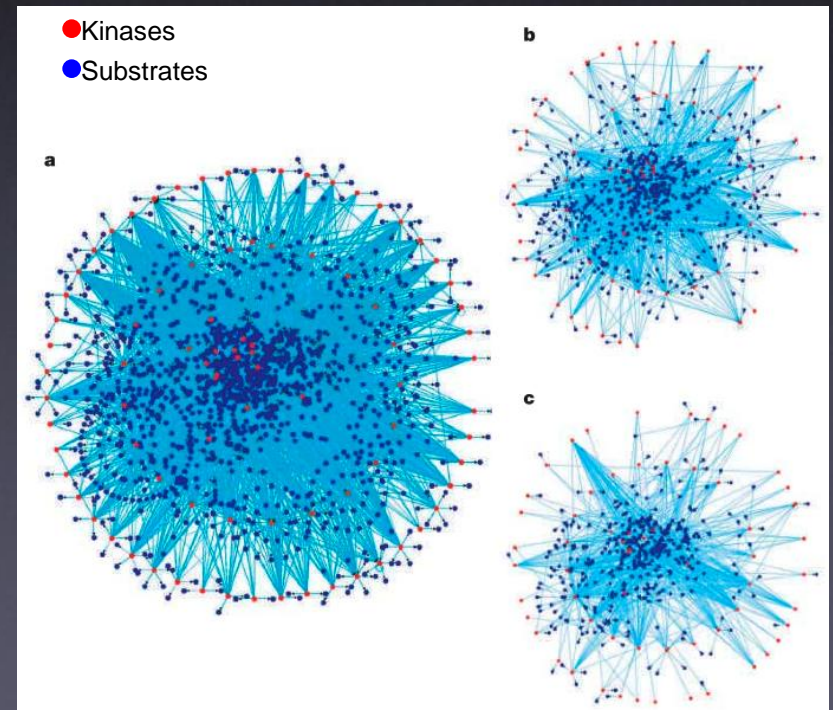
- Προσδιορισμός αλληλεπιδράσεων μεταξύ πρωτεϊνών
- Ανάλυση υποστρωμάτων κινασών και άλλων ενζύμων
- Αναγνώριση πρωτεϊνικών στόχων φαρμακευτικών ουσιών

# Global analysis of protein phosphorylation using protein microarrays

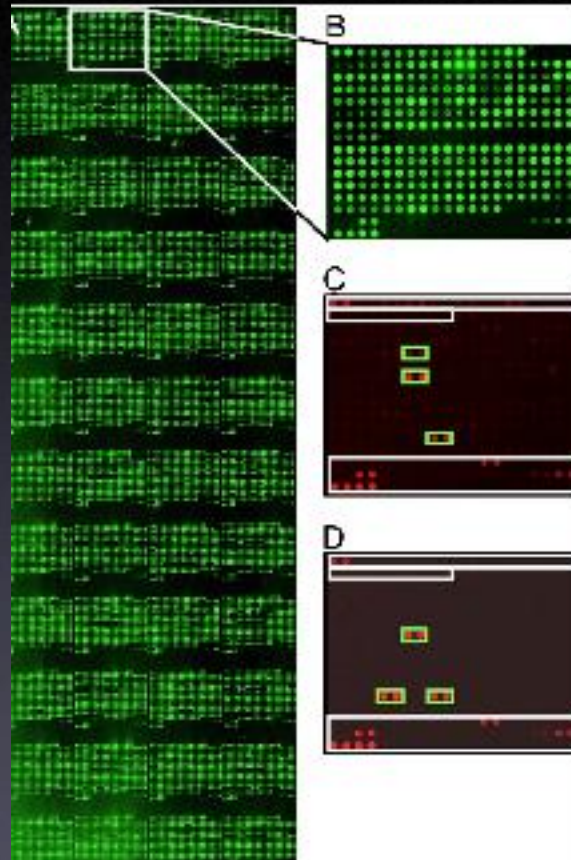


Proteome array:  
4400 yeast proteins spotted on a chip

87 kinases (out of 122 total)  
4200 phosphorylation events  
1325 proteins



# Identification of differentially expressed proteins in ovarian cancer using high-density protein microarrays



Detection of immobilized GST-fusion proteins

Detection with sera from normal controls

Detection with sera from patients



# Μελλοντικές κατευθύνσεις

Αυτοματισμός, διεκπεραίωση πολλών δειγμάτων

Καλύτερες μέθοδοι διαχωρισμού και ανίχνευσης πρωτεϊνών

Καλύτερη ποσοτικοποίηση διαφορικής έκφρασης

Καλύτερα εργαλεία βιοπληροφορικής

- Συνδυασμός:**
- μικροσυστοιχιών πρωτεϊνών (high-throughput, αυτοματισμός)
  - βιοαισθητήρων (ανίχνευση χωρίς σήμανση)
  - φασματογράφων μάζας (ευαισθησία)



# Βιβλιογραφία

1. Pandey & Mann (2000) Proteomics to study genes and genomes. Nature 405: 837 - 846
2. Shevchenko, A., et al. (1996) Linking genome and proteome by mass spectrometry: Large scale identification of yeast proteins from two dimensional gels. Proc. Natl.Acad.Sci.USA 93:14440-14445.
3. Wulfkuhle JD, Liotta LA, Petricoin EF. (2003) Proteomic applications for the early detection of cancer. Nat Rev Cancer. 3:267-75.
4. Aebersold R, Mann M. (2003) Mass spectrometry-based proteomics. Nature.; 422: 198-207.
5. Mann M, Jensen ON. Proteomic analysis of post-translational modifications. Nat Biotechnol. 2003 Mar; 21(3): 255-61. Review.
6. Focus on mass spectrometry in proteomics. Nature Methods, 4, 781-833 (2007)
7. <http://www.nature.com/reviews/focus/proteomics/index.html>