



Μέθοδοι Πρωτεωμικής Ανάλυσης

Αγγελική Κατσαφάδου DVM, MSc, PhD

Αναπληρώτρια Καθηγήτρια, Τμήμα Δημόσιας και Ενιαίας Υγείας
Πανεπιστήμιο Θεσσαλίας

Fulbright Alumna, Yale School of Public Health

Στόχοι

- Οπics & Πρωτεωμική
- Ανάλυση αλληλουχιών σε πρωτεώματα και πρωτεΐνες
- Εφαρμογές



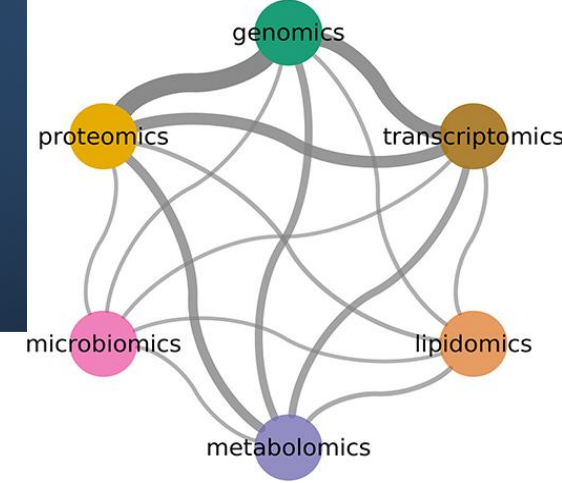
Ποιες είναι οι
ωμικές (omics)
τεχνολογίες;;;



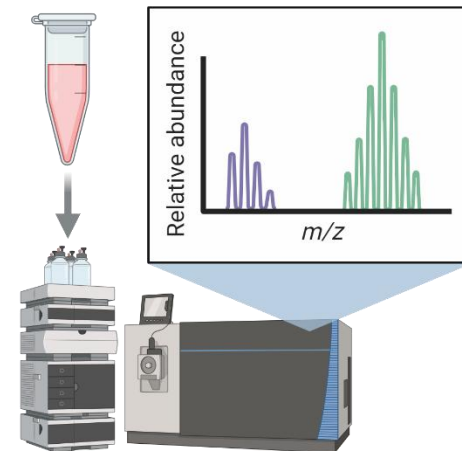
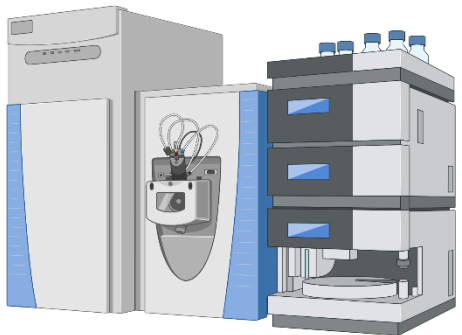
Γιατί τις
μελέταμε;;



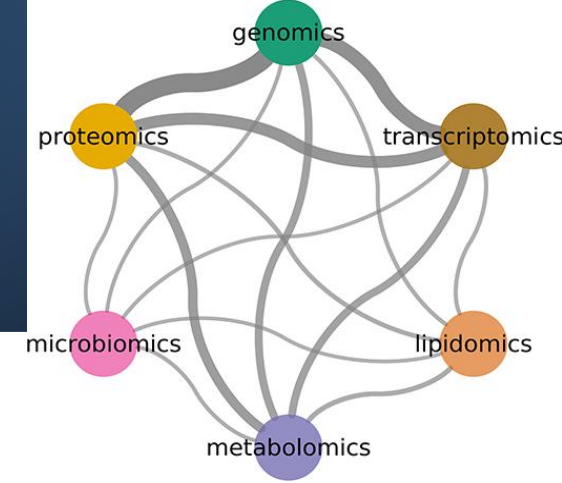
OMICS



Τεχνολογίες υψηλής απόδοσης
(high-throughput)
που χρησιμοποιούνται για την ολιστική
ανάλυση των βιολογικών μορίων



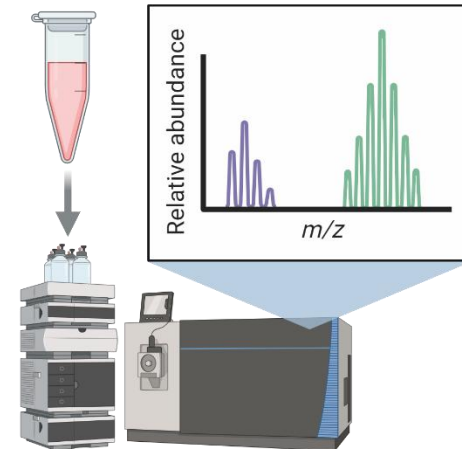
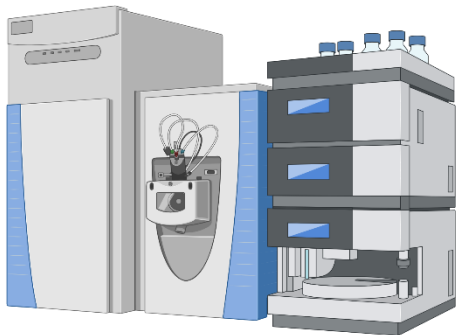
OMICS



✓ προσέγγιση μελέτης

-Ωμική:

Η κατάληξη αφορά τη μελέτη ενός συνόλου καθολικά (προηγούμενου συνθετικού της λέξης)



OMICS

 National Human Genome Research Institute

[ABOUT GENOMICS](#)[RESEARCH FUNDING](#)[RESEARCH AT NHGRI](#)[ABOUT HEALTH](#)[CAREERS & TRAINING](#)[NEWS & EVENTS](#)[ABOUT NHGRI](#)

...Begin your search here

The Human Genome Project

The Human Genome Project (HGP) is one of the greatest scientific feats in history. The project was a voyage of biological discovery led by an international group of researchers looking to comprehensively study all of the DNA (known as a genome) of a select set of organisms. Launched in October 1990 and completed in April 2003, the Human Genome Project's signature accomplishment – generating the first sequence of the human genome – provided fundamental information about the human blueprint, which has since accelerated the study of human biology and improved the practice of medicine.

Learn more about the Human Genome Project below.



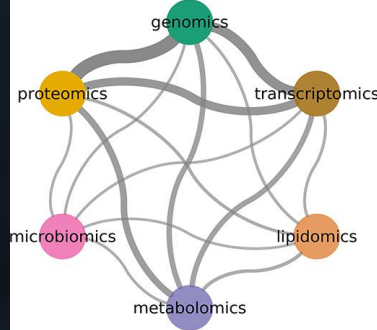
**HUMAN
PROTEOME
PROJECT**

TRANSLATING THE CODE OF LIFE

HPP

twitter | linkedin

-Ωμικές -Omics



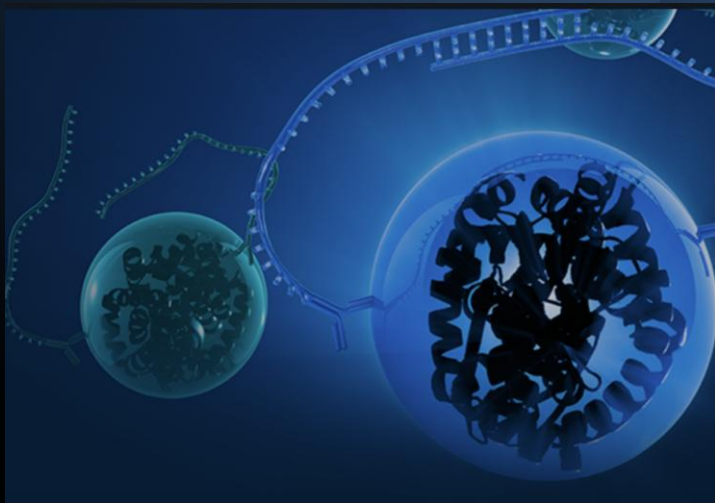
Genomics

Transcriptomics

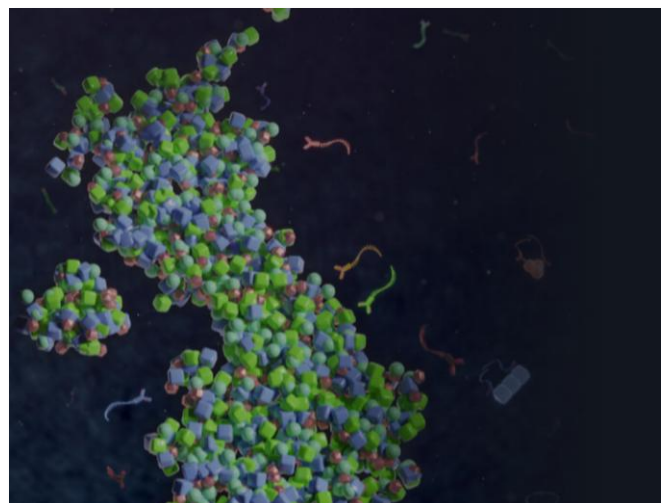
Proteomics

Metabolomics

Multi-omics



Proteogenomics

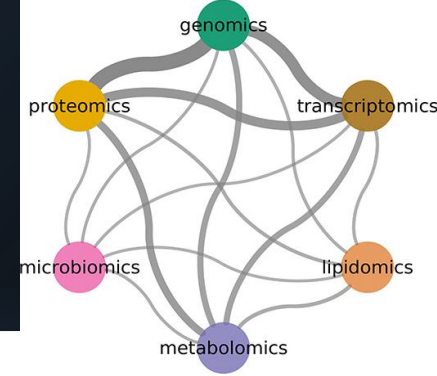


Realizing the potential of
population-scale
proteogenomics

Landmark articles in Nature present impressive findings from the UK
Biobank Pharma Proteomics Project

[Learn more](#) >

Multi-omics



Συνδυασμός διαφόρων τεχνολογιών omics



Πλεονεκτήματα στην ανάλυση βιολογικών υγρών



Ταυτόχρονη ταυτοποίηση γονιδίων, πρωτεϊνών και μεταβολιτών

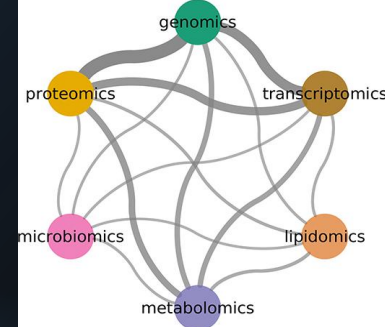


Διευκρίνιση / αποσαφήνιση των μοριακών τους μονοπατιών



Βελτιωμένα προγνωστικά μοντέλα με τη χρήση αλγορίθμων μηχανικής μάθησης

-Ωμικές -Omics



- ❖ >200 ‘*-omics*’ υποκατηγορίες έχουν πια εμφανιστεί
- ❖ Ολιστική μελέτη των σχετικών μορίων
- ❖ Μόδα

epigenomics

lipidomics

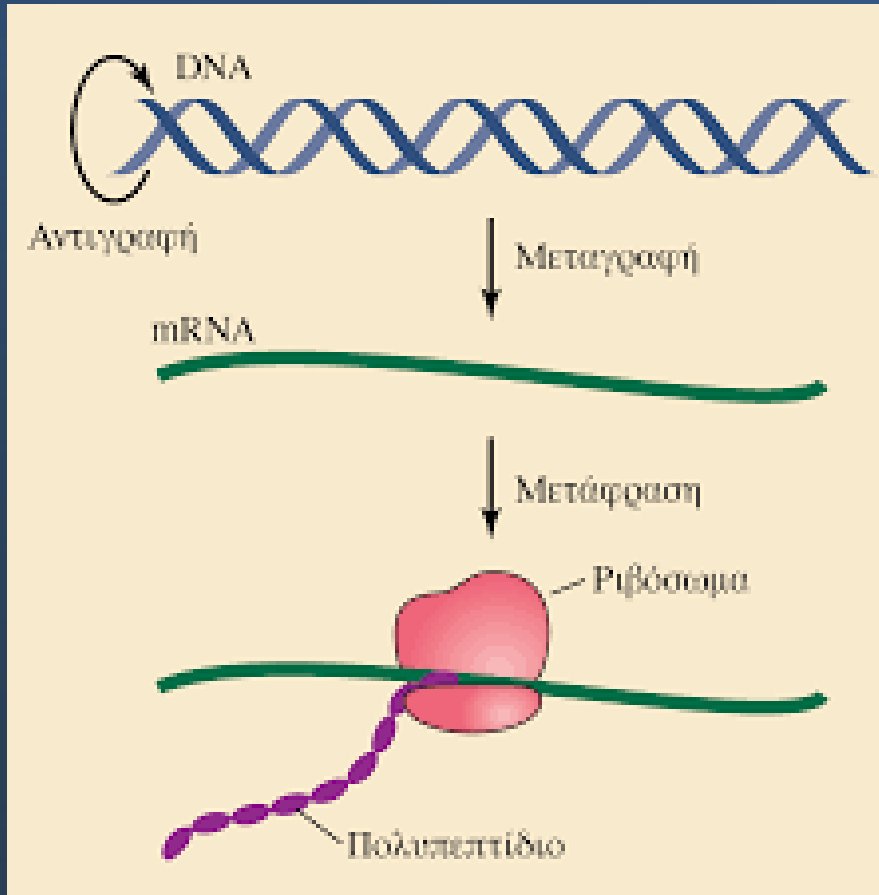
glycomics

interactomics

nutrigenomics

foodomics

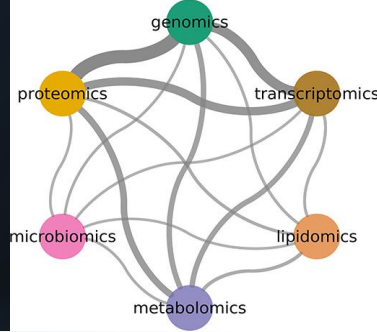
Ροή γενετικής πληροφορίας



Μετα-
μεταφραστικές
τροποποιήσεις



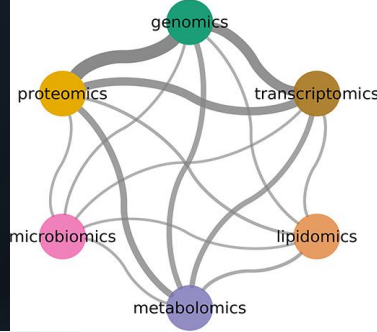
Ροή γενετικής πληροφορίας



- ❖ **DNA:** αυτό που ενδέχεται να συμβεί
- ❖ **RNA:** αυτό που πιθανότατα θα συμβεί
- ❖ **Πρωτεΐνες:** αυτό που συμβαίνει

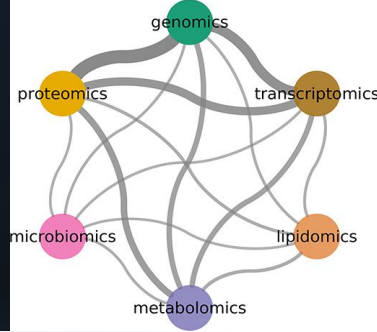
Γονίδια φορείς οδηγιών -> πρωτεΐνες μόρια που εκτελούν τις οδηγίες

Εφαρμογές omics



- ❖ Χρησιμοποιούνται στην ιατρική ακριβείας (precision medicine) ή εξατομικευμένη ιατρική
- ❖ Χαρτογράφηση ανθεκτικότητας νέων ασθενειών
- ❖ Χαρτογράφηση ιών (π.χ. covid)

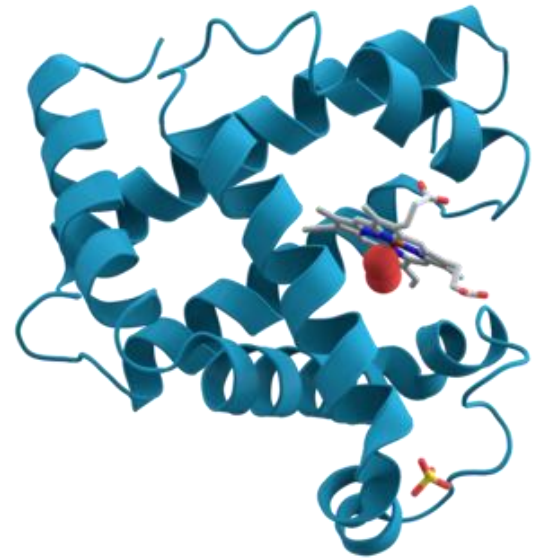
Εφαρμογές omics



- ❖ Μελέτες παθογένειας νοσημάτων
- ❖ Εύρεση βιοδεικτών
- ❖ Ταυτοποίηση φαρμακευτικών στόχων

Πρωτεωμική (Proteomics)

= protein + genome
(Marc Wilkins 1996)



➤ Ανάλυση και μελέτη του συνόλου των πρωτεϊνών ενός ιστού σε μια δεδομένη χρονική στιγμή.

Πρωτεωμική: γιατί;

Πρωτεΐνες: ενεργά και κυρίαρχα μόρια που αντικατοπτρίζουν φυσιολογικές ή παθολογικές καταστάσεις

Αναγνώριση πρωτεϊνών
που υπάρχουν σε
δείγματα ενδιαφέροντος

Ποσοτικοποίηση των
μεταβολών στην έκφραση
πρωτεϊνών κατά τη
διάρκεια παθολογικών ή
διαφορετικών
καταστάσεων

Ανίχνευση Βιοδεικτών

Γονιδιωματική vs. Πρωτεωμική



Ίδιο **γονιδίωμα...**



...διαφορετικά **πρωτεώματα**

Γονιδιωματική vs. Πρωτεωμική

- Το ανθρώπινο γονιδίωμα περιλαμβάνει περίπου **30.000 γονίδια**
- Το πρωτέωμα ενός ατόμου σε μια δεδομένη χρονική στιγμή μπορεί να περιλαμβάνει περίπου **100.000** προϊόντα γονιδίων

Το πρωτέωμα ενός ατόμου καθ' όλη τη διάρκεια της ζωής του μπορεί να περιλαμβάνει περίπου **1.000.000** προϊόντα γονιδίων

Πρωτεωμική

- Ποιοτική & ποσοτική εκτίμηση των πρωτεϊνών ως αποτέλεσμα έκφρασης του γονιδιώματος
- Μαζική & ταυτόχρονη μελέτη μεγάλου αριθμού πρωτεϊνών, ειδικότερα:
 - ✓ της δομής
 - ✓ των μετα-μεταφραστικών μεταβολών
 - ✓ των αλληλεπιδράσεων με άλλα βιολογικά μόρια

Πρωτεωμική

□ Εφαρμογές:

- ✓ διαλεύκανση παθογενετικών μηχανισμών νοσημάτων
- ✓ ταυτοποίηση βιοδεικτών για πρόιμη και ακριβή διάγνωση νοσημάτων
- ✓ παραγωγή εμβολίων

Μεθοδολογίες

- ❑ Ηλεκτροφόρηση 2 διαστάσεων (2-DE ή 2-D gel electrophoresis)
- ❑ Υγρή χρωματογραφία (LC ή HPLC)
- ❑ Φασματομετρία Μάζας (MS ή MS/MS)
- ❑ Βιοπληροφορική (Bioinformatics)

Υπολογιστική Ανάλυση Αλληλουχιών;;

- Η ταυτοποίηση πρωτεϊνών με φασματομετρία μάζας μπορεί να ιδωθεί καθαρά ως πρόβλημα ανάλυσης αλληλουχιών.
- Σε αντίθεση με το DNA ή το RNA, δεν παρατηρούμε ποτέ άμεσα την αλληλουχία της πρωτεΐνης.
- Αυτό που έχουμε είναι έμμεσες μετρήσεις μάζας.

*“We never see the protein sequence directly
— we infer it.”*



Proteomic analysis-methodologies

- Protein separation



2D gel electrophoresis
(gel dependent)

Gel free methods (LC-MS/MS)

Bottom-up proteomics

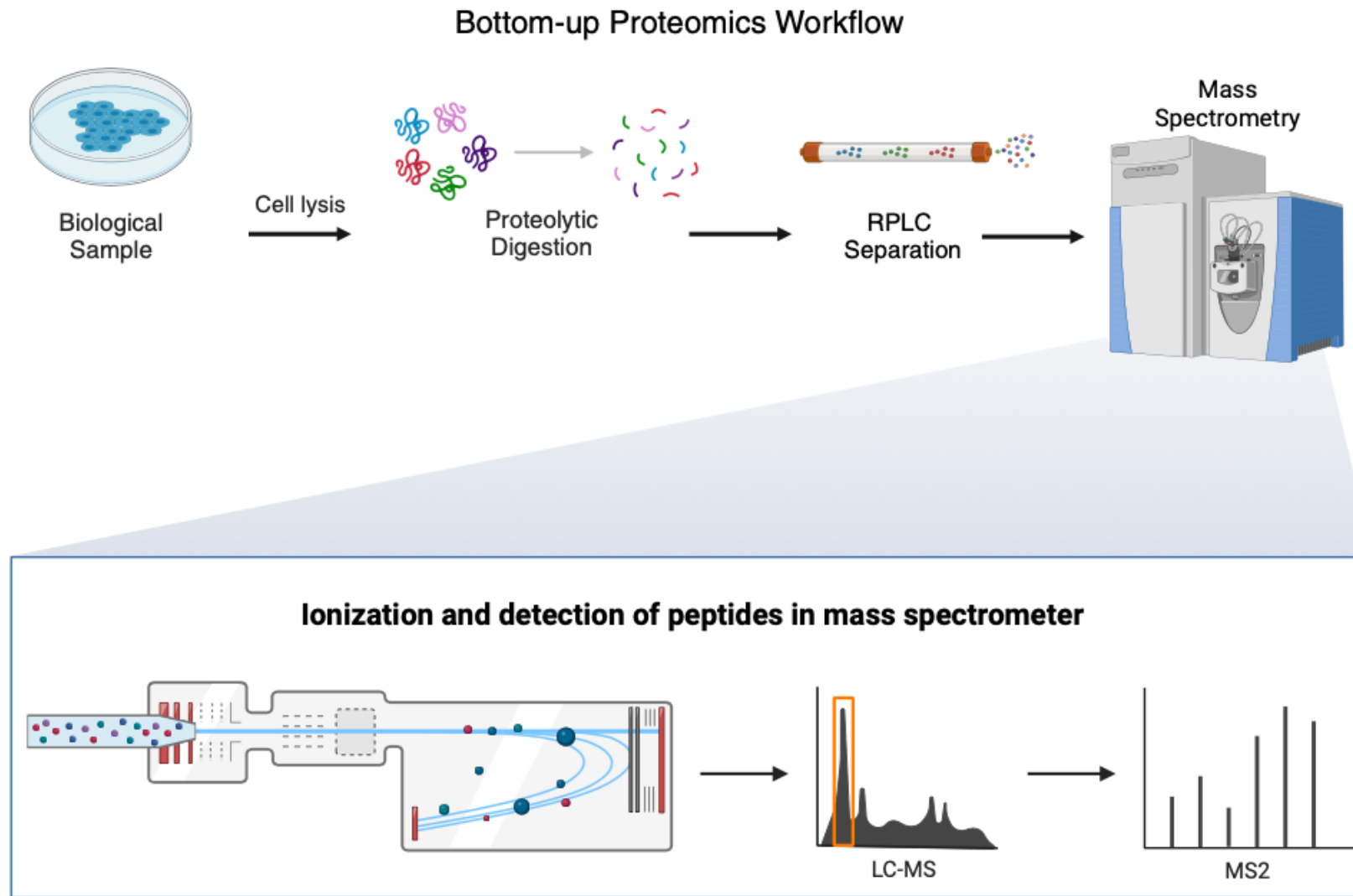
- Προσέγγιση από κάτω-προς-τα-πάνω (**bottom-up proteomics**):
ανάλυση των πεπτιδίων ή των θραυσμάτων τους διότι ξεκινά με τα **κομμάτια της πρωτεΐνης** και όχι με ολόκληρη την πρωτεΐνη.
- Υδρόλυση με **θρυψίνη** (tryptic digest peptides):
υδρολύει τις πρωτεΐνες, κόβοντας την αλληλουχία τους στην καρβοξυλική περιοχή της λυσίνης (K) ή της αργινίνης (R), όταν δεν ακολουθούνται από το αμινοξύ προλίνη (P)

NRRPCHSHTK ECESAWK NR PCHSHTKK PCHSHTKK NRKVWK IPPFFW

↓ trypsin digest

~~NR~~ PCHSHTK ~~NR~~ ~~K~~ ~~K~~ ~~K~~ ~~IPPFFW~~
R ECESAWK PCHSHTK PCHSHTK ~~NR~~ VWK

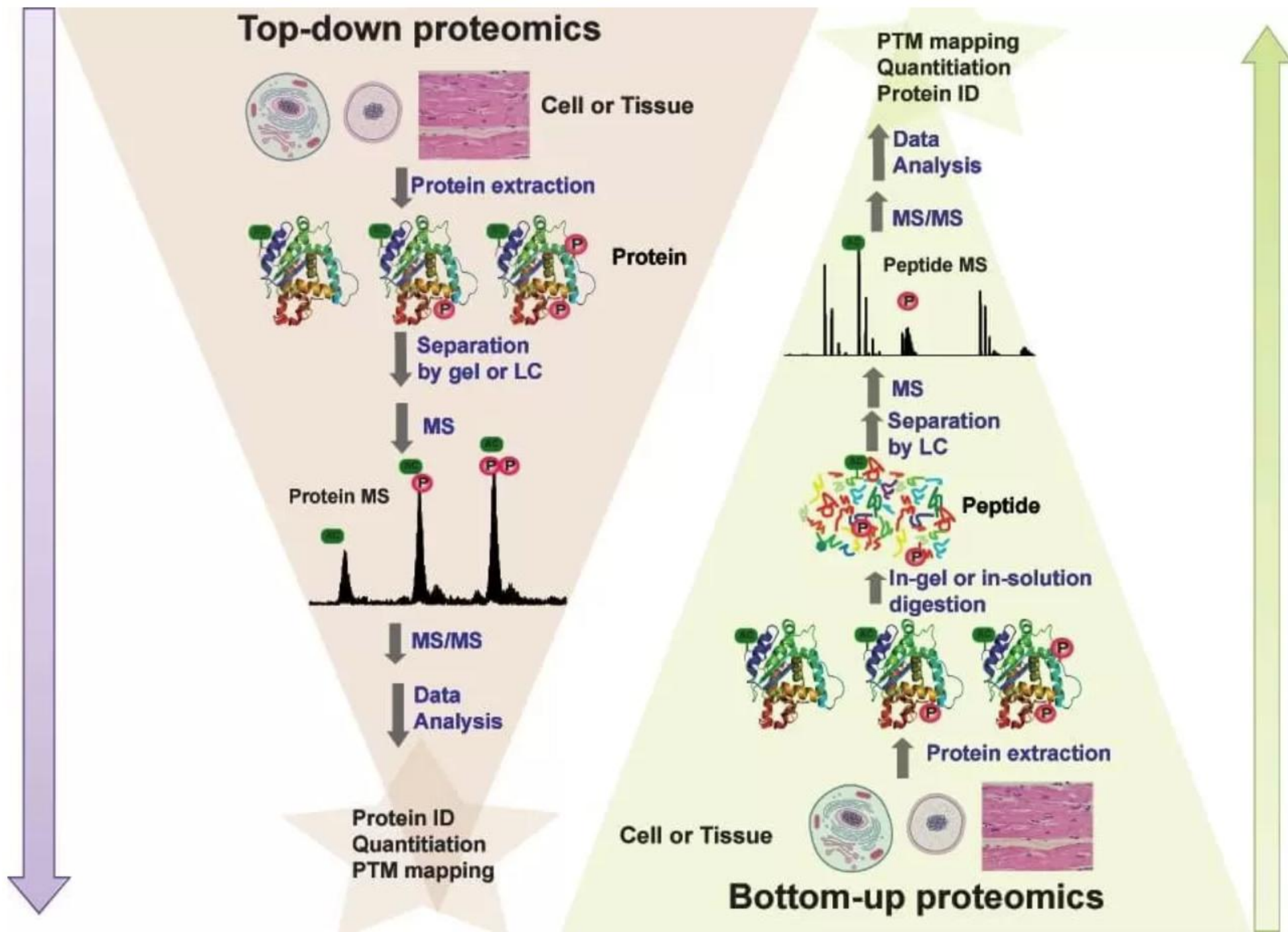
Bottom-up προσέγγιση: Protein → peptides → spectra → sequences → protein



Denaturation – Reduction – Alkylation -> Πέψη με τρυψίνη -> Sequence link -> Προβλέψιμα πεπτιδικά άκρα *In silico* digestion από databases

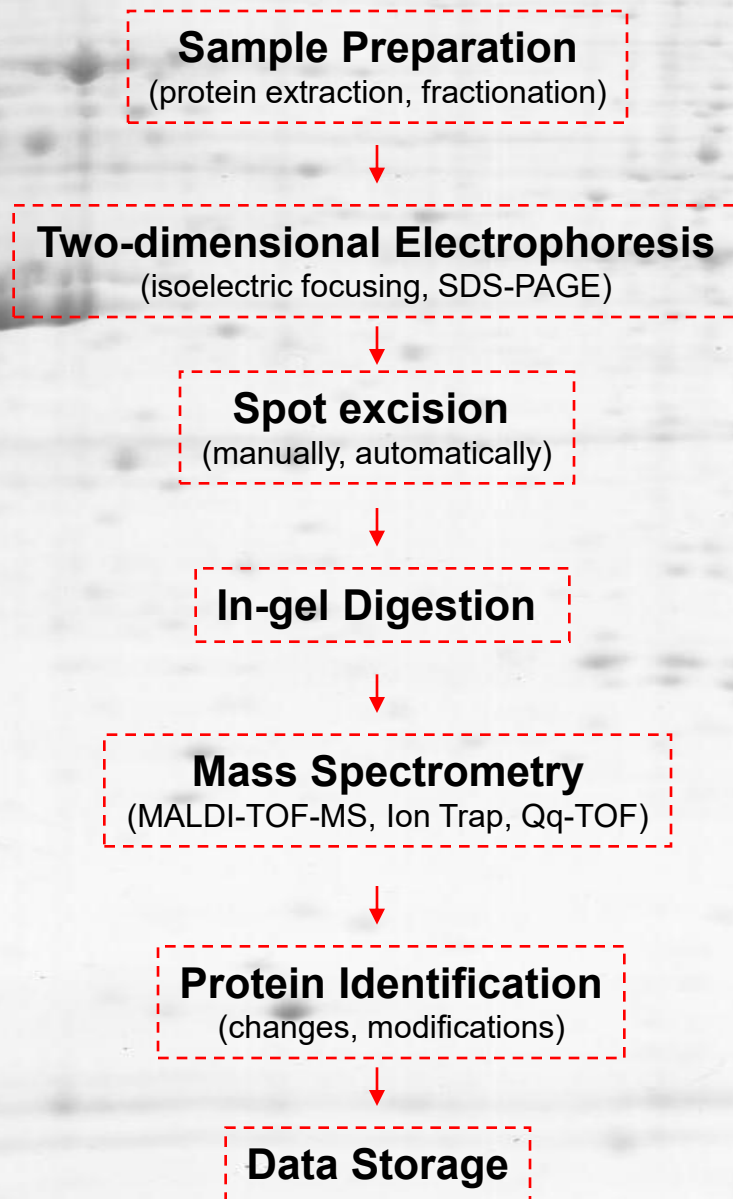
Top-down proteomics

- Προσέγγιση από πάνω-προς-τα-κάτω (**top-down proteomics**): αρχίζει με την ακέραιη πρωτεΐνη
- Πρωτεΐνες μοριακού βάρους < 50 KDa μπορούν να ιονιστούν και να κατακερματιστούν μέσω μιας διαδικασίας ανάλογης της ανάλυσης θραυσμάτων πεπτιδικών ιόντων, με τη διαφορά ότι δημιουργούνται μεγαλύτερα θραύσματα.
- μεγαλύτερη πρόκληση για πρωτεΐνες των οποίων η μάζα ξεπερνά σε μέγεθος τα 50KDa (κατά προσέγγιση 500 αμινοξέα)- ανώτερο αξιόπιστο όριο



Workflow in Proteomics

2D GEL DEPENDENT METHODS



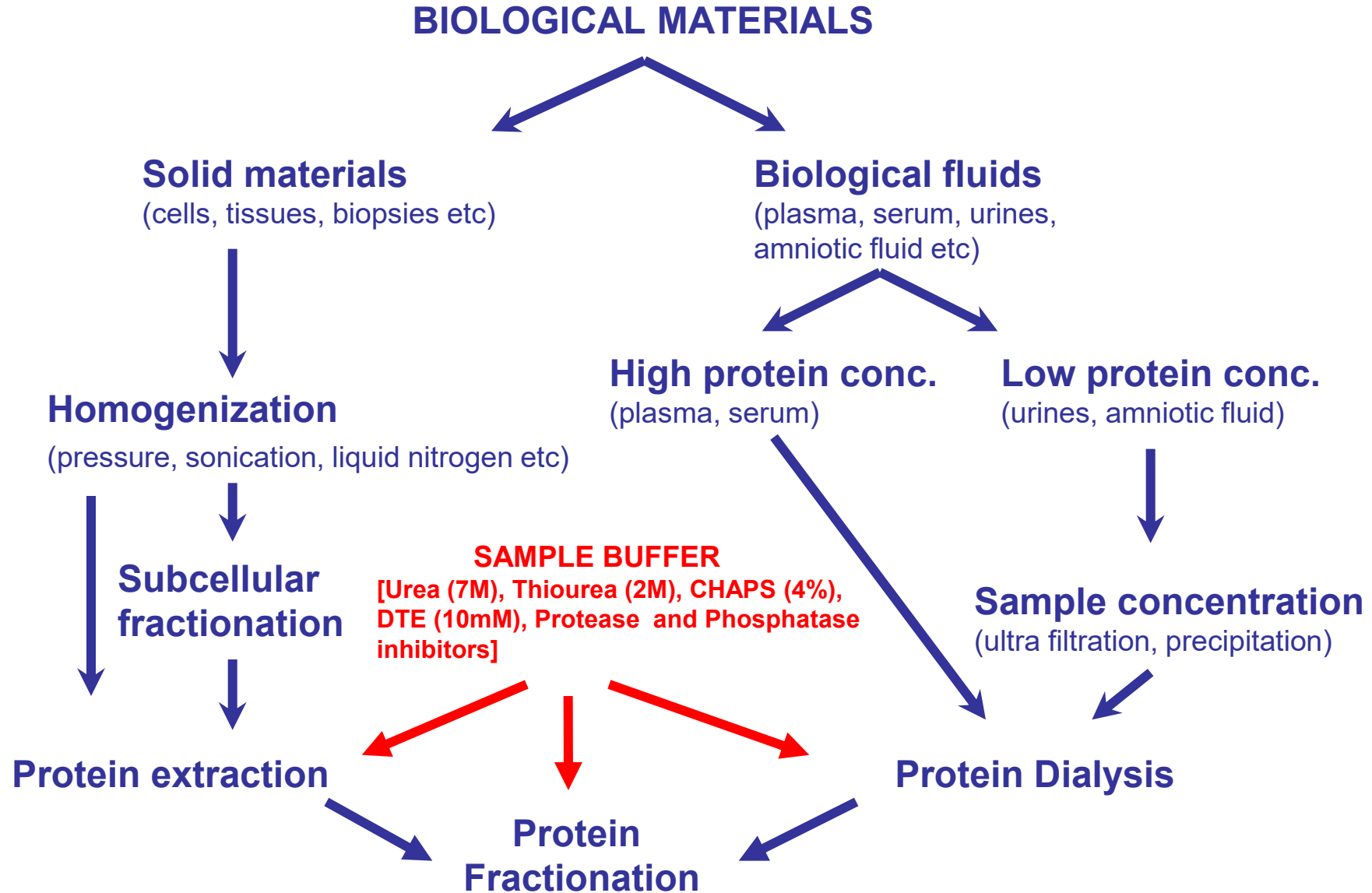
Ισοηλεκτρικό σημείο πρωτεΐνης

- Τα αμινοξέα μπορούν να εμφανίσουν τόσο τον **όξινο** (-COOH) όσο και το **βασικό** (-NH₂) χαρακτήρα, έχουν δηλαδή **αμφολυτικό χαρακτήρα**.
- Σε *ουδέτερο* pH, στις πρωτεΐνες, κάποια αμινοξικά κατάλοιπα είναι γενικά **θετικά φορτισμένα**, ενώ άλλα κατάλοιπα είναι γενικά **αρνητικά φορτισμένα**.
- Η ισορροπία μεταξύ αρνητικών και θετικών φορτίων σε ένα πρωτεϊνικό μόριο εξαρτάται από το pH του περιβάλλοντος

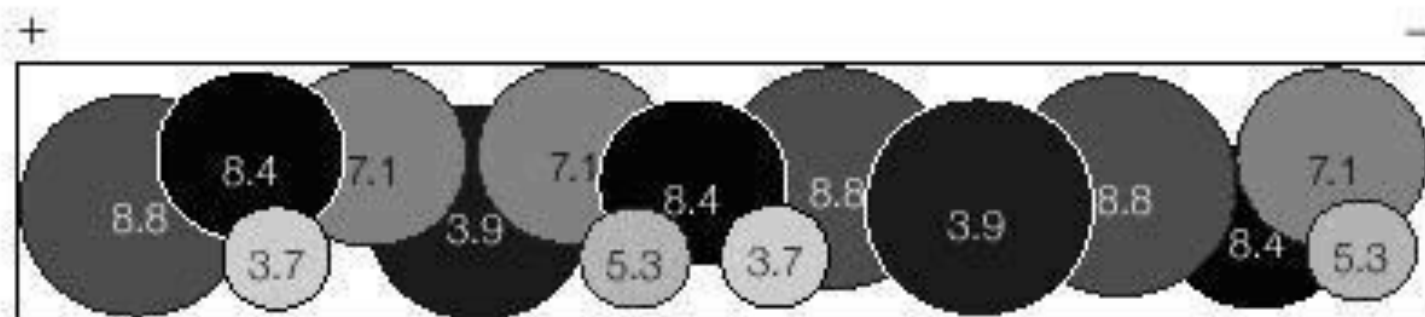
Ισοηλεκτρικό σημείο πρωτεΐνης (pI):
το pH όπου το δίκτυο φορτίων της πρωτεΐνης είναι μηδέν

Workflow in Proteomics

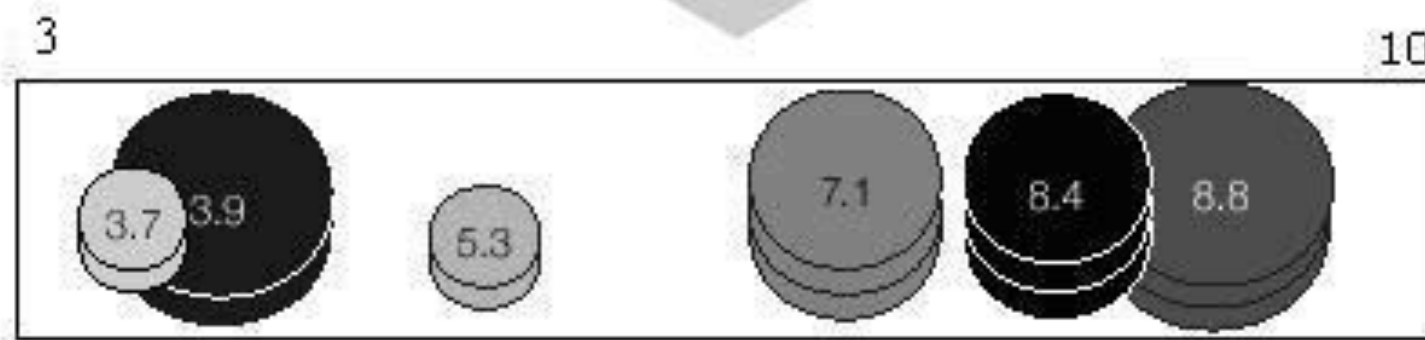
SAMPLE PREPARATION



Ισοηλεκτρική εστίαση - πρώτη διάσταση



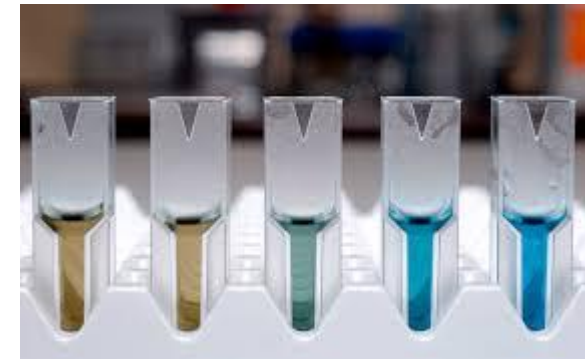
Πρίν τον εστιασμό



Μετά τον εστιασμό

Ποσοτικοποίηση πρωτεϊνών

- *1mg/ml*
- **Bradford protein assay**

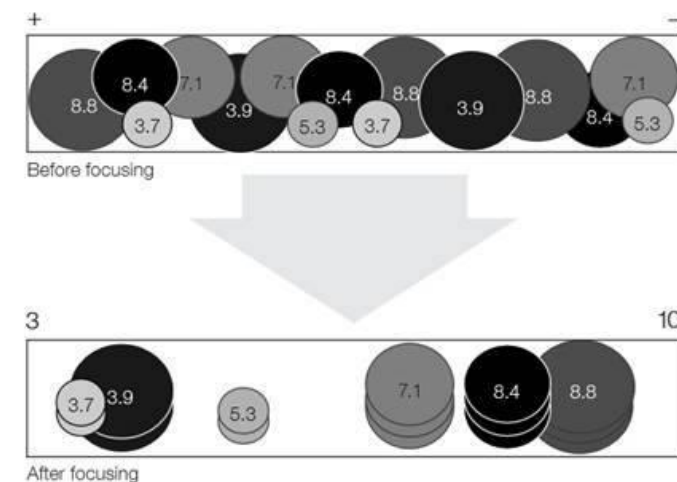
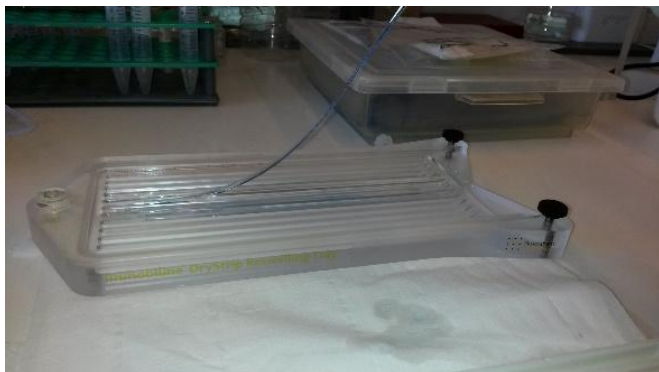


Μεθοδολογίες- Ηλεκτροφόρηση δυο διαστάσεων

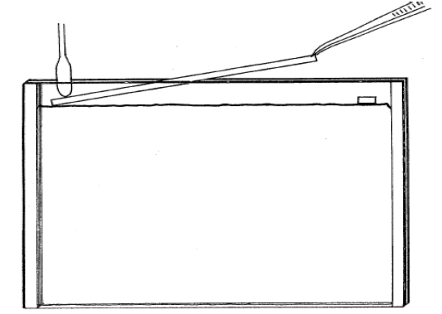


Ηλεκτροφόρηση δύο διαστάσεων (2-D gel)

1η διάσταση: Ισοηλεκτρική εστίαση
οριζόντιος διαχωρισμός πρωτεϊνών 1mg/ml τεχνική που διαχωρίζει τις
πρωτεΐνες σύμφωνα με το ηλεκτρικό τους φορτίο

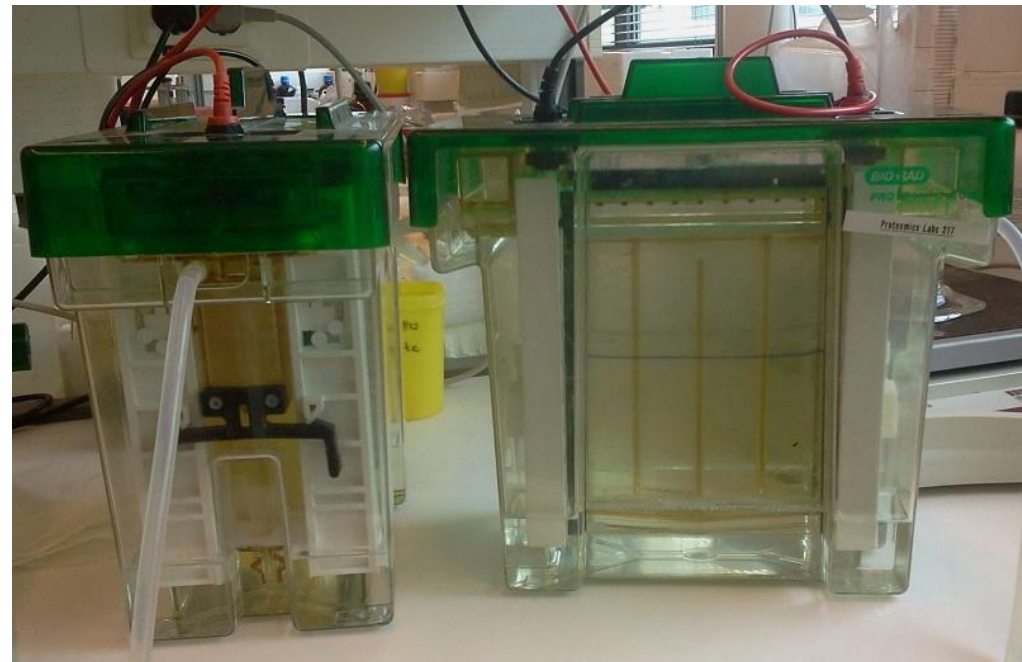
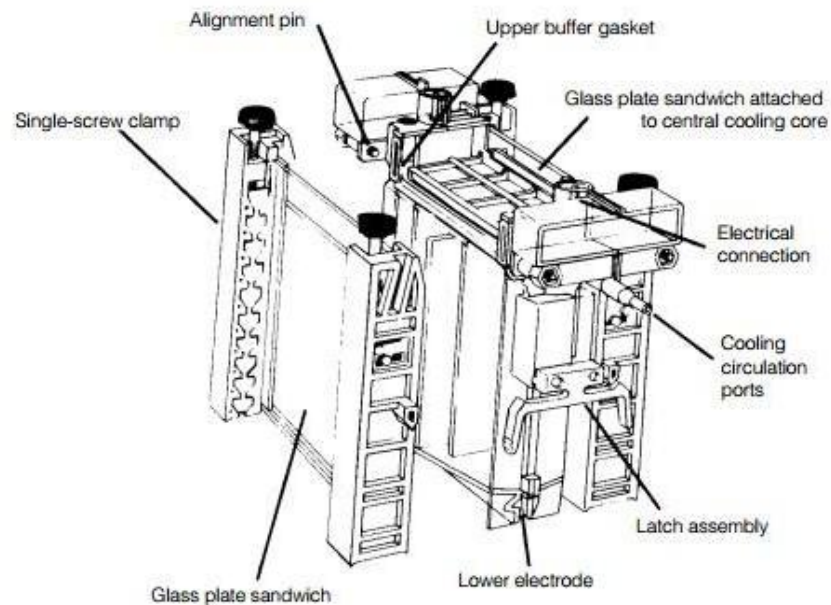


Μεθοδολογίες- Ηλεκτροφόρηση δυο διαστάσεων



Δεύτερη διάσταση

- 2η διάσταση: κατακόρυφος διαχωρισμός πρωτεϊνών σε διάλυμα 12% gel polyacrylamide



pH 3

pH10



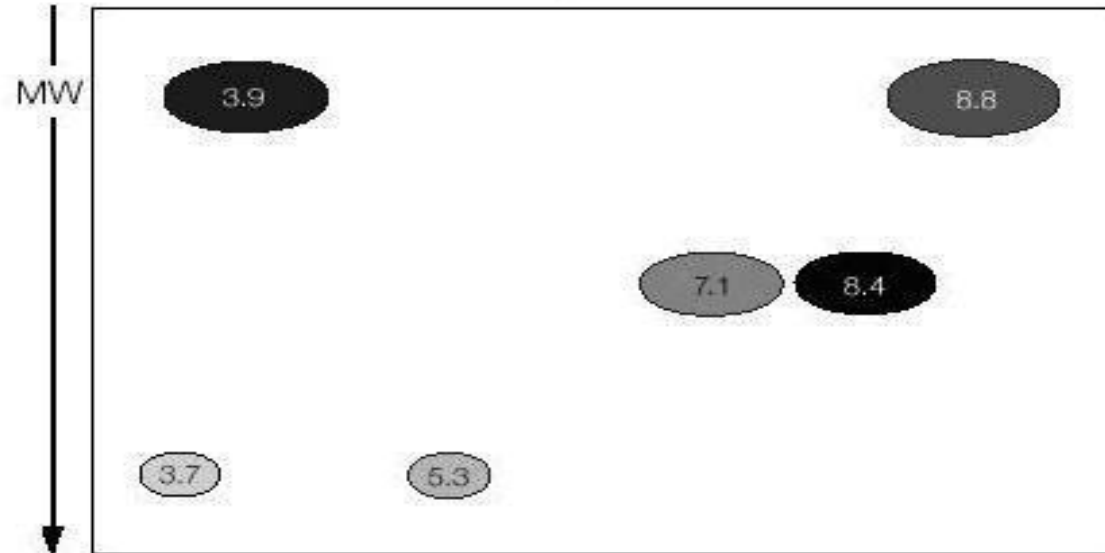
εστιασμένες πρωτεΐνες

εξισορρόπηση σε SDS και αναγωγικά μέσα
για σταθερό πρωτεϊνικό σχήμα, μονές
υπομονάδες, σταθερό αρνητικό
 m/z λόγο



φορτισμένες με SDS πρωτεΐνες στο IPG strip

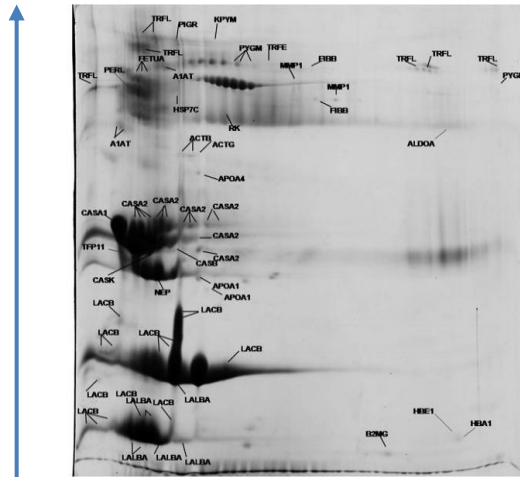
εφαρμογή στο SDS-PAGE gel



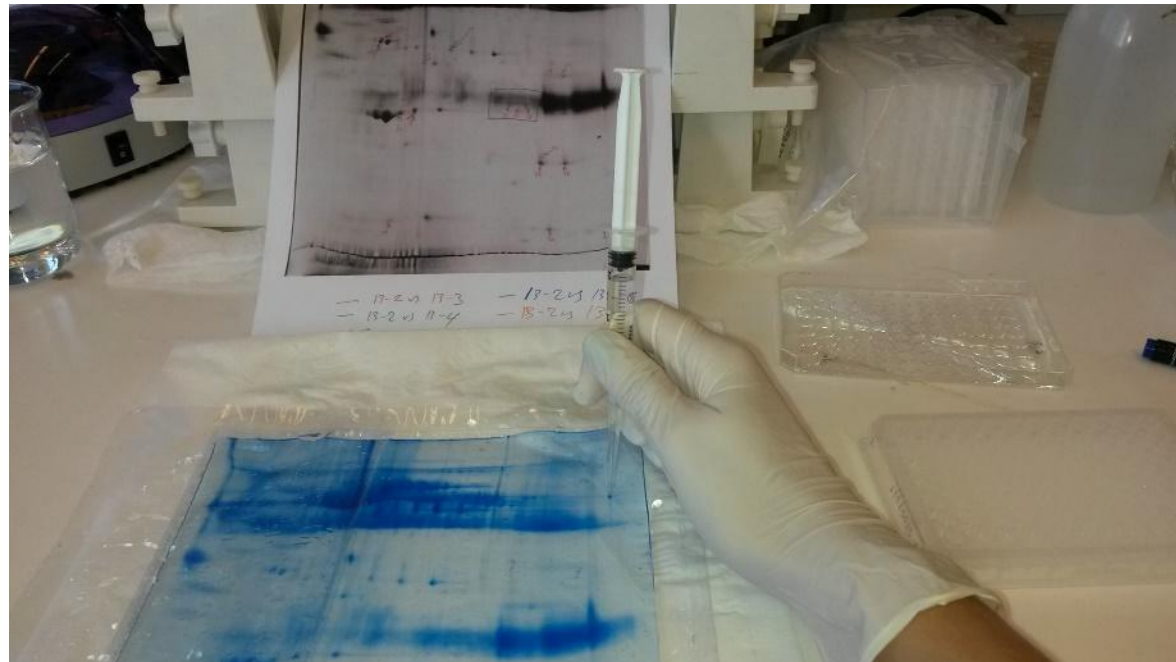
οι φορτισμένες με SDS πρωτεΐνες διαχωρίζονται ανάλογα με το μέγεθός τους

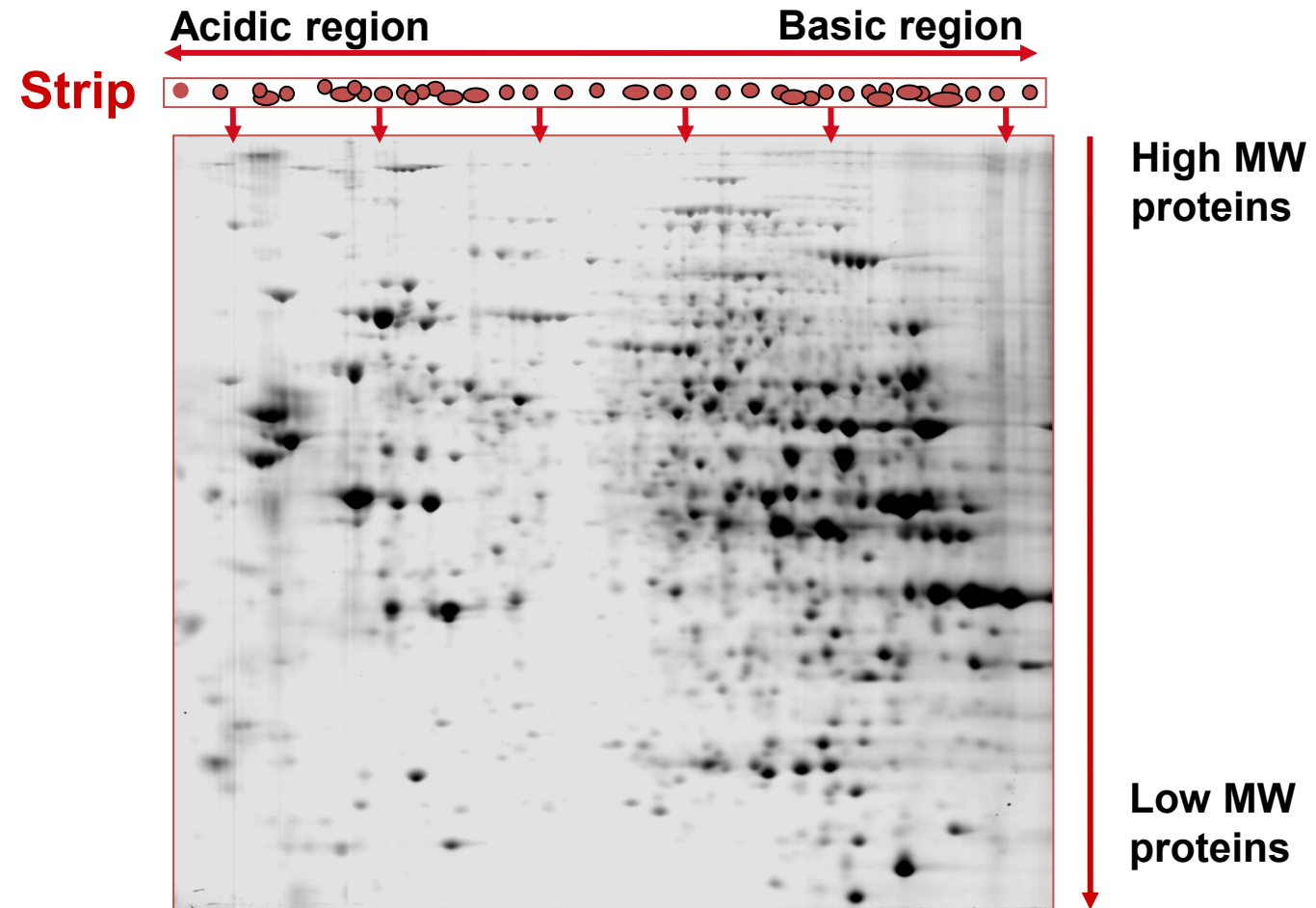
Μεθοδολογίες- Ηλεκτροφόρηση δυο διαστάσεων

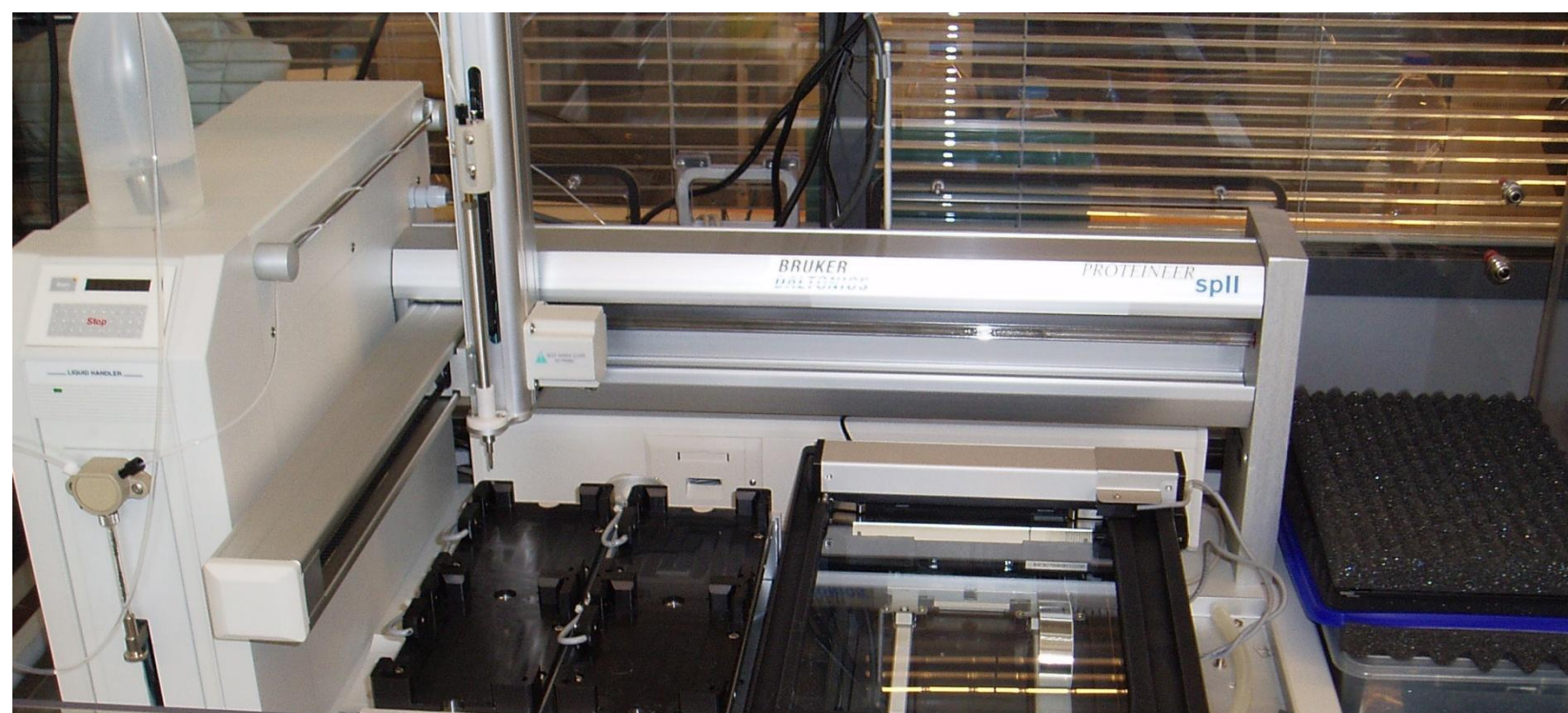
Χρώση των πηκτωμάτων (gel) με κυανό της κουμαρίνης – 2-DE gels

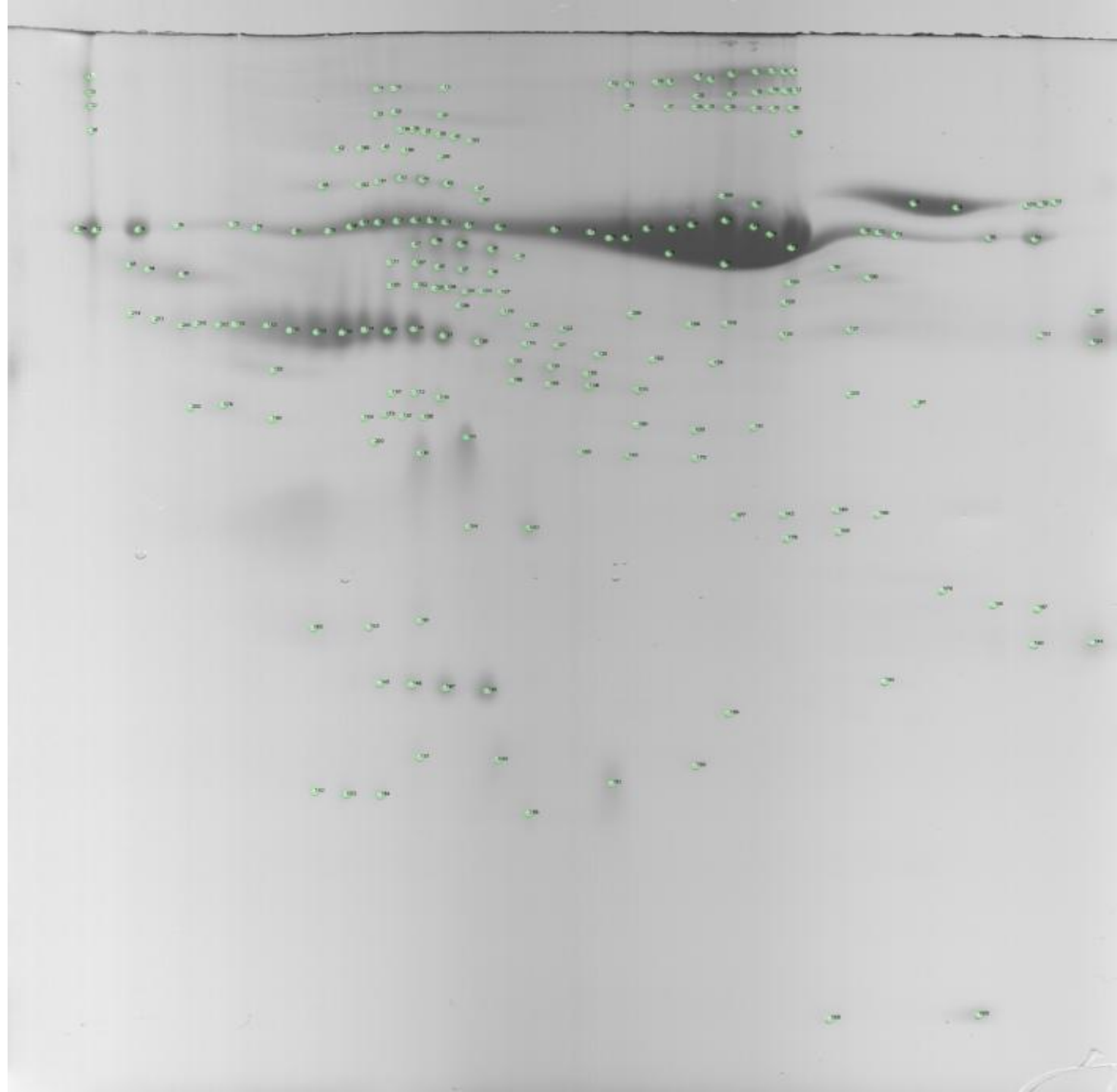


2 άξονες X, Y

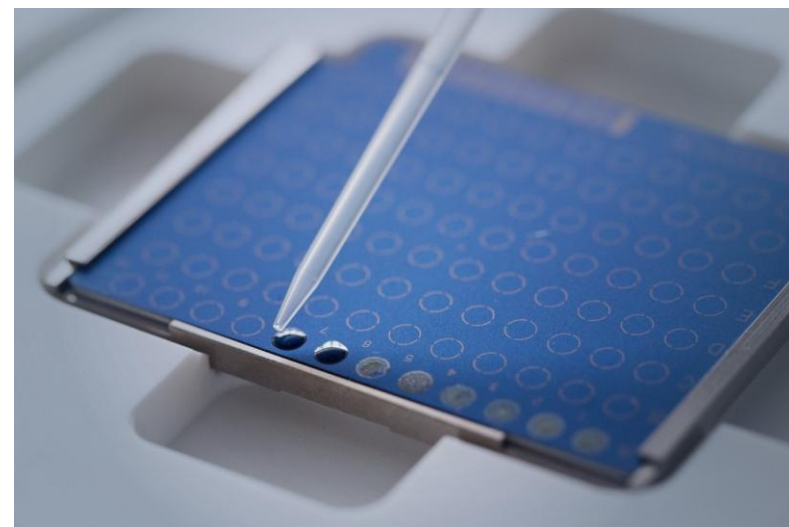
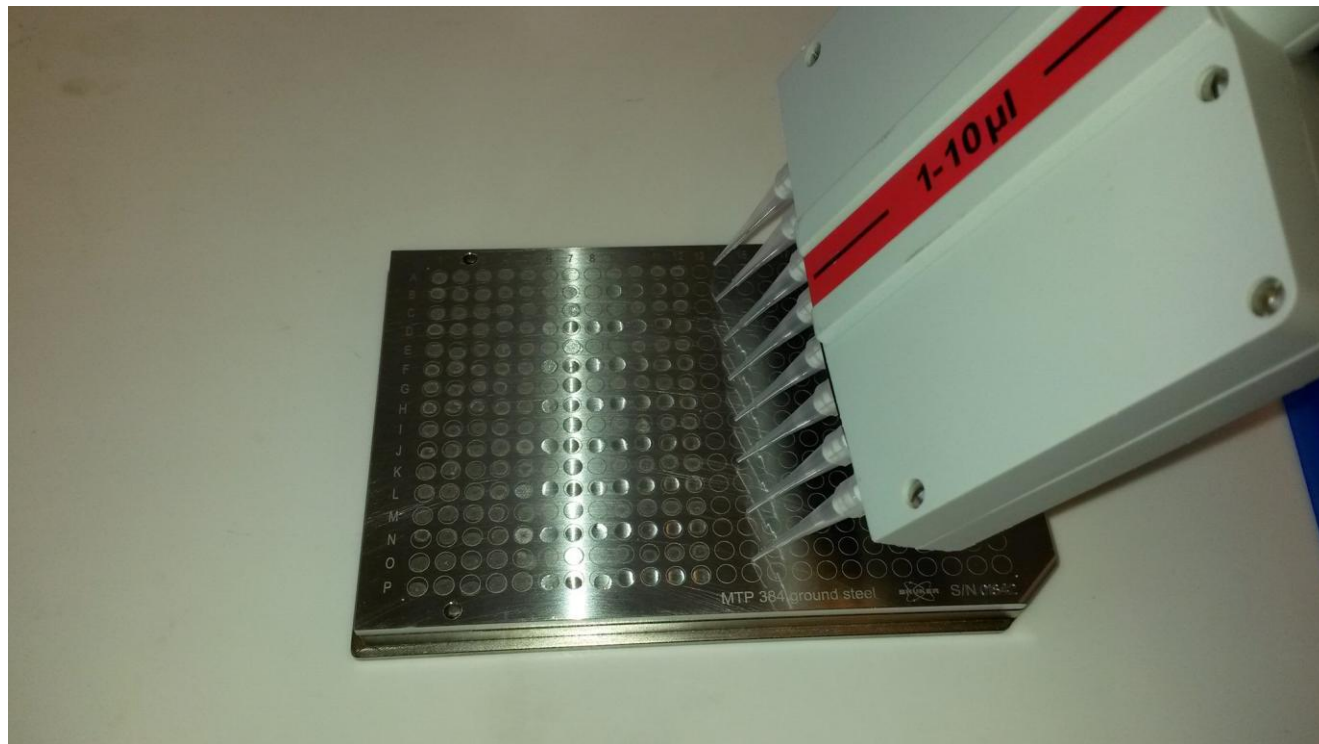




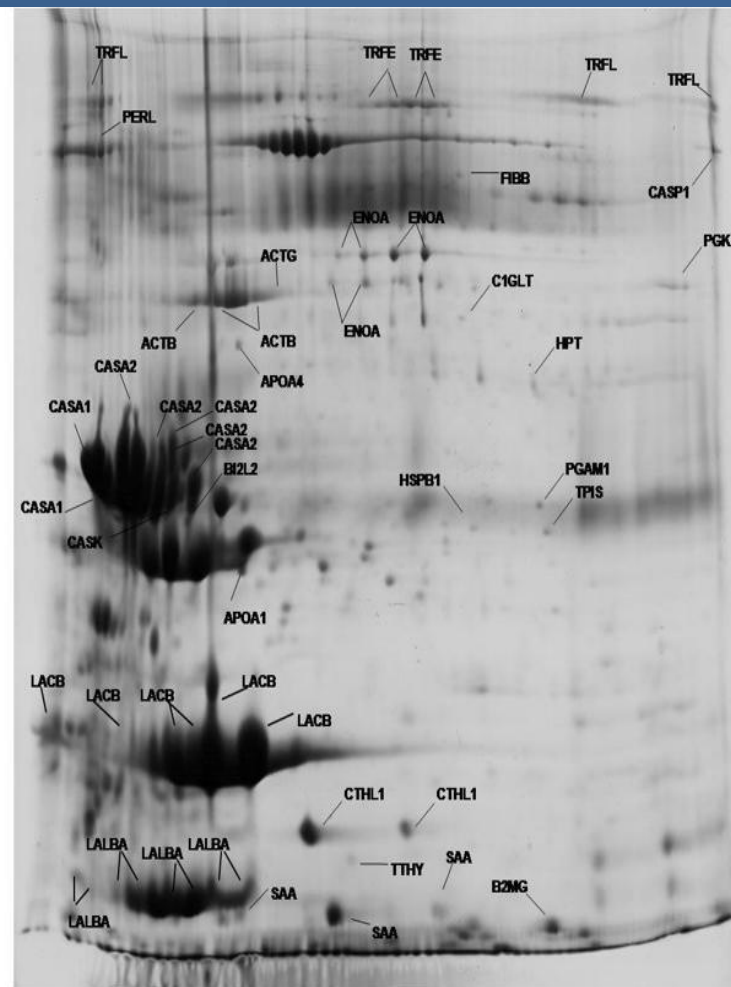
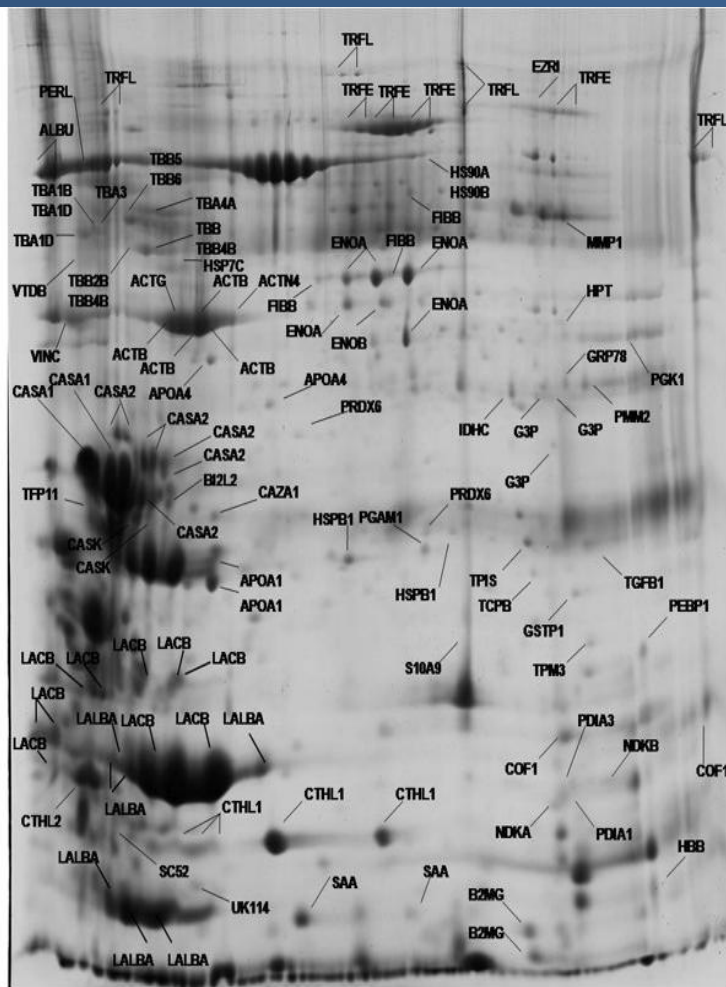
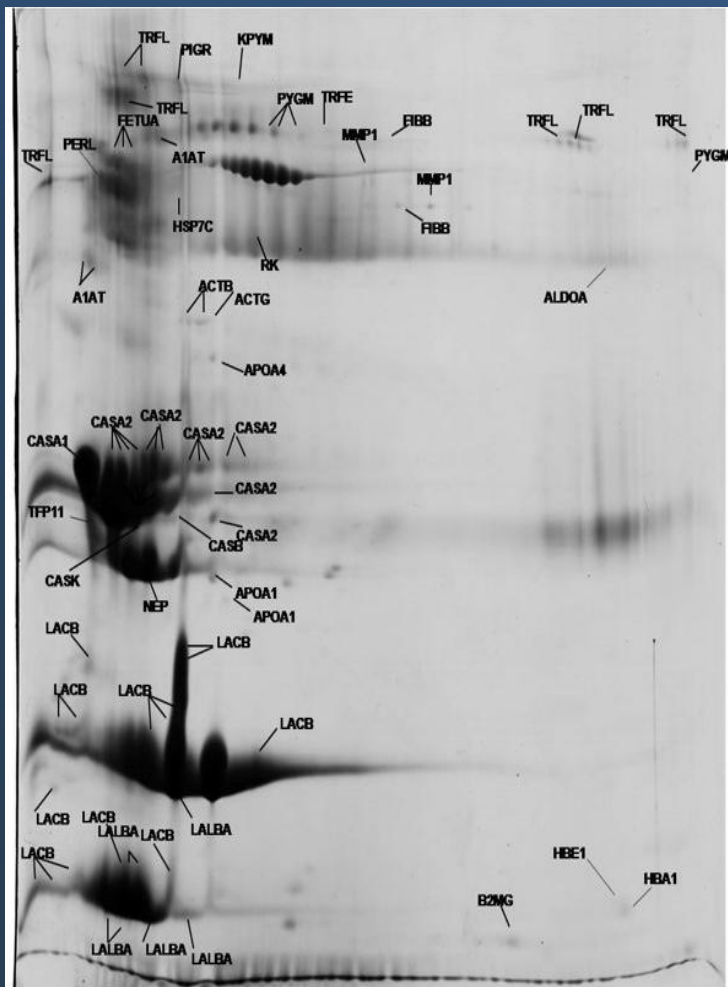




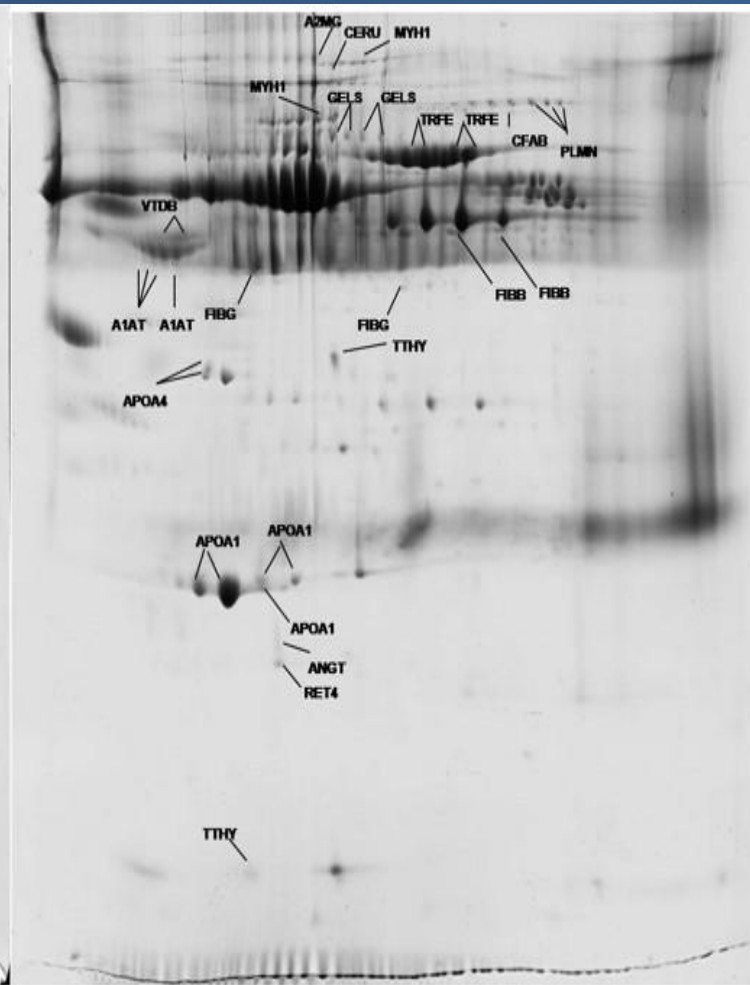
MALDI Target Plate



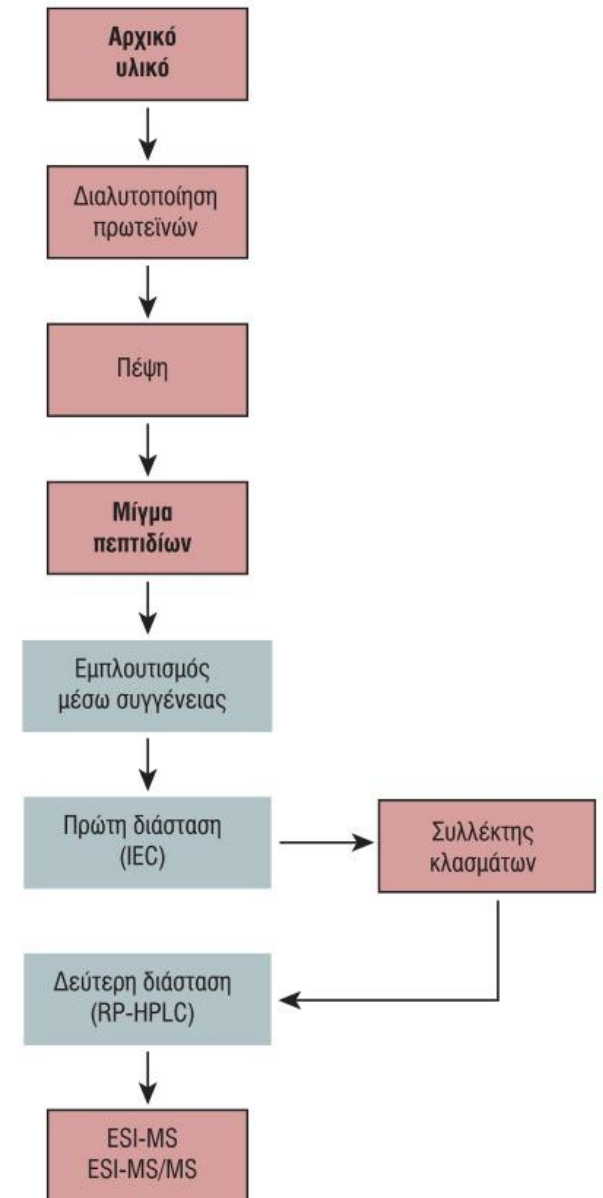
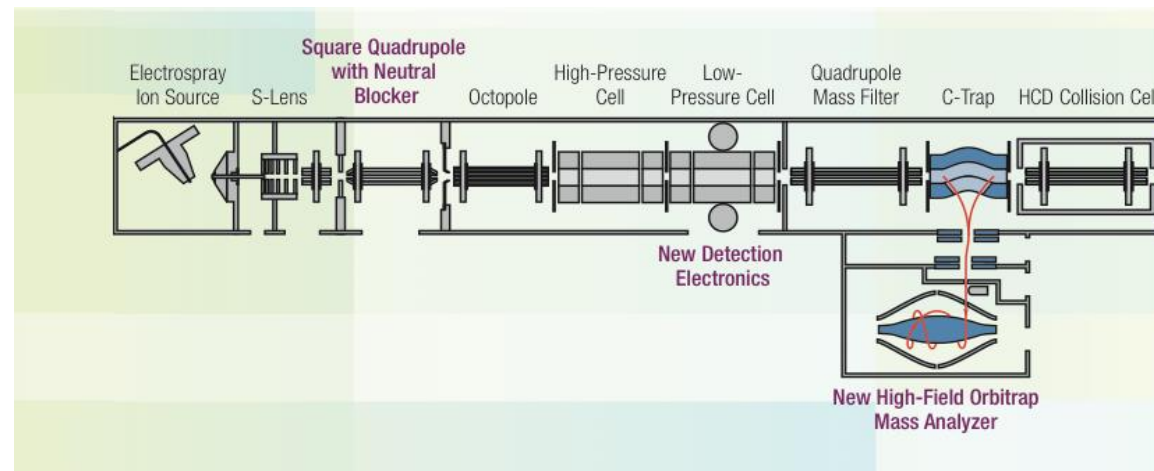
Πρωτεύωνικός χάρτης γάλακτος



Πρωτεωμικός χάρτης αίματος



Ανάλυση με Υγρή Χρωματογραφία (nano LC-MS/MS)

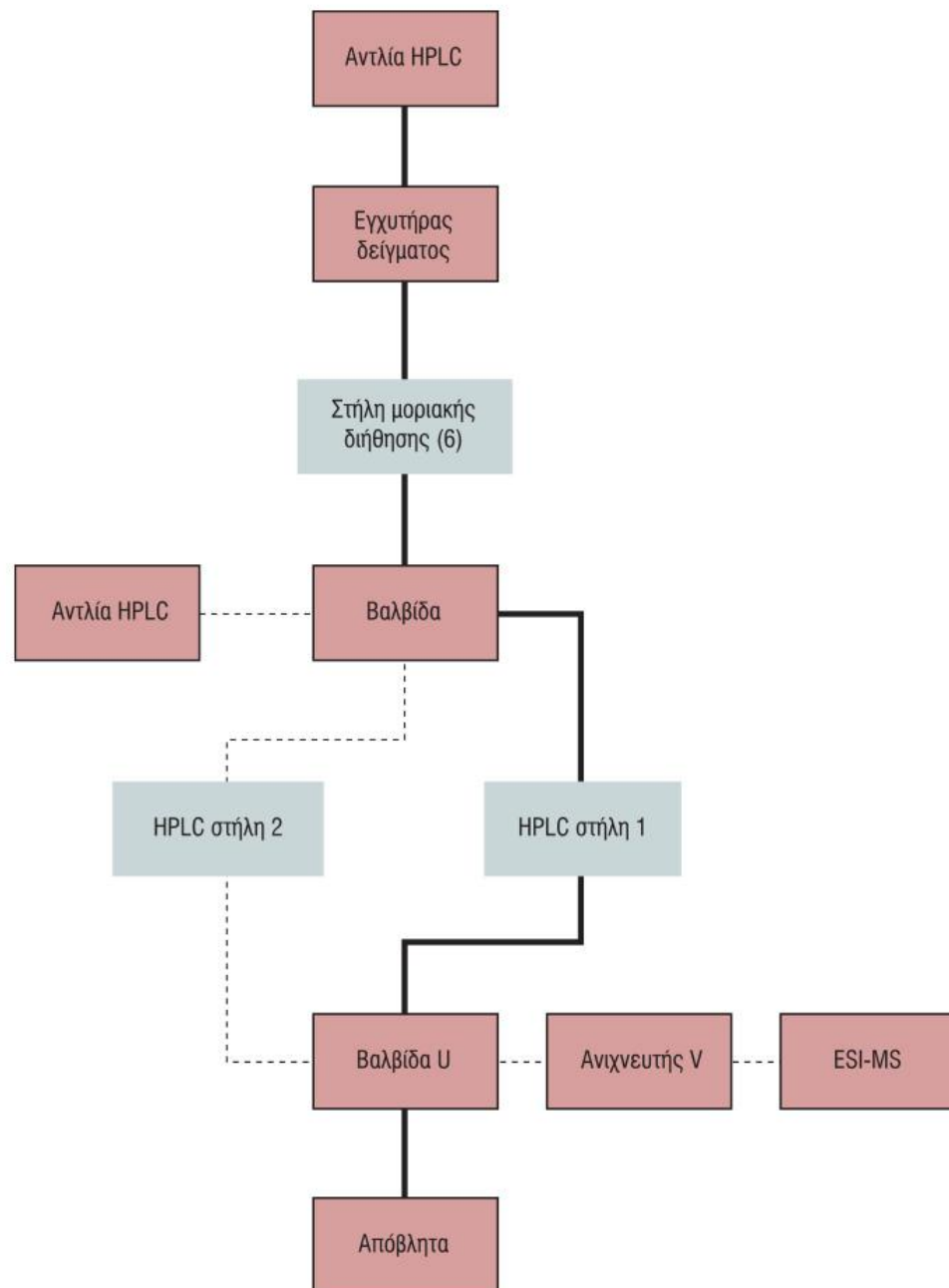


Thermo Scientific Orbitrap Elite
Hybrid Mass Spectrometer



ΕΙΚΟΝΑ 2.15

Συνεχόμενη πολυδιάστατη χρωματογραφία με αλλαγή στήλης. Σε αυτό το παράδειγμα, δύο HPLC στήλες που δουλεύουν παράλληλα προσλαμβάνουν εναλλακτικά εκλούσματα από μία συστοιχία έξι στηλών μοριακής διήθησης. Μετά την εφαρμογή του δείγματος και το διαχωρισμό με χρωματογραφία μοριακής διήθησης, κλάσματα από τη στήλη μοριακής διήθησης κατευθύνονται στη HPLC στήλη 1 χρησιμοποιώντας μία βαλβίδα τεσσάρων διαδρομών (παχειά γραμμή). Ενώ τα πεπτίδια παγιδεύονται σε αυτή τη στήλη, η HPLC στήλη 2 εκλούεται και το δείγμα κατευθύνεται στον ανιχνευτή και το συλλέκτη κλασμάτων (διακεκομμένη γραμμή). Αφού ξεπλυθεί και εξισορροπηθεί η στήλη 2, οι βαλβίδες αναστρέφονται επιτρέποντας στη στήλη 2 να φορτωθεί με το επόμενο κλάσμα από διαχωρισμό μοριακής διήθησης ενώ η στήλη 1 εκλούεται. Αυτός ο κύκλος συνεχίζεται μέχρι τα κλάσματα από τη μοριακή διήθηση εξαντληθούν. (Adapted from Opiteck GJ, Ramirez SM, Jorgenson JW & Moseley MA 3rd (1998) Anal. Biochem. 258, 349. (Κατόπιν αδείας από την Elsevier).



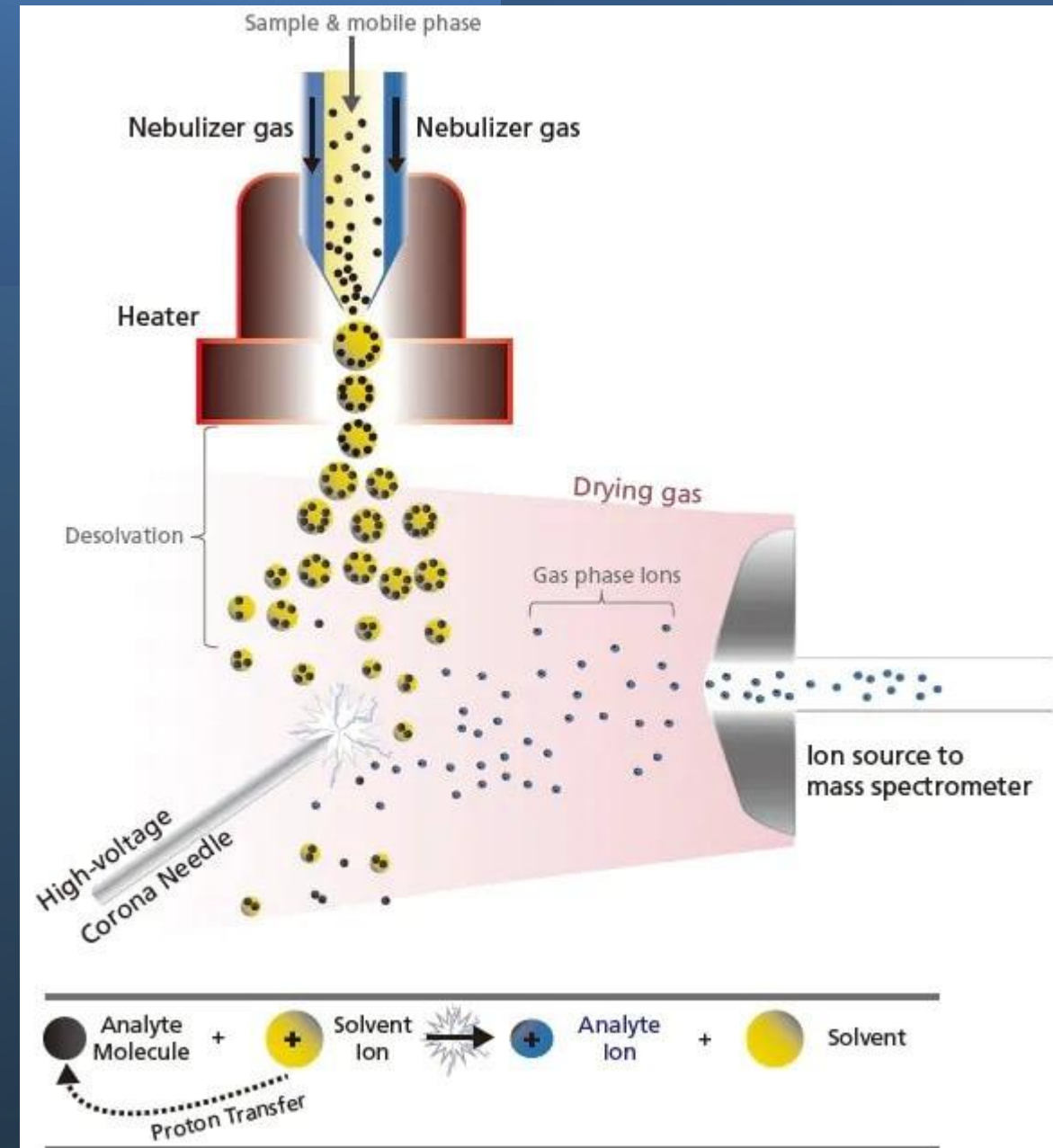
Υγρή χρωματογραφία

Μείωση πολυπλοκότητας

- Χρονική διάσταση πληροφορίας

👉 Κάθε peptide → συγκεκριμένο retention time

What happens if we skip LC?



Proteomic analysis-methodologies

- Protein separation

2D gel electrophoresis

(gel dependent)

Gel free methods (LC-MS/MS)

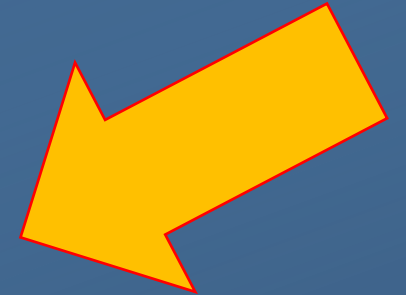
- Protein Identification

Mass spectrometry

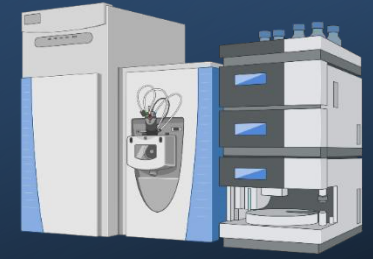
- MALDI-TOF-TOF MS
- LTQ-Orbitrap MS
- Tandem MS (Qq-TOF)
- SELDI MS

- Data analysis-Bioinformatics

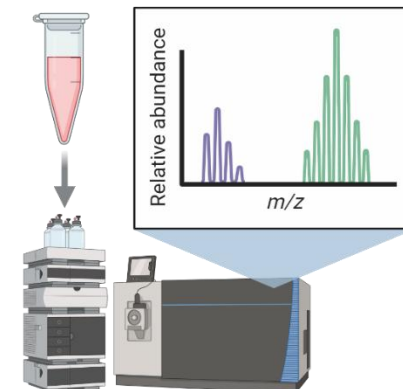
- Interpretation



Φασματομετρία μάζας – Βασικές αρχές & όργανα

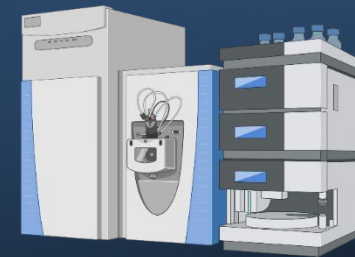


- Η φασματομετρία μάζας διαχωρίζει μόρια με βάση τη σχέση μάζα/φορτίο τους (m/z) των ιόντων στο κενό
- μπορεί να προσδιοριστεί η μάζα των μορίων με υψηλό βαθμό ακρίβειας, καθιστώντας εφικτό τον προσδιορισμό της μοριακής σύνθεσης ενός συνόλου πεπτιδίων

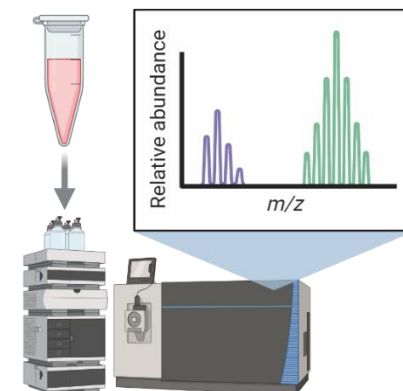


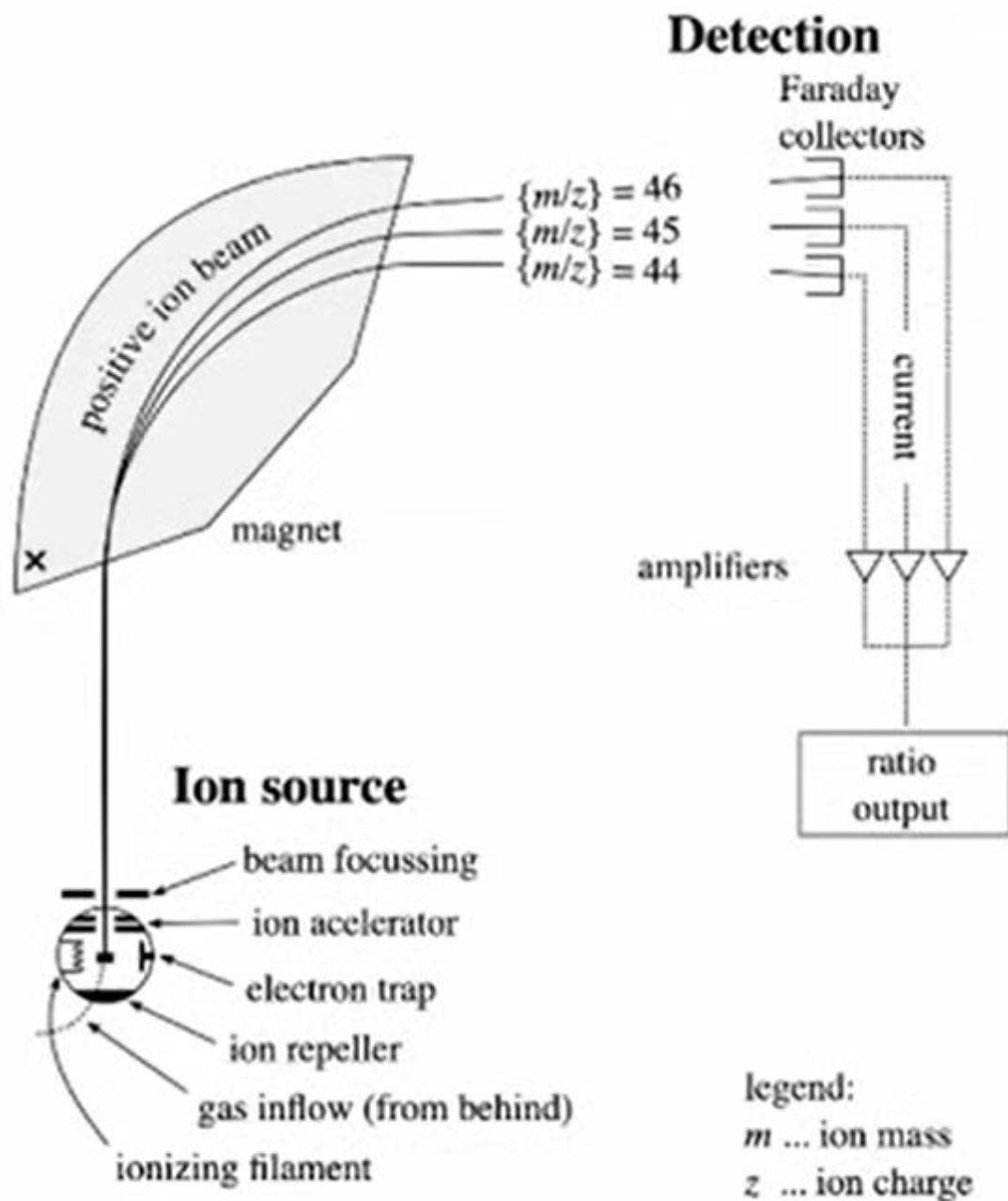
Η MS-based proteomics είναι εφαρμοσμένη sequence analysis σε θορυβώδη δεδομένα

Φασματομετρία μάζας – Βασικές αρχές & όργανα



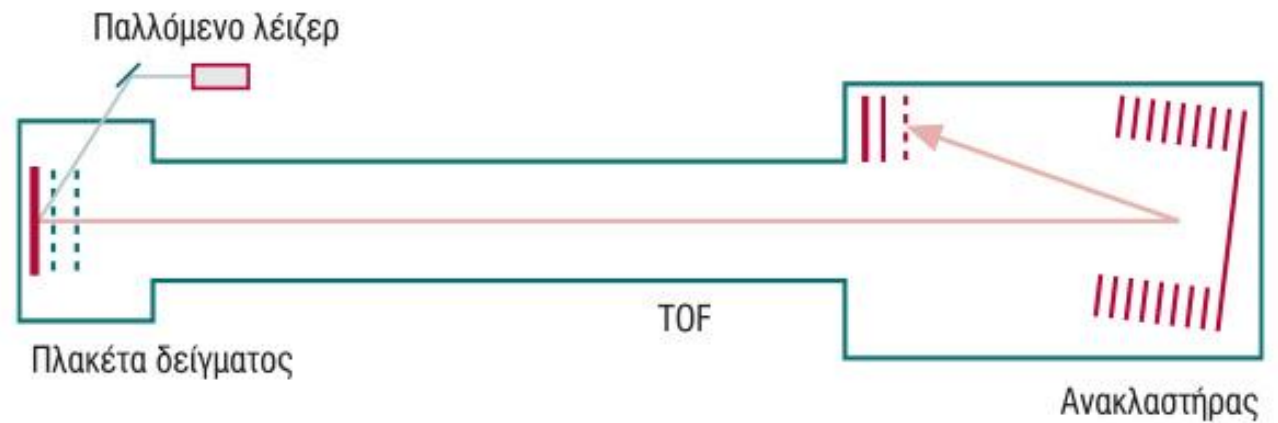
- έχουν τρία βασικά λειτουργικά στοιχεία:
 - **μια πηγή ιόντων** (μετατρέπει το αναλυόμενο δείγμα σε ιόντα στην αέρια φάση σε συνθήκες κενού)
 - **έναν αναλυτή μάζας** (ιόντα επιταχύνονται σε ένα ηλεκτρικό πεδίο προς τον αναλυτή, ο οποίος τα διαχωρίζει ανάλογα με το λόγο m/z που έχει το καθένα καθώς κατευθύνονται στον ανιχνευτή)
 - **ανιχνευτή ιόντων** (καταγράφει το σήμα κάθε ιόντος ξεχωριστά).





MALDI-TOF (MS)

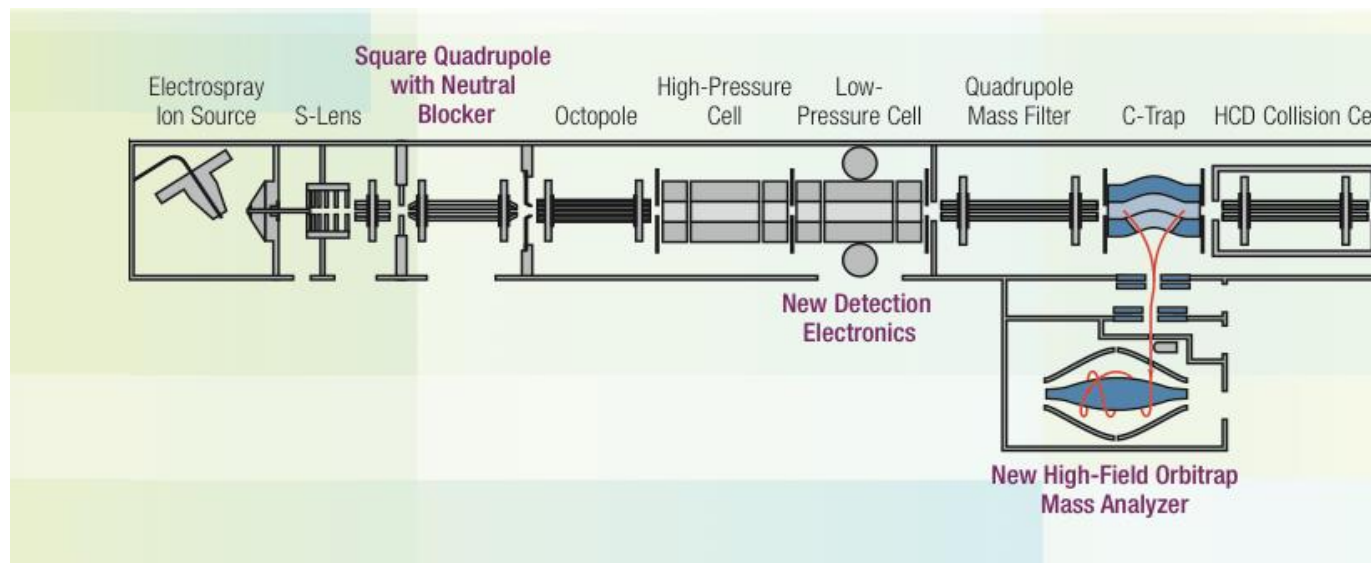
Το όνομά του είναι συντόμευση της λειτουργίας του:
Matrix Assisted Laser Desorption/Ionization Time Of Flight



MALDI TOF-TOF MS-MS



LTD-Orbitrap MS



Thermo Scientific Orbitrap Elite
Hybrid Mass Spectrometer

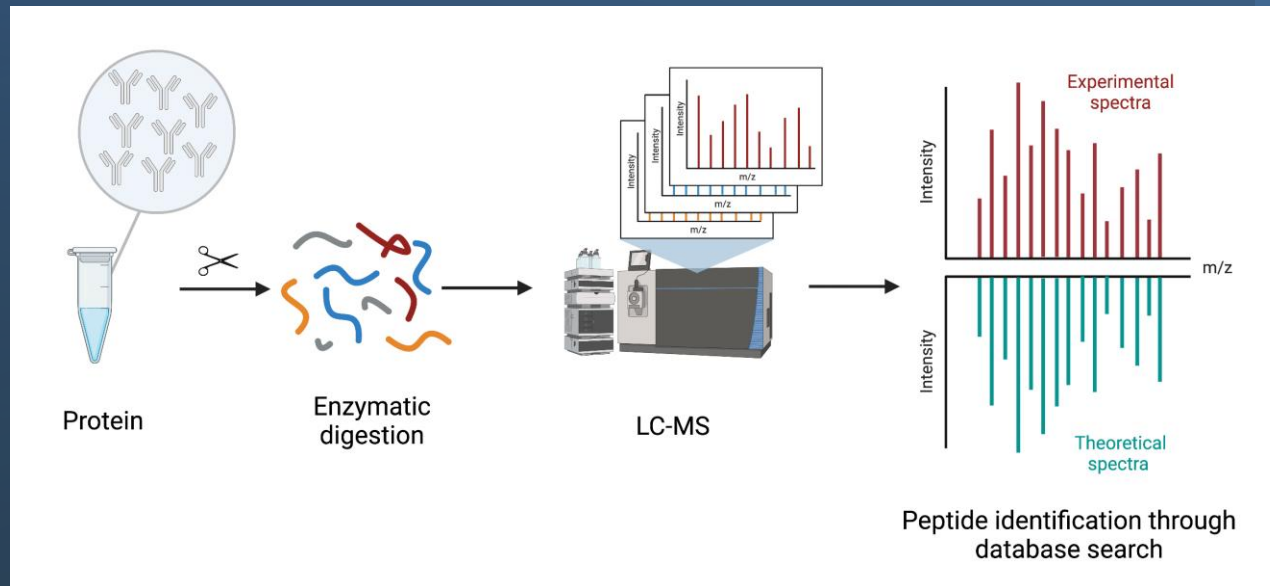


Πώς φτάνουμε
στην
ταυτοποίηση...

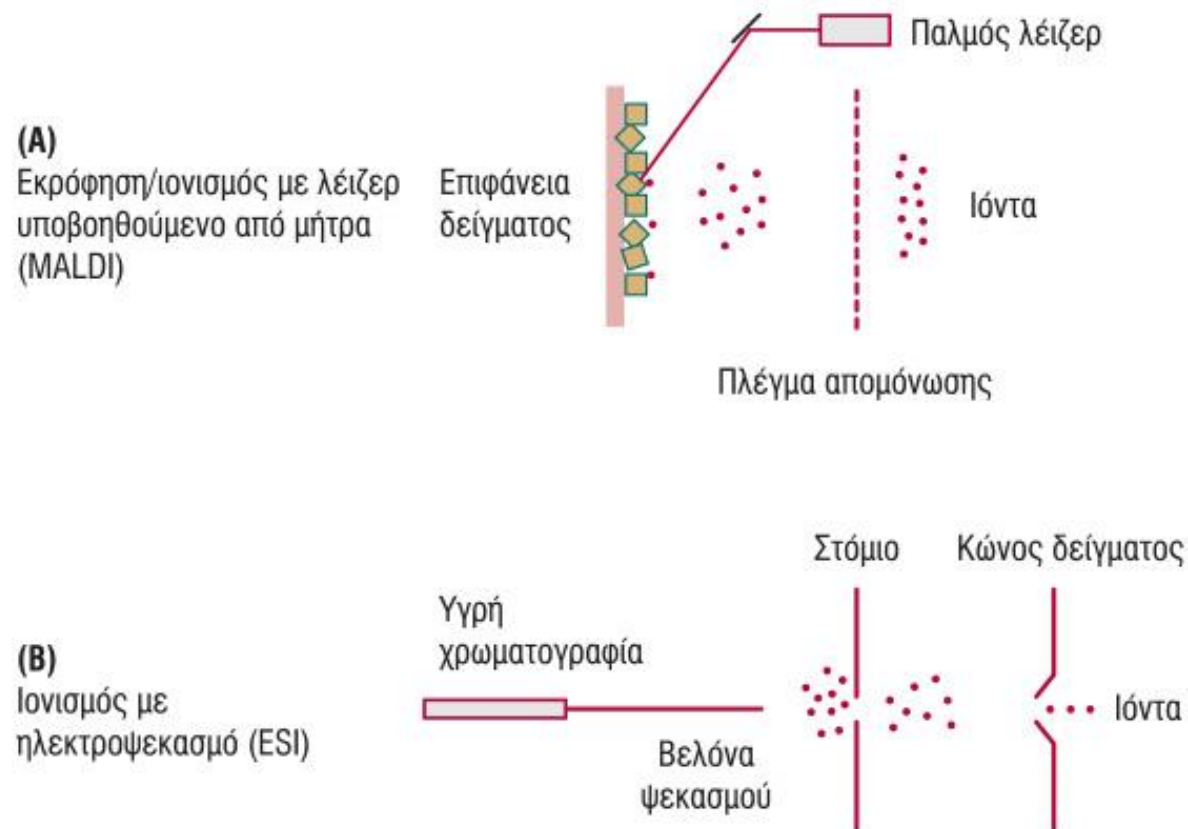


Pipeline

Protein → peptides → spectra → sequences → protein



Ionization

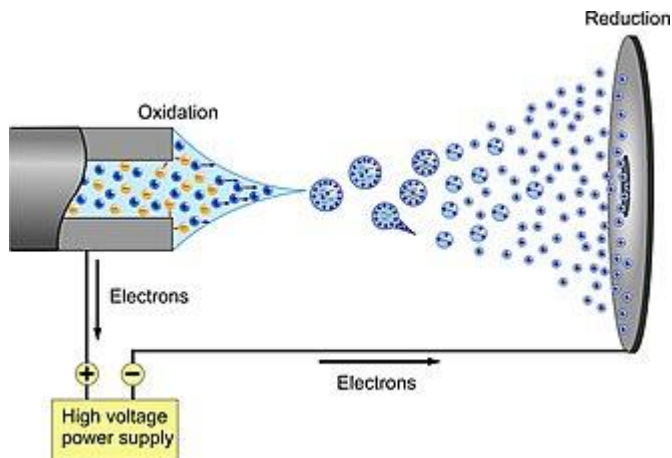


ΕΙΚΟΝΑ 3.7

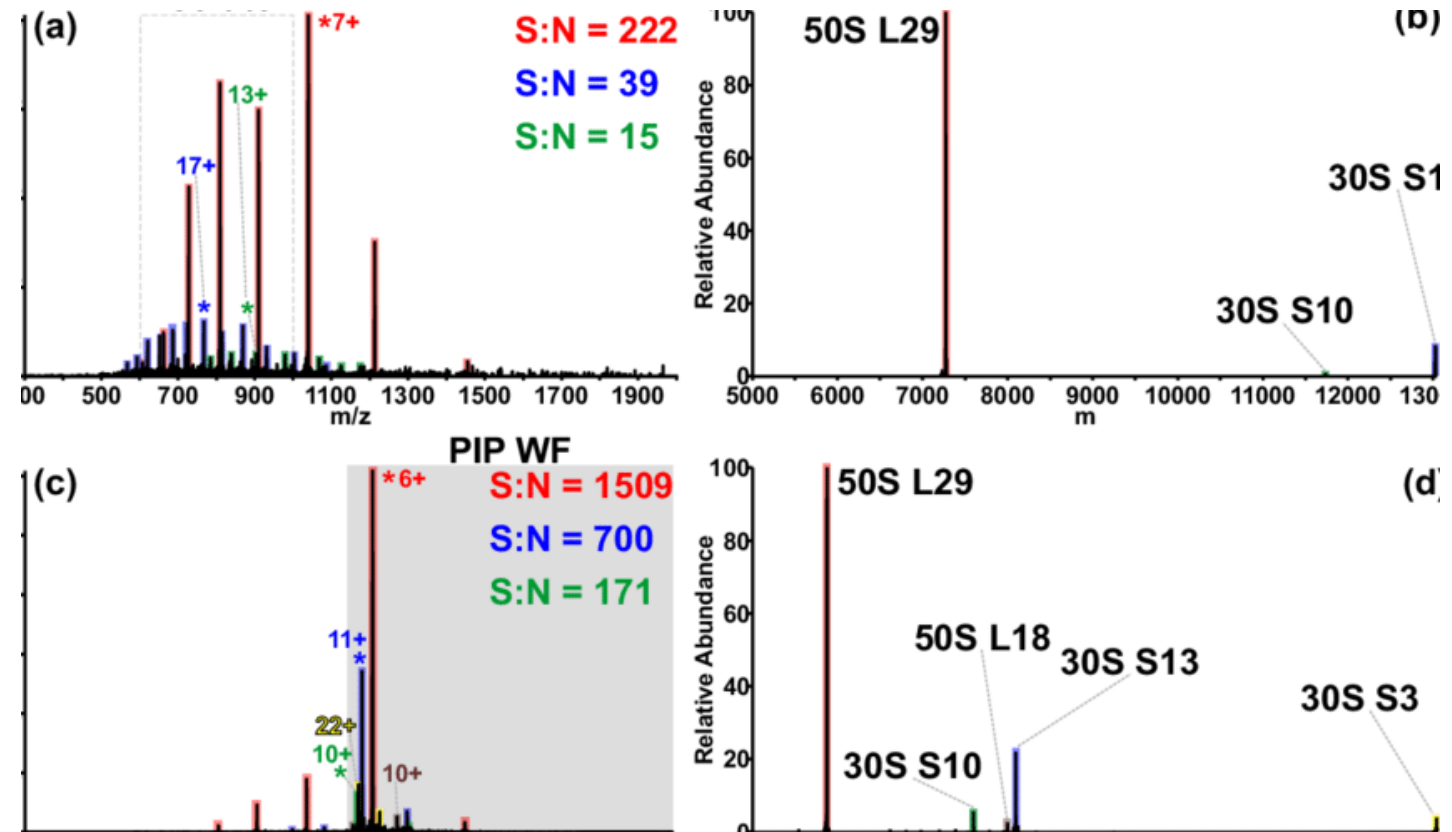
Μέθοδοι ήπιου ιονισμού στην πρωτεωμική.

(A) Η μέθοδος MALDI περιλαμβάνει θερμαινόμενους κρυστάλλους της αναλυόμενης ουσίας σε μια πλάκα δείγματος με χρήση παλμών λέιζερ. **(B)** Η ESI περιλαμβάνει την εξαναγκασμένη διέλευση του δείγματος μέσω μιας λεπτής βελόνας που δημιουργεί ένα λεπτό εκνέφωμα (*spray*) ιόντων.

Ionization & MS¹



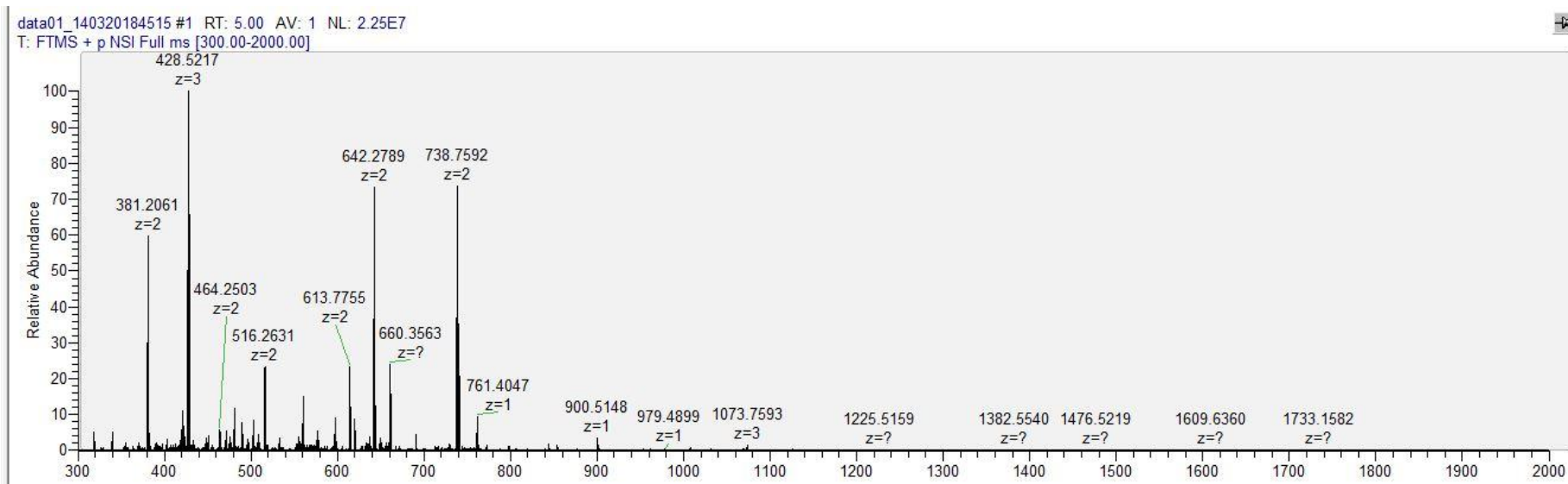
- m/z
- charge states
- isotopic envelopes
- ⚠ Δεν δίνει αλληλουχία
- 👉 Μόνο υποψήφια peptide features



👉 *We still don't know the sequence.*

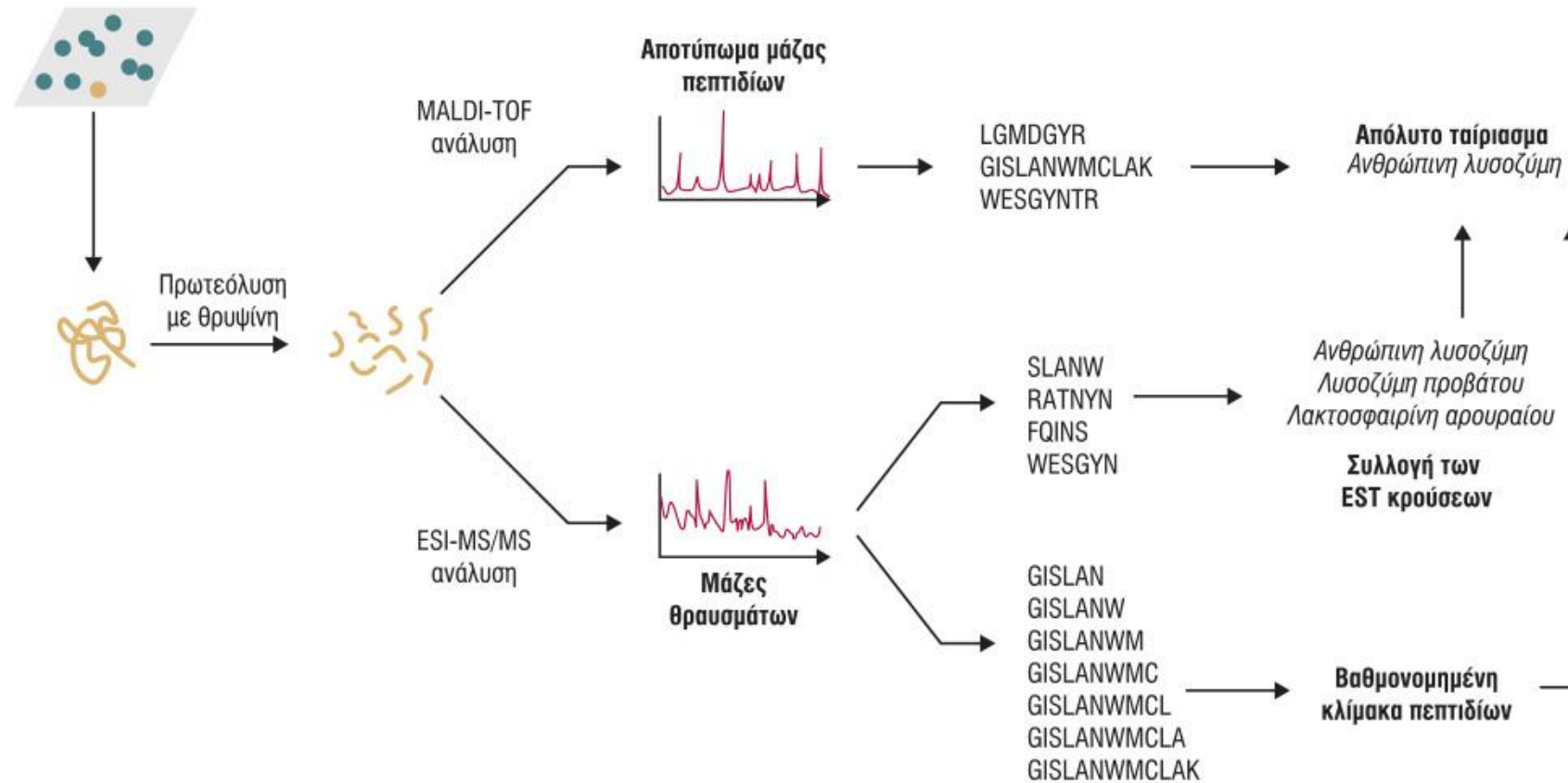
Φάσματα μάζας

Γραφική παράσταση της έντασης του ιοντικού σήματος ως συνάρτηση του λόγου μάζας προς φορτίο (m/z) των ιόντων που ανιχνεύονται



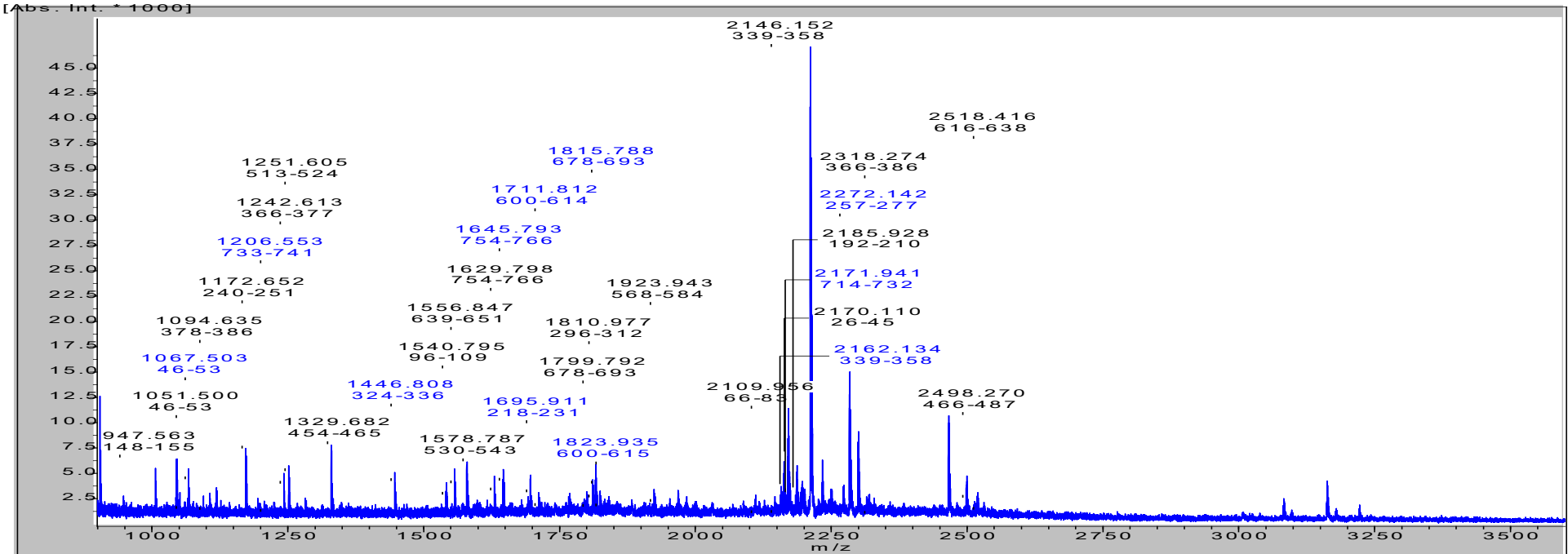
- **x-axis:** mass-to-charge ratio (m/z)
- **y-axis:** ion intensity (relative abundance)

Each **peak** represents an ion detected by the instrument



Η πρώτη ευρέως χρησιμοποιούμενη μέθοδος για την ταυτοποίηση πρωτεϊνών με τη χρήση φασματομετρίας μάζας ήταν το **δακτυλικό αποτύπωμα πεπτιδικών μαζών** (*peptide mass fingerprinting, PMF*), η οποία αφορά την ταυτοποίηση πρωτεϊνών μέσω της χρήσης δεδομένων από τις μάζες των ακέραιων πεπτιδίων

Peak Identification



Peptide Mass Fingerprint (PMF)



Protein Identification

Πρόγραμμα	URL
Αποτύπωμα πεπτιδικών μαζών	
Mascot	http://www.matrixscience.com/search_form_select.html
MS-Fit	http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msfitstandard
ProFound	http://prowl.rockefeller.edu/prowl-cgi/profound.exe
MassSearch	http://www.cbrg.ethz.ch/services/MassSearch_new

MS/MS φάσματα

- **κατάτμηση ακέραιων πεπτιδίων** και εξάγοντας περισσότερα δεδομένα από τις μάζες των τυχαίως παραγόμενων ιοντικών θραυσμάτων.
- Precursor ion: Μια πρόδρομη ιοντική μορφή (precursor ion) είναι ένα ιόν που ανιχνεύεται στο φάσμα MS1, το οποίο απομονώνεται και υφίσταται κατακερματισμό ώστε να παραχθεί ένα φάσμα MS2 (MS/MS).
- Το MS/MS φάσμα με άγνωστη αλληλουχία μπορεί τότε να συγκριθεί με θεωρητικά φάσματα των τμημάτων που προκύπτουν από πεπτίδια σε όποια βάση επιθυμούμε.

Γιατί η MS¹ πληροφορία δεν επαρκεί για sequence identification, ενώ η MS/MS είναι κρίσιμη;

Γιατί η MS¹ πληροφορία δεν επαρκεί για sequence identification, ενώ η MS/MS είναι κρίσιμη;

MS¹ (Precursor Mass Information)

- Παρέχει μόνο m/z , charge state και ισοτοπική κατανομή
- Πολλά διαφορετικά πεπτίδια μπορεί να έχουν:
 - ίδια ή σχεδόν ίδια μάζα
 - ισοβαρή ή ισομερή αμινοξέα (π.χ. Leu / Ile)
- Δεν περιέχει πληροφορία για τη **σειρά** των αμινοξέων

👉 Η MS¹ πληροφορία είναι **αναγκαία αλλά όχι επαρκής** για ταυτοποίηση αλληλουχίας

MS/MS (Fragmentation Information)

- Η ελεγχόμενη θραύση του πεπτιδίου παράγει **σειρές θραυσμάτων (b/y ions)**
- Οι διαφορές m/z μεταξύ διαδοχικών ιόντων αντιστοιχούν σε:
 - συγκεκριμένα αμινοξέα
 - τμήματα της πεπτιδικής αλυσίδας
- Το MS/MS φάσμα κωδικοποιεί **πληροφορία αλληλουχίας**

👉 Το MS/MS μετατρέπει τη φασματομετρία μάζας σε **πρόβλημα sequence matching**

Η MS¹ περιορίζει τον χώρο αναζήτησης, ενώ η MS/MS επιτρέπει την ανακατασκευή της αλληλουχίας

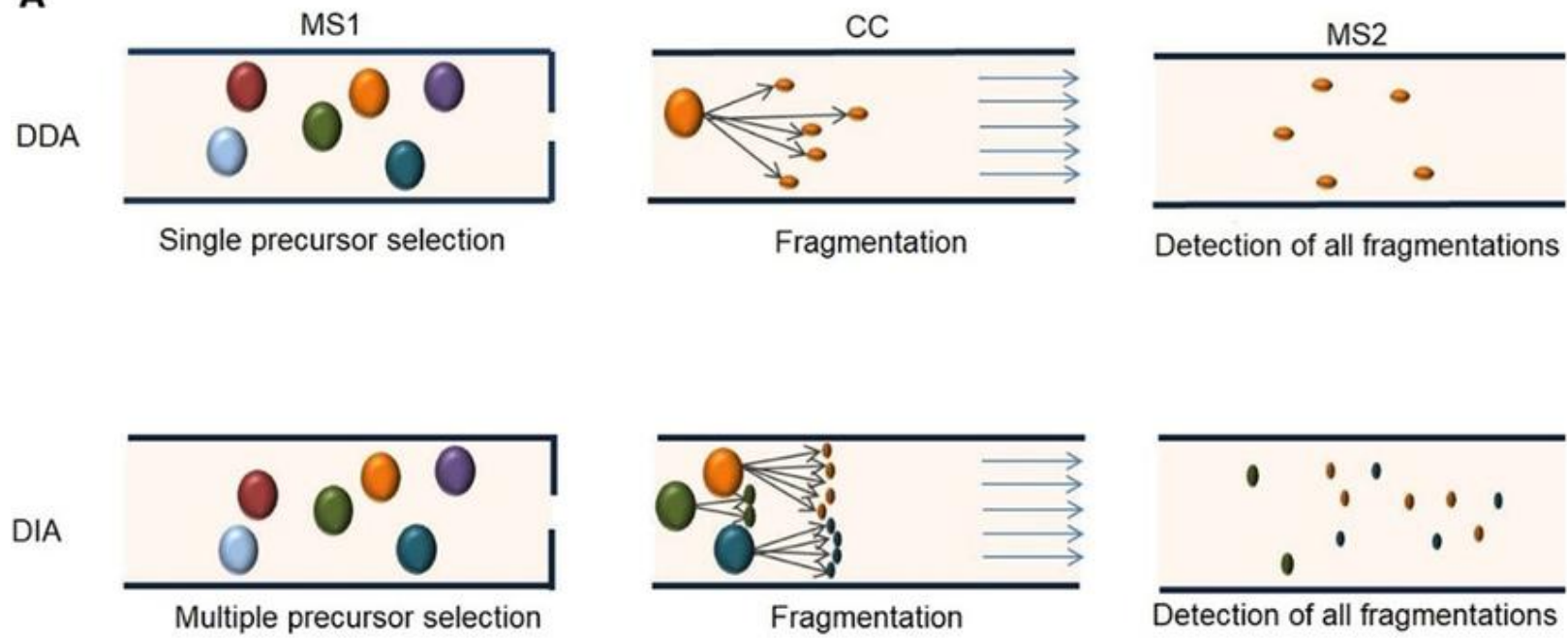
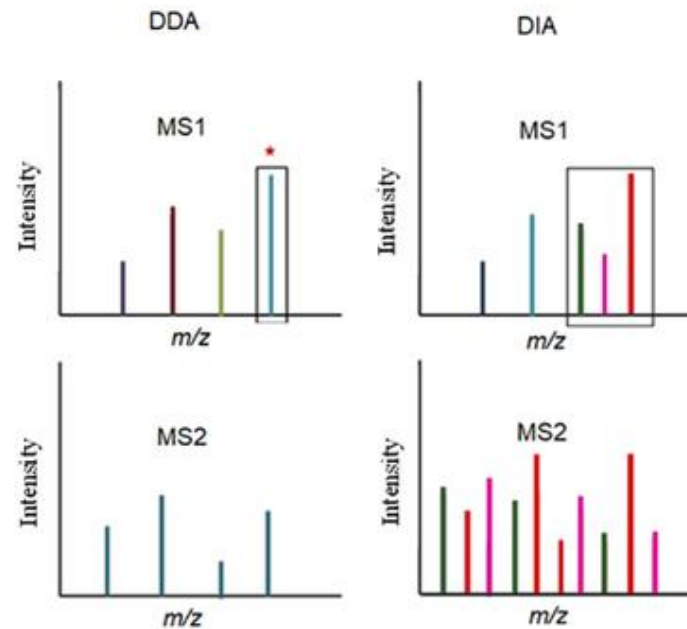
Πώς επιλέγεται το Precursor ion?

- One ion (DDA)
- A range of ions (DIA)

Χαρακτηριστικό	DDA (Data-Dependent Acquisition)	DIA (Data-Independent Acquisition)
Επιλογή προδρόμων ιόντων	Επιλογή των Top N πιο έντονων ιόντων	Όλα τα ιόντα εντός προκαθορισμένων παραθύρων m/z
Φάσματα MS/MS	Καθαρά, χαμηλής πολυπλοκότητας	Πολύπλοκα, πολυπλεγμένα
Κάλυψη πρωτεώματος	Μέτρια	Υψηλή
Αναπαραγωγικότητα	Χαμηλότερη	Υψηλότερη
Ελλείπουσες τιμές	Συχνές	Σπάνιες
Ευαισθησία σε χαμηλής αφθονίας πεπτίδια	Περιορισμένη	Βελτιωμένη
Ποσοτική συνέπεια	Μέτρια	Υψηλή
Ανάλυση δεδομένων	Απλούστερη	Υπολογιστικά απαιτητική
Φασματικές βιβλιοθήκες	Προαιρετικές	Συχνά απαραίτητες
Τυπική χρήση	Ανακάλυψη, ταυτοποίηση	Ποσοτικοποίηση, μελέτες μεγάλης κλίμακας

•DDA → *Who is there?*

•DIA → *How much is there, reproducibly?*

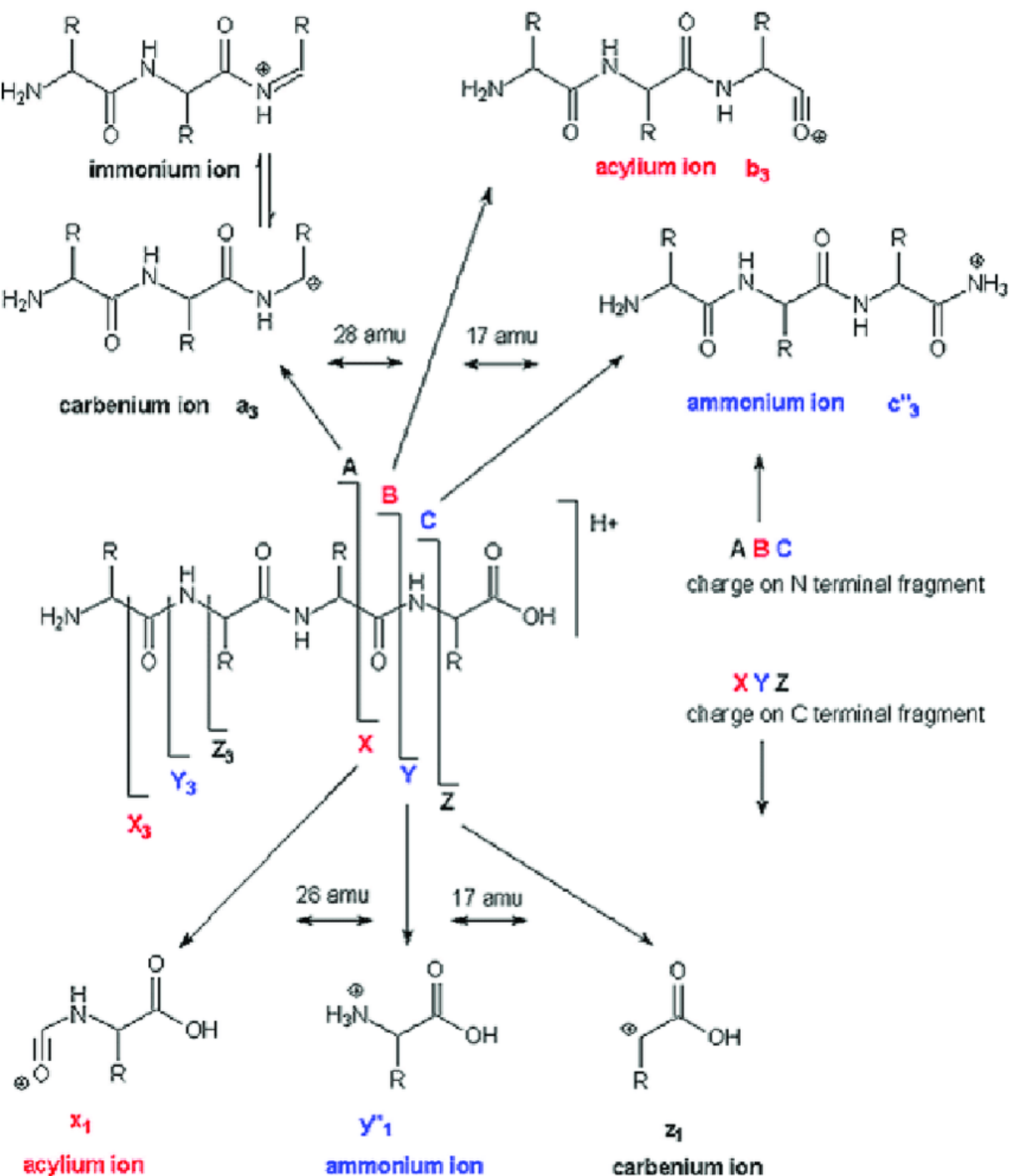
A**B**

Comparison of MS1 and MS2 Spectra

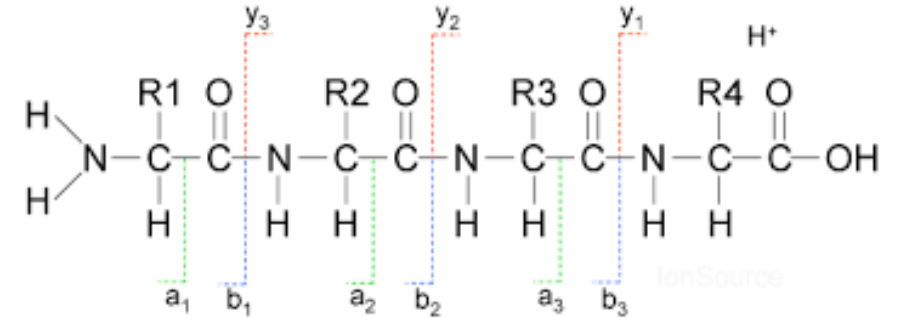
Feature	MS1 Spectrum	MS2 Spectrum (MS/MS)
What is measured	Intact (precursor) ions	Fragment ions of a selected precursor
Acquisition stage	First stage of MS	Second stage after fragmentation
m/z represents	Whole peptides (or metabolites)	Peptide fragments (b/y ions in proteomics)
Fragmentation	No	Yes (CID, HCD, ETD, etc.)
Spectral complexity	Lower	Higher
Main purpose	Quantification, precursor selection	Identification, sequence inference
Information obtained	Mass and abundance of precursors	Amino acid sequence information
Used for	Feature detection, DIA windows, isotope patterns	Database searching, peptide-spectrum matching
Typical output	Chromatographic peaks	Fragment ion patterns
Example question answered	How much is present?	What is it?

- MS1 → Detect and quantify intact molecules
- MS2 → Identify molecules by their fragments

- Fragmentation produces **smaller ions** (b-ions, y-ions)
- These are called **fragment ions** (or product ions)



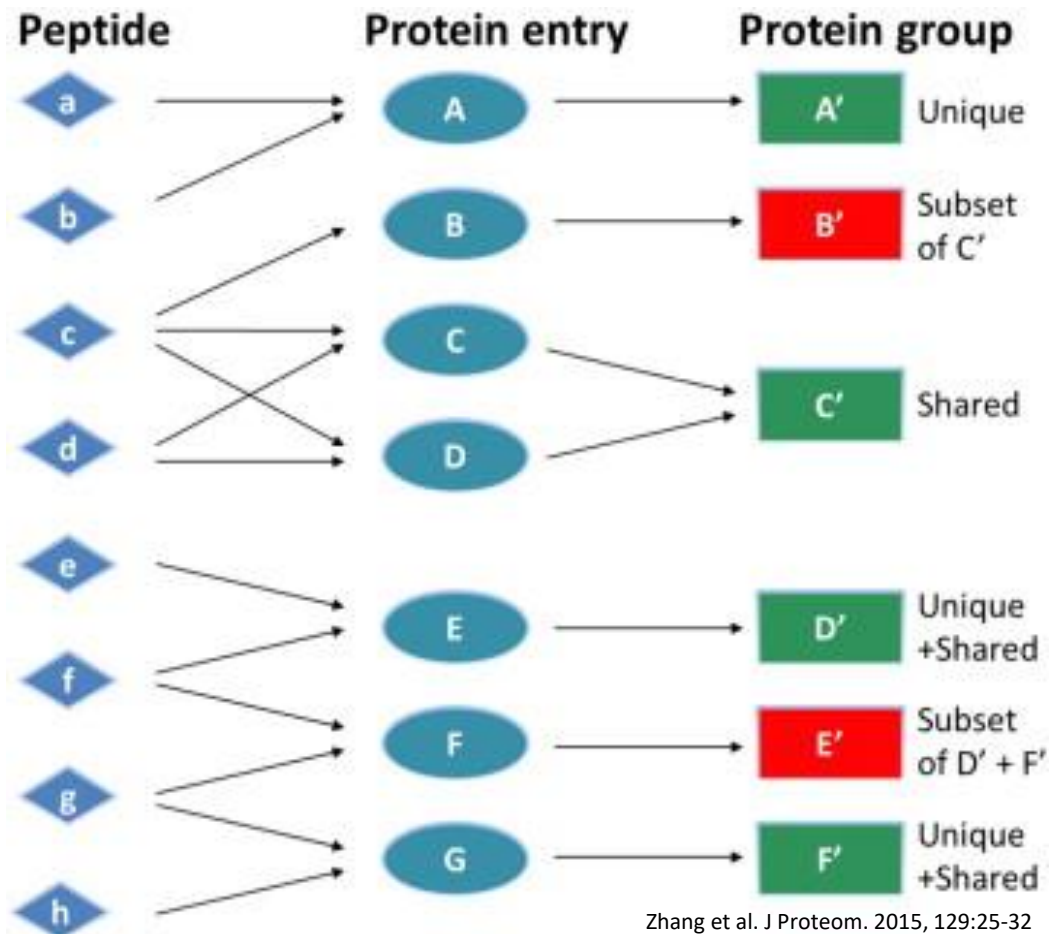
Tandem MS (MS/MS): Where Sequence Appears



- **Επιλογή precursor** (ion detected in the MS1 spectrum that is isolated and fragmented to generate an MS2 spectrum)
- Θραύση (CID / HCD / ETD)
- Παραγωγή **b/y ions**

👉 Το MS/MS φάσμα περιέχει πληροφορία αλληλουχίας
m/z differences → amino acids

From Spectrum to Sequence

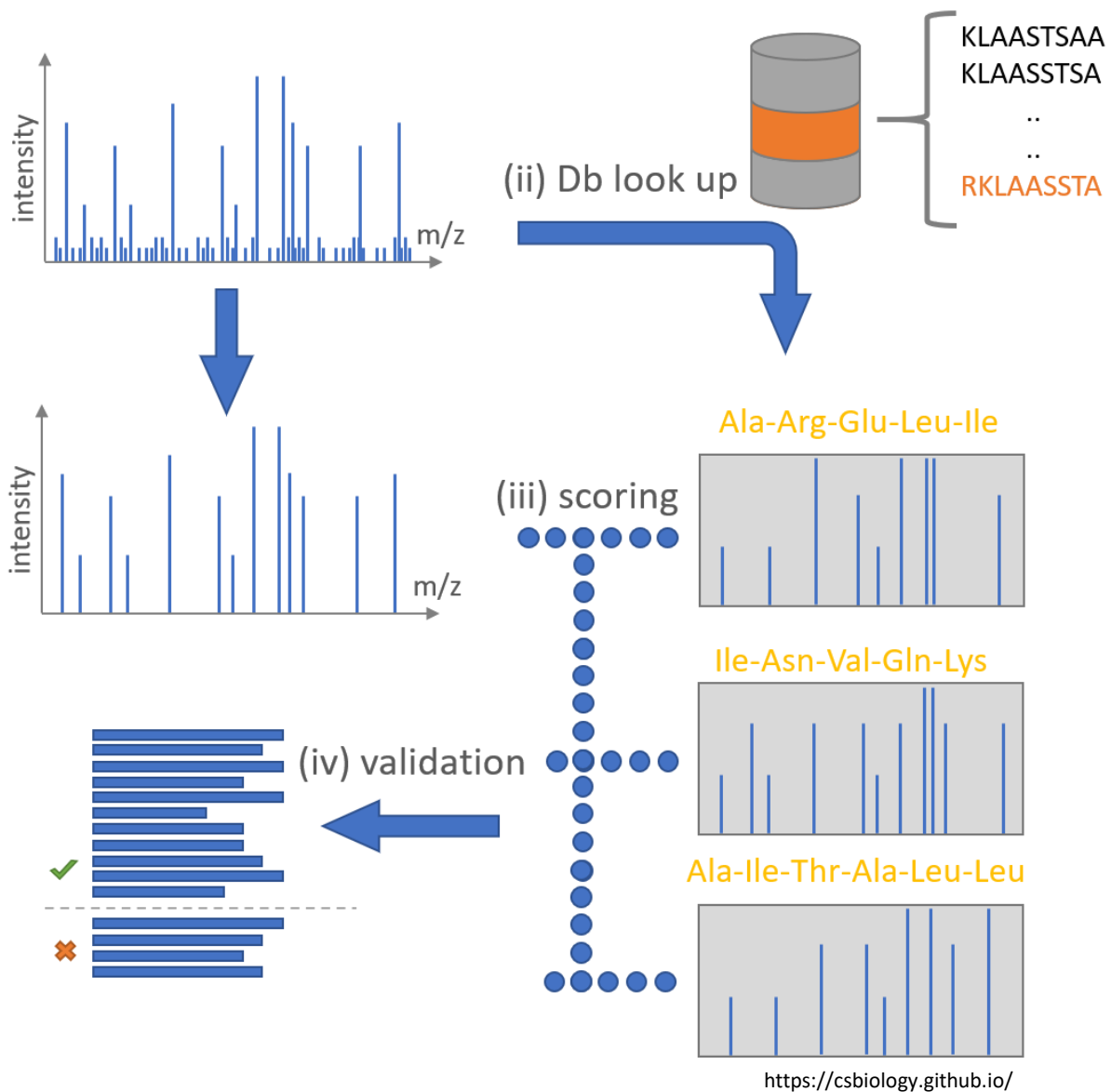


Δυσκολίες

- Shared peptides
 - Isoforms
 - Redundancy databases
- 👉 Protein ID ≠ απλή ένωση peptides

Protein Inference Problem

From Spectrum to Sequence



Υπολογιστικό πρόβλημα

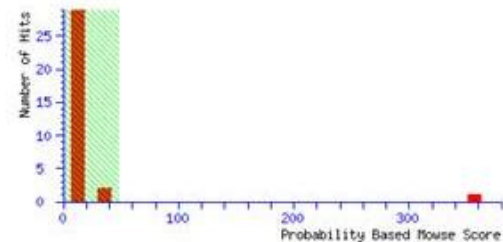
- Experimental spectrum
- Theoretical spectra από database peptides
- Scoring & matching

👉 **Sequence alignment υπό θόρυβο**

Database Searching & Validation

Probability Based Mowse Score

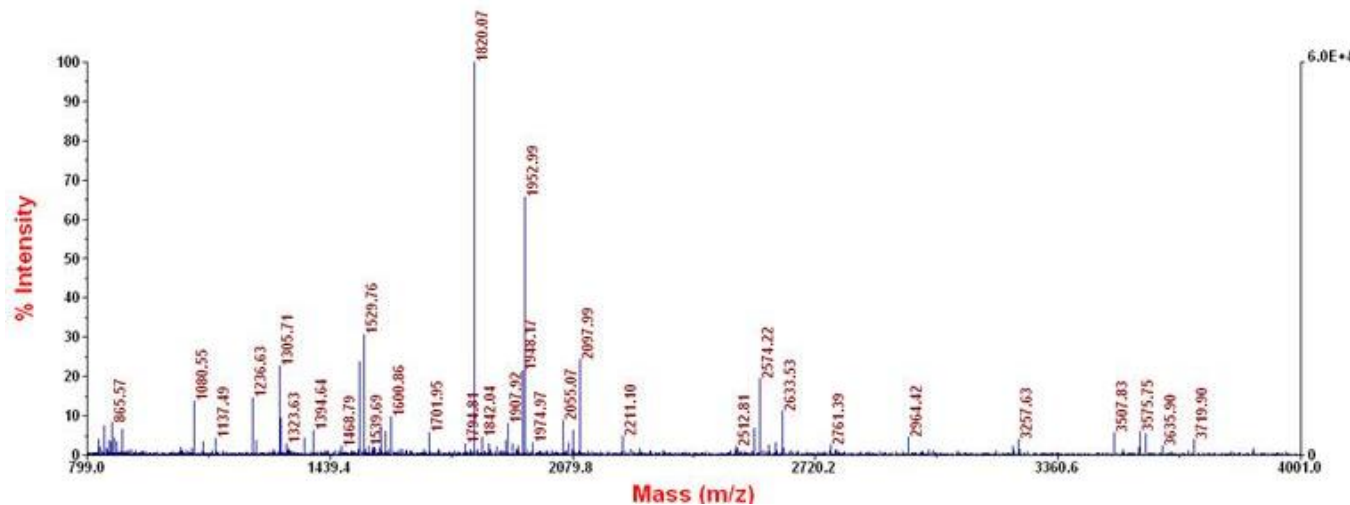
Ions score is $-10 \cdot \log(P)$, where P is the probability that the observed match is a random event. Protein scores greater than 48 are significant ($p < 0.05$).



Number of mass values searched: 44
Number of mass values matched: 36
Sequence Coverage: 65%

Matched peptides shown in Bold Red

1	AKSIIVNDEA	RPALERGRDI	LARAVAVTLG	PRGRNVVLEK	WFGSPQIIND
51	GITIASKRIEL	EDVENTOVVS	LIRQAASKTN	DVAGDOTTFA	TVLAAIAIVKE
101	GLRNVAAGAN	FISLGRGIDK	ATDFLVARIK	ERAQPVODSK	AIAQVGAISA
151	GNDEEVGQMI	ANAMDKVQGE	GVISLEEGKS	HTTELRIEEO	MRFDKVIYSP
201	YFVTDAERME	AVLEDPRIIL	TDKRINLVQD	LVPILQVAR	QWPLLIIEAE
251	DIEKKALATL	VVNRIRGVIN	VAAVKAPGFG	DRKKMLEDI	ATLTGGQVIS
301	EDAGLKLESA	TVDSLGSAAR	INITKDNITII	VAEGNEAAVK	SRCEQIRRIQI
351	EETDSSYDKE	KLQERLAKLA	GGVAVIRVGA	ATSTEGSDRK	LRLRDAINAT
401	KAAVEEGIVF	GGDTTLAKLA	PQLEDNATON	LKDELTGAL	IVARALPAFL
451	KRIARNAQGN	GAVISERVKE	KEFNVGYNAA	SLEYVDMIAA	GIVDPARVIR
501	SALQMAASIA	GNVLTTEICV	VDRPEKEKAP	AGAPOGDFDY	



👉 Protein ID is probabilistic

- Peptide–spectrum matches (PSMs) are assigned **statistical scores**
- Protein IDs are reported with **confidence measures** (e.g. *p-values*, *FDR*)
- Multiple proteins can share peptides → **ambiguity**
- Final identification reflects the **most likely explanation** of the data

ΠΙΝΑΚΑΣ 3.2 Βασικά εργαλεία στην ανάλυση δεδομένων φασματομετρίας μάζας

Πρόγραμμα	URL
Αποτύπωμα πεπτιδικών μαζών	
Mascot	http://www.matrixscience.com/search_form_select.html
MS-Fit	http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msfitstandard
ProFound	http://prowl.rockefeller.edu/prowl-cgi/profound.exe
MassSearch	http://www.cbrg.ethz.ch/services/MassSearch_new
Αναζητήσεις σε βάση δεδομένων MS/MS	
Sequest	http://fields.scripps.edu/sequest/
Mascot	http://www.matrixscience.com/search_form_select.html
ProteinProspector	http://prospector.ucsf.edu
ProbiD	http://tools.proteomecenter.org/wiki/index.php?title=Software:ProbiD
X! Tandem	http://www.thegpm.org
SpectrumMill	http://www.chem.agilent.com/
Phoenyx	http://www.genebio.com/products/phenyx/
VEMS	http://yass.sdu.dk/
ProteinPilot	http://www.absciex.com
MyriMatch	http://fenchurch.mc.vanderbilt.edu/software.php
PepSplice	http://www.ti.inf.ethz.ch/pw/software/pepsplice/
RAId_DbS	http://www.ncbi.nlm.nih.gov/CBBresearch/qmbp/raid_dbs/
Mass Matrix	http://www.massmatrix.net/mm-cgi/home.py
Ιχνηθέτες αλληλουχίας/υβριδικές μέθοδοι	
InsPecT	http://proteomics.ucsd.edu/Software/Inspect.html
Popitam	http://code.google.com/p/popitam/
TagRecon	http://fenchurch.mc.vanderbilt.edu/software.php
ByOnic	http://proteinmetrics.com/software-products/byonic-software/
Spectral Networks	http://proteomics.ucsd.edu/Software/SpectralNetworks.html
MODi	http://www.massmatrix.net/mm-cgi/home.py
MS-Tag	http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=mstagstandard

<i>Αλληλούχηση de novo</i>	
Lutefisk	http://www.hairyfatguy.com/Lutefisk
PepNovo	http://proteomics.ucsd.edu/Software/PepNovo.html
PEAKS	http://www.bioinformaticssolutions.com
Sequit	http://www.sequit.org/
<i>Φασματική αντιστοίχιση</i>	
SpectraST	http://www.peptideatlas.org/spectrast/
X! P3	http://ppp.thegpm.org/tandem/thegpm_ppp.html
BiblioSpec	https://skyline.gs.washington.edu/labkey/project/home/software/BiblioSpec/begin.view
<i>Ανάλυση μετά την αναζήτηση δεδομένων</i>	
PeptideProphet and ProteinProphet	http://www.proteomecenter.org/software.php
Scaffold	http://www.proteomesoftware.com/
IDPicker	http://fenchurch.mc.vanderbilt.edu/software.php
MassSieve	http://www.ncbi.nlm.nih.gov/staff/slottad/MassSieve/
MS-GF	http://proteomics.ucsd.edu/Software/MSGeneratingFunction.html
MaxQuant	http://maxquant.org
PeptideClassifier	http://www.mop.unizh.ch/software.html

Φασματογράφος Μάζας (MS)

Thermo Proteome Discoverer 1.4.0.288

File Search Report Quantification Processing Workflow Editor Administration Tools Window Help

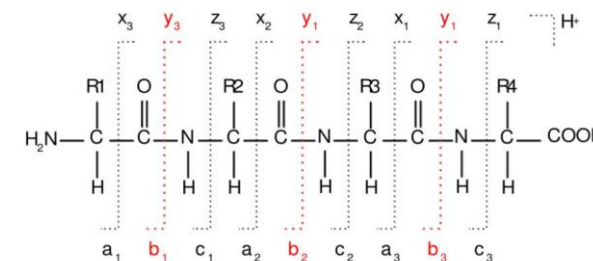
Administration Loading... AA plasma work flow.msf

Proteins	Peptides	Search Input	Result Filters	Peptide Confidence	Search Summary																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
----------	----------	--------------	----------------	--------------------	----------------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Τα αποτελέσματα
προκύπτουν έπειτα
από τη χρήση
κατάλληλου Software



MS/MS Fragmentation Pattern



DTGILDSIGR

ChemCalc

Sequence: DTGILDSIGR

Ionizations: H+, (H+)², (H+)³

Select the fragments to simulate:

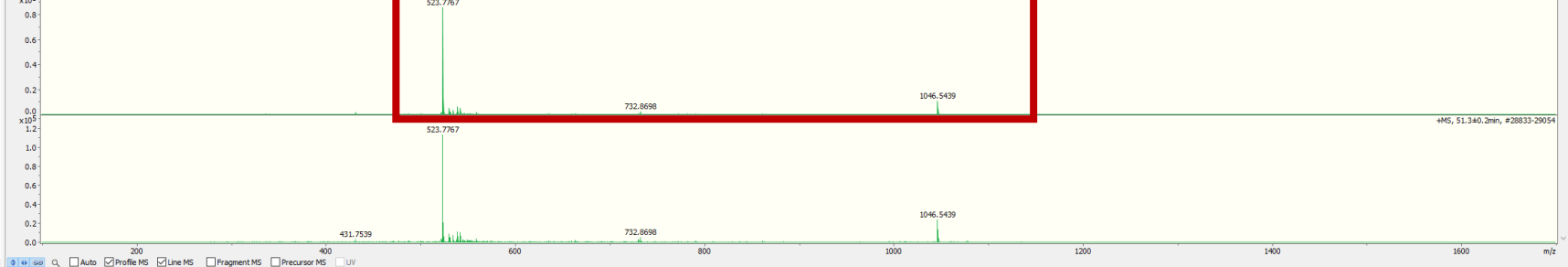
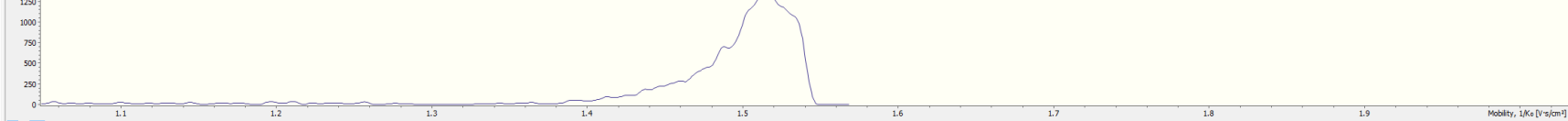
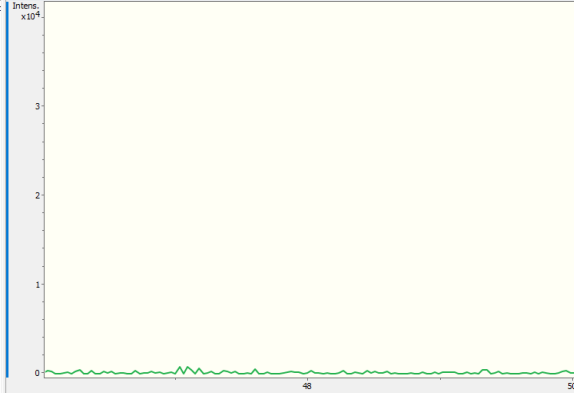
- ☐ Fragment A
- ☒ Fragment Y
- ☒ Fragment B
- ☐ Fragment Z

ANALYSIS

LIST ISOTOPIC DISTRIBUTION

#	Fragment	Type	MF
1	HAsp(+1)	b1	C ₄ H ₆ NO ₃ ⁺
2	HAsp(+1)	b1	C ₄ H ₆ NO ₃ ⁺
3	HAsp(+1)	b1	C ₄ H ₆ NO ₃ ⁺
4	H2(+1)ArgOH	y1	C ₆ H ₁₅ N ₄ O ₂ ⁺
5	H2(+1)ArgOH	y1	C ₆ H ₁₅ N ₄ O ₂ ⁺
6	H2(+1)ArgOH	y1	C ₆ H ₁₅ N ₄ O ₂ ⁺
7	HAspThr(+1)	b2	C ₈ H ₁₃ N ₂ O ₅ ⁺
8	HAspThr(+1)	b2	C ₈ H ₁₃ N ₂ O ₅ ⁺
9	HAspThr(+1)	b2	C ₈ H ₁₃ N ₂ O ₅ ⁺
10	H2(+1)GlyArgOH	y2	C ₈ H ₁₆ N ₅ O ₃ ⁺
11	H2(+1)GlyArgOH	y2	C ₈ H ₁₆ N ₅ O ₃ ⁺
12	H2(+1)GlyArgOH	y2	C ₈ H ₁₆ N ₅ O ₃ ⁺
13	HAspThrGly(+1)	b3	C ₁₀ H ₁₆ N ₃ O ₆ ⁺
14	HAspThrGly(+1)	b3	C ₁₀ H ₁₆ N ₃ O ₆ ⁺
15	HAspThrGly(+1)	b3	C ₁₀ H ₁₆ N ₃ O ₆ ⁺
16	H2(+1)IleGlyArgOH	y3	C ₁₄ H ₂₉ N ₆ O ₄ ⁺

DTGILDSIGR



PEPTIDES

Sequence

N_{term}

DTGILDSIGR

C_{term}

Ionizations

H⁺, (H⁺)₂, (H⁺)₃

Select the fragments to simulate:

☐ Fragment A

☐ Fragment B

☐ Fragment C

☐ Fragment X

☐ Fragment Y

☐ Fragment Z

☐ Immonium ions

☐ Allow neutral choice

ANALYSIS

?

LIST

ISOTOPIC DISTRIBUTION

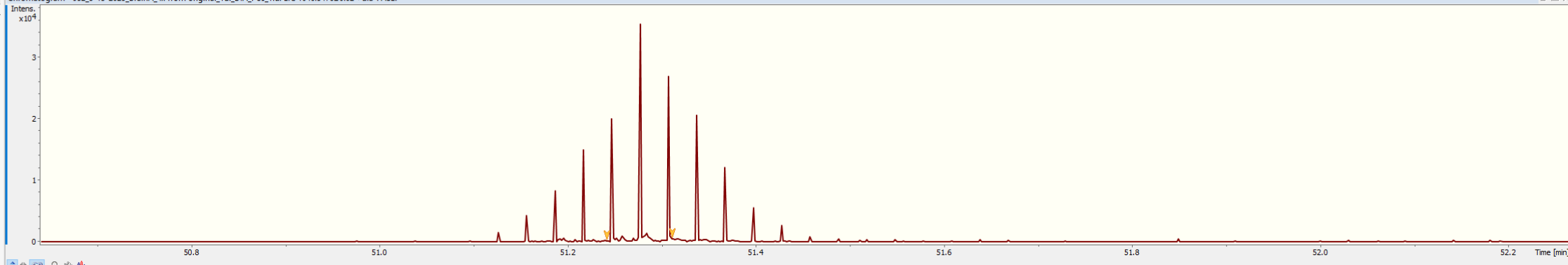
#	Fragment	Type	MF	Monoisotopic mass		Ionization	m/z	
				min	max		min	max
1	HAspThrGlylleLeuAspSerlleGlyArgOH		C ₄₃ H ₇₅ N ₁₃ O ₁₇	1045.54039		H ⁺	1046.54766	
2	HAspThrGlylleLeuAspSerlleGlyArgOH		C ₄₃ H ₇₅ N ₁₃ O ₁₇	1045.54039		(H ⁺) ₂	523.77747	
3	HAspThrGlylleLeuAspSerlleGlyArgOH		C ₄₃ H ₇₅ N ₁₃ O ₁₇	1045.54039		(H ⁺) ₃	349.52074	

BrainA_4x from original_1uL_DIA_Pos

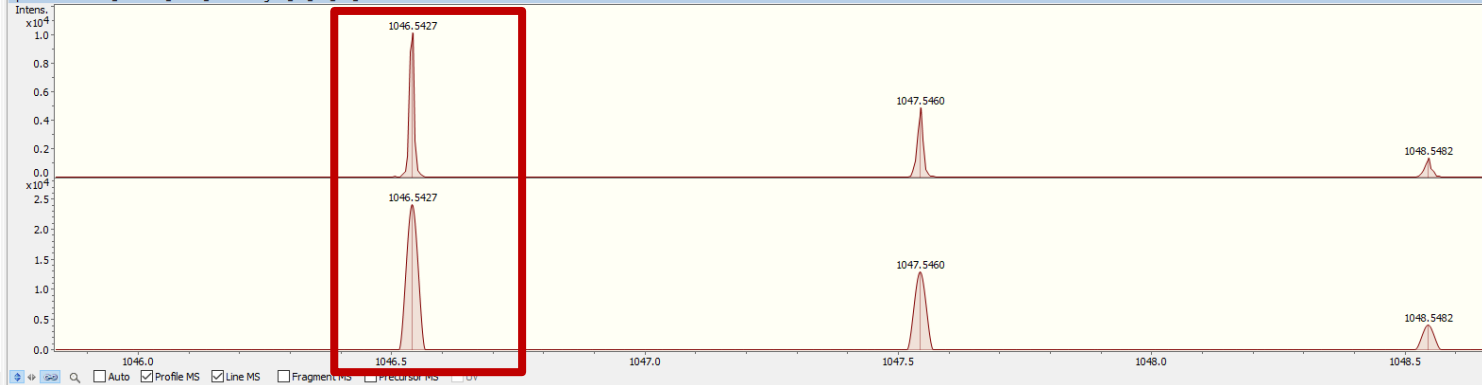
Analysis List

- ☒ 952_5-13-2025_BrainA_4x from original_1uL_DIA_Pos
- ☒ Chromatograms
 - ☐ BPC + All MS
 - ☐ EIC 1046.5476±0.02 + All MS
 - ☐ EIC 523.7775±0.02 + All MS
 - ☐ EIC 349.5207±0.02 + All MS
 - ☒ EIC 1046.5476±0.02 +dia-PASEF
 - ☐ EIC 523.7775±0.02 +dia-PASEF
 - ☐ EIC 349.5207±0.02 +dia-PASEF
- ☒ Mobiliograms
- ☒ Compound Spectra
 - ☒ +dia-PASEF, 51.3min, #28909-28946

Chromatogram - 952_5-13-2025_BrainA_4x from original_1uL_DIA_Pos_1.d: EIC 1046.5476±0.02 +dia-PASEF

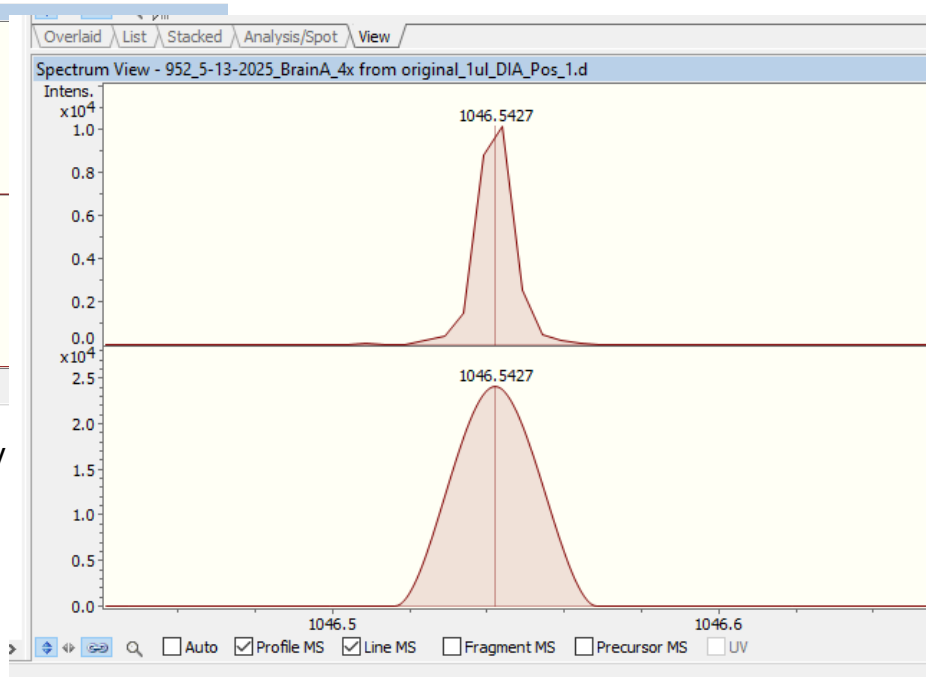


Spectrum View - 952_5-13-2025_BrainA_4x from original_1uL_DIA_Pos_1.d



~5ppm accuracy

1	sno	fragment	type	monois	ionizati	mf	mz
2	55	HAspThrGlylleLeuAspSerlleGlyArgOH		1045.54	H+	C43H75N1	1046.548
3	56	HAspThrGlylleLeuAspSerlleGlyArgOH		1045.54	(H+) ²	C43H75N1	523.7775



Άλλες εφαρμογές
ανάλυσης
αλληλουχιών σε
πρωτεΐνες &
πρωτεώματα

Διερεύνηση μοναδικών πεπτιδίων της καθελιδίνης-1 για την πρώιμη διάγνωση της μαστίτιδας — *in silico* ανάλυση

- Εντοπισμός μοναδικών πεπτιδίων → ενδεχόμενοι
στόχοι
για ακριβή ανίχνευση της πρωτεΐνης
- Εντοπισμός μοναδικών πεπτιδίων σε πεπτίδια
προερχόμενα μετά από χρήση θρυψίνης → βελτίωση
της ακρίβειας ταυτοποίησης της πρωτεΐνης
κατά την ανάλυση με φασματομετρία μάζας



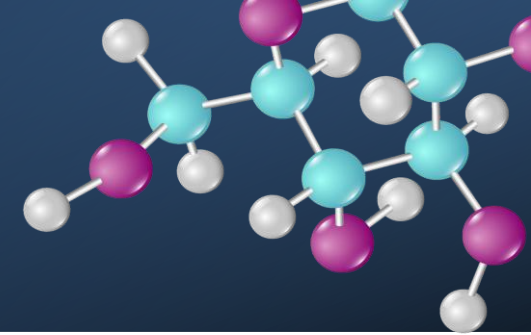
Article

Unique Peptides of Cathelicidin-1 in the Early Detection of Mastitis—In Silico Analysis

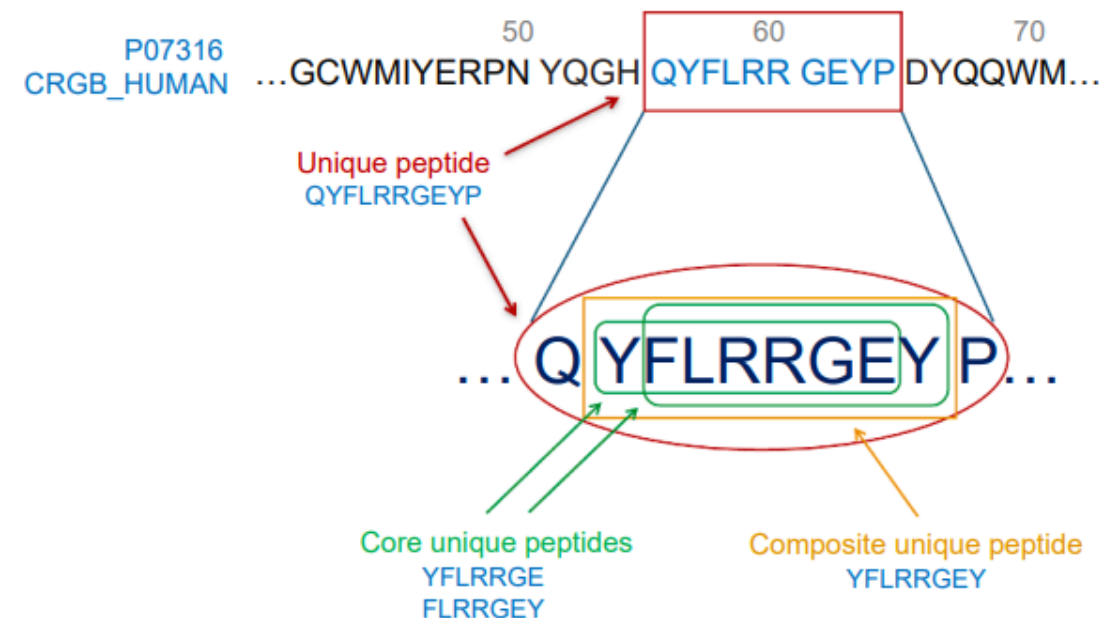
Maria V. Bourganou ^{1,2,†}, Evangelos Kontopodis ^{2,†}, George Th. Tsangaris ^{2,†} , Vasileios Pierros ²,
Natalia G. C. Vasileiou ³, Vasia S. Mavrogianni ⁴ , George C. Fthenakis ^{4,†}  and Angeliki I. Katsafadou ^{1,*,†} 

Int. J. Mol. Sci. **2023**, *24*(12), 10160; <https://doi.org/10.3390/ijms241210160>

Μοναδικό πεπτίδιο

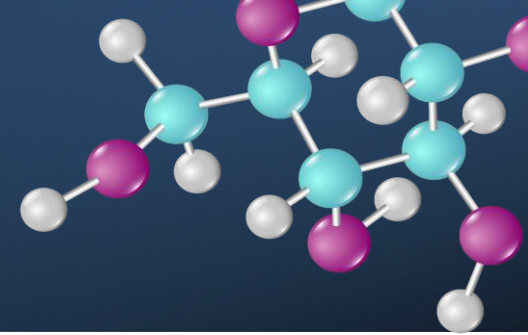


- Μοναδικό πεπτίδιο (Unique peptide): πεπτίδιο, του οποίου η αμινοξική αλληλουχία εμφανίζεται μόνον σε μία πρωτεΐνη σε κάποιο πρωτόωμα

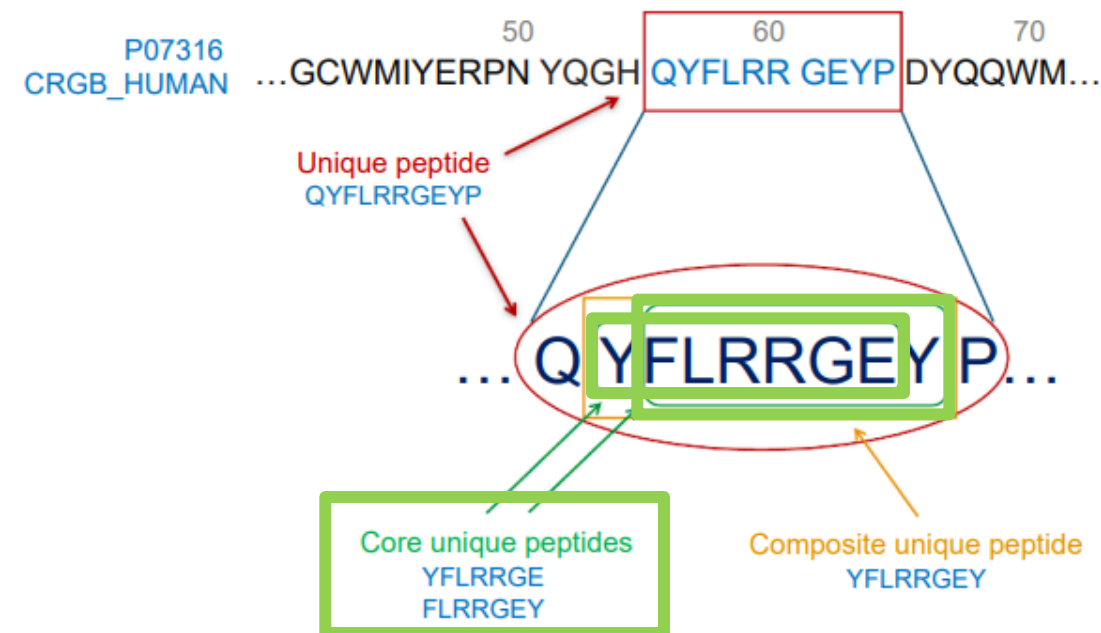


Peptide	Unique	Proteins in which the peptide is present		Core Unique
		Accession Number	Name	
...QYFLRRGEYP...	YES	P07316	CRGB_HUMAN	NO
QYFLRRGEY	YES	P07316	CRGB_HUMAN	NO
QYFLRRGE	YES	P07316	CRGB_HUMAN	NO
YFLRRGEYP	YES	P07316	CRGB_HUMAN	NO
YFLRRGEY	YES	P07316	CRGB_HUMAN	NO
FLRRGEYP	YES	P07316	CRGB_HUMAN	NO
YFLRRGE	YES	P07316	CRGB_HUMAN	YES
FLRRGEY	YES	P07316	CRGB_HUMAN	YES
FLRRGE	NO	O75443	TECTA_HUMAN	-
		P56199	ITA1_HUMAN	-
		P07316	CRGB_HUMAN	-
		Q9HB09	B2L12_HUMAN	-
YFLRRG	NO	P07320	CRGD_HUMAN	-
		P07316	CRGB_HUMAN	-
		P11844	CRGA_HUMAN	-
LRRGEY	NO	P07316	CRGB_HUMAN	-
		P48736	PK3CG_HUMAN	-
		P07315	CRGC_HUMAN	-

Ελάχιστου μήκους μοναδικό πεπτίδιο

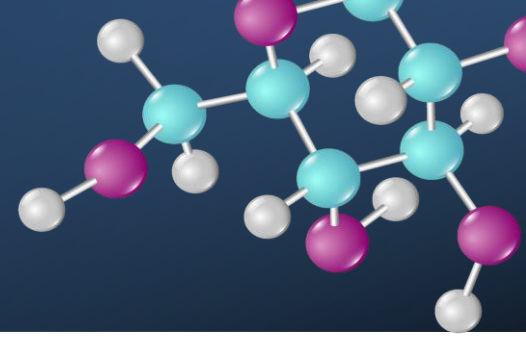


- Ελάχιστου μήκους μοναδικό πεπτίδιο (Core unique peptide, CrUP): πεπτίδιο με ελάχιστο μήκος αλληλουχίας αμινοξέων που εμφανίζεται μόνον σε μια πρωτεΐνη σε κάποιο πρωτέωμα

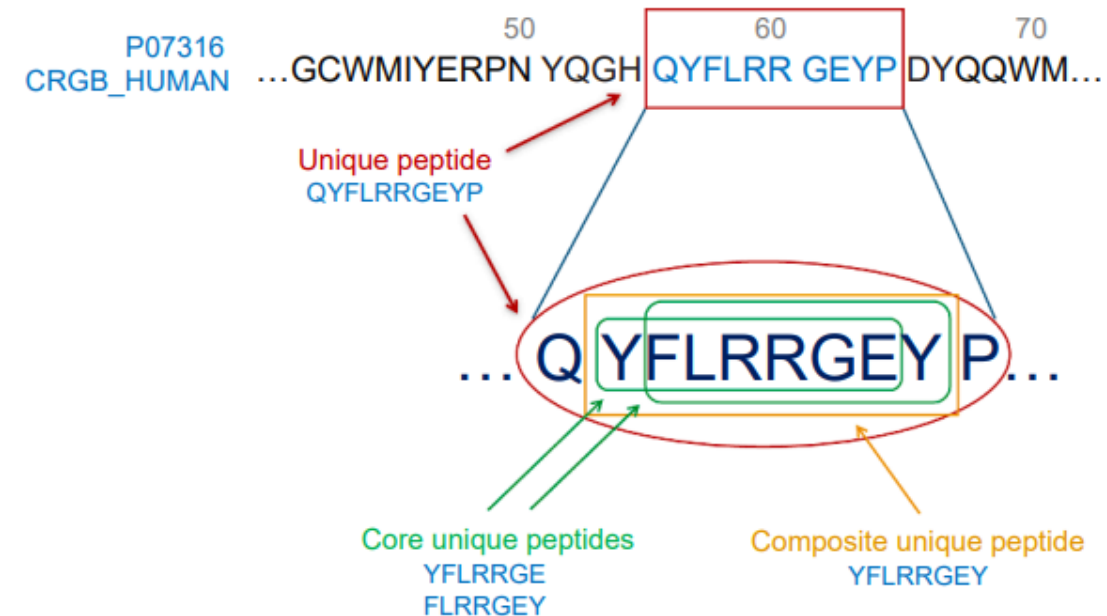


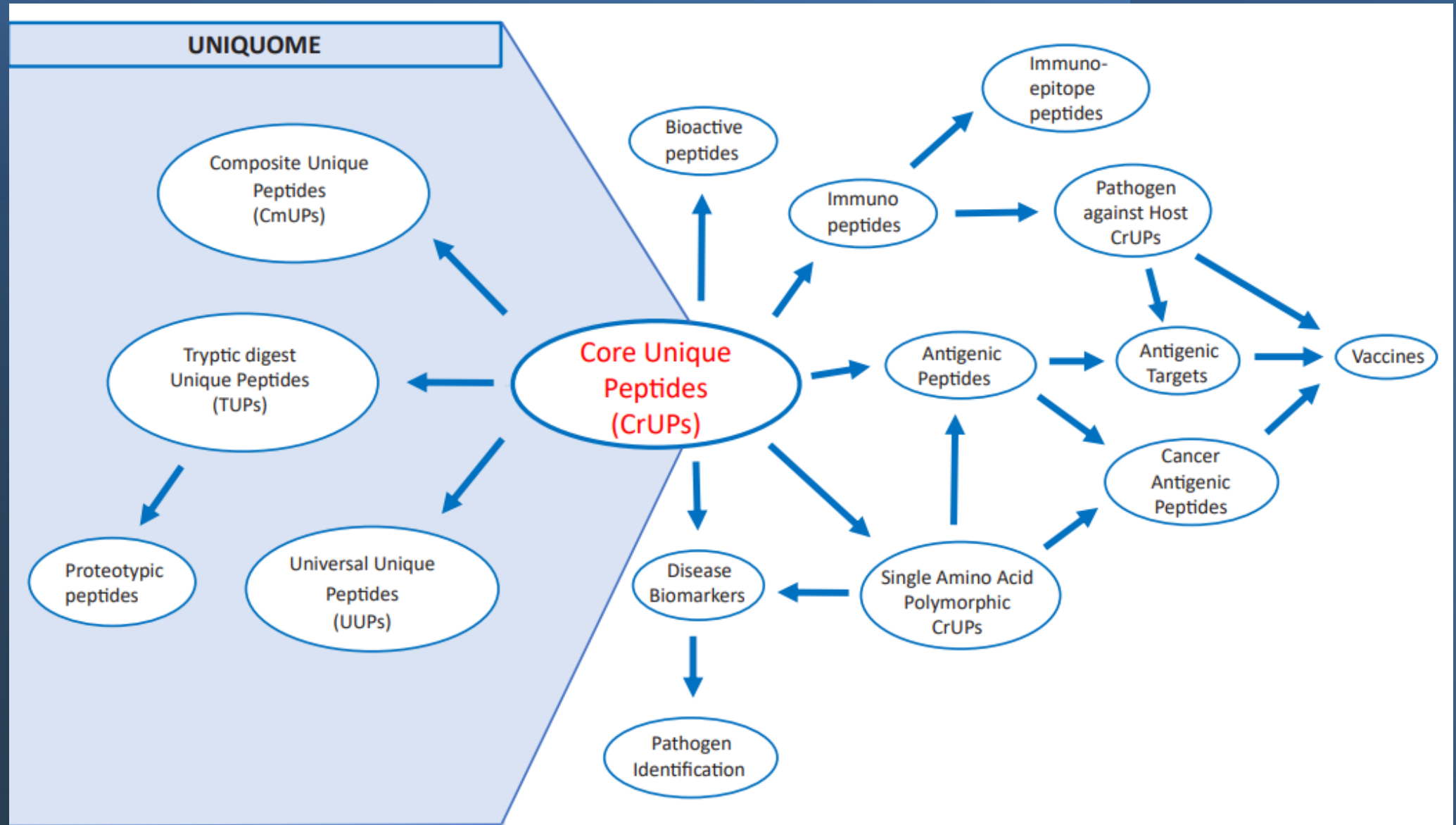
Peptide	Unique	Proteins in which the peptide is present		Core Unique
		Accession Number	Name	
...QYFLRRGEYP...	YES	P07316	CRGB_HUMAN	NO
QYFLRRGEY	YES	P07316	CRGB_HUMAN	NO
QYFLRRGE	YES	P07316	CRGB_HUMAN	NO
YFLRRGEYP	YES	P07316	CRGB_HUMAN	NO
YFLRRGEY	YES	P07316	CRGB_HUMAN	NO
FLRRGEYP	YES	P07316	CRGB_HUMAN	NO
YFLRRGE	YES	P07316	CRGB_HUMAN	YES
FLRRGEY	YES	P07316	CRGB_HUMAN	YES
FLRRGE	NO	O75443	TECTA_HUMAN	-
		P56199	ITA1_HUMAN	-
		P07316	CRGB_HUMAN	-
		Q9HB09	B2L12_HUMAN	-
YFLRRG	NO	P07320	CRGD_HUMAN	-
		P07316	CRGB_HUMAN	-
		P11844	CRGA_HUMAN	-
LRRGEY	NO	P07316	CRGB_HUMAN	-
		P48736	PK3CG_HUMAN	-
		P07315	CRGC_HUMAN	-

Σύνθετο μοναδικό πεπτίδιο

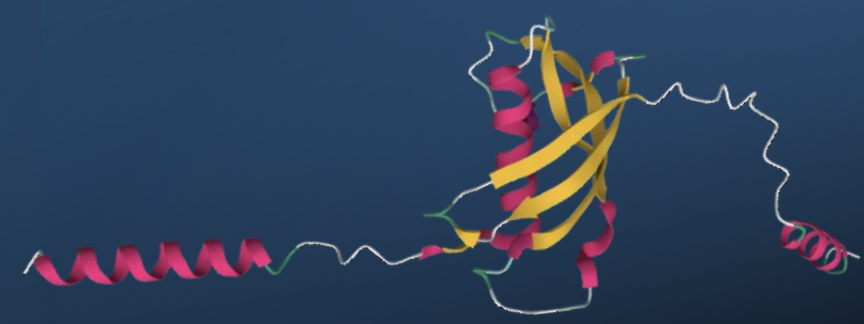


- Σύνθετο μοναδικό πεπτίδιο (Composite unique peptide, CmUP): πεπτίδιο το οποίο σχηματίζεται από την ένωση δύο ή περισσότερων μοναδικών πεπτιδίων ελάχιστου μήκους, όταν το ένα επικαλύπτει το δεύτερο

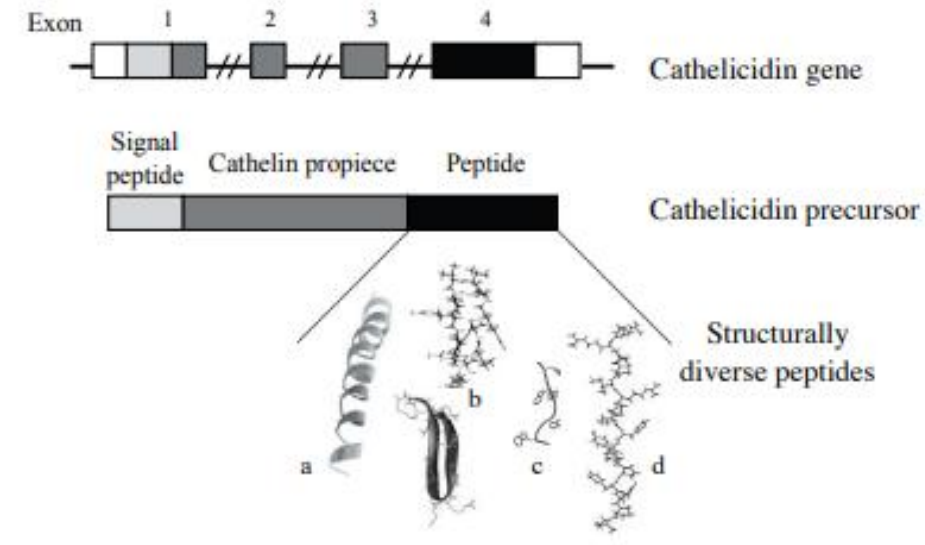




Cathelicidins



- Κύρια ομάδα αντιμικροβιακών πεπτιδίων με άμεση δράση έναντι πολλών παθογόνων
- Εντοπίζονται στα δευτερογενή κοκκία των ουδετερόφιλων και απελευθερώνονται μετά την ενεργοποίηση
- Απαντώνται σε ουδετερόφιλα και επιθηλιακά κύτταρα του μαστικού αδένου, με αυξημένα επίπεδα στα πρώιμα στάδια μαστίτιδας



Ανάλυση αλληλουχιών Cathelicidins



- Αναζήτηση της καθελιδίνης-1 στο πρωτέωμα και στις πρωτεΐνες των προβάτων, των γιδιών και των βοοειδών που λήφθηκαν από τη βάση δεδομένων Uniprot (<https://www.uniprot.org/>)
- **Ανάλυση:** εργαλείο βιοπληροφορικής που είχε βασιστεί σε αλγόριθμο μεγάλων δεδομένων

📄 P54230 · CTHL1_SHEEP

Protein ⁱ	Cathelicidin-1
Gene ⁱ	CATHL1A; CATHL1B
Status ⁱ	📄 UniProtKB reviewed (Swiss-Prot)
Organism ⁱ	Ovis aries (Sheep)

Amino acids	155 (go to sequence)
Protein existence ⁱ	Evidence at transcript level
Annotation score ⁱ	📄 3/5

Input: Proteome, minimum and maximum peptide lengths

Goal: Identify peptides unique to a single protein

Method (Rolling Window Approach):

- Process window sizes from $W_{\min}$ to $W_{\max}$
- For each protein (length L), generate peptides using a rolling window
$$C = L - W + 1$$

Filtering Rule (Priority):

- Discard any peptide that **contains a previously identified CUP** from **smaller window sizes** (same protein)

Uniqueness Test:

- Search remaining peptides against **all other proteins**
- **Stop immediately** if a match is found

CUP Definition:

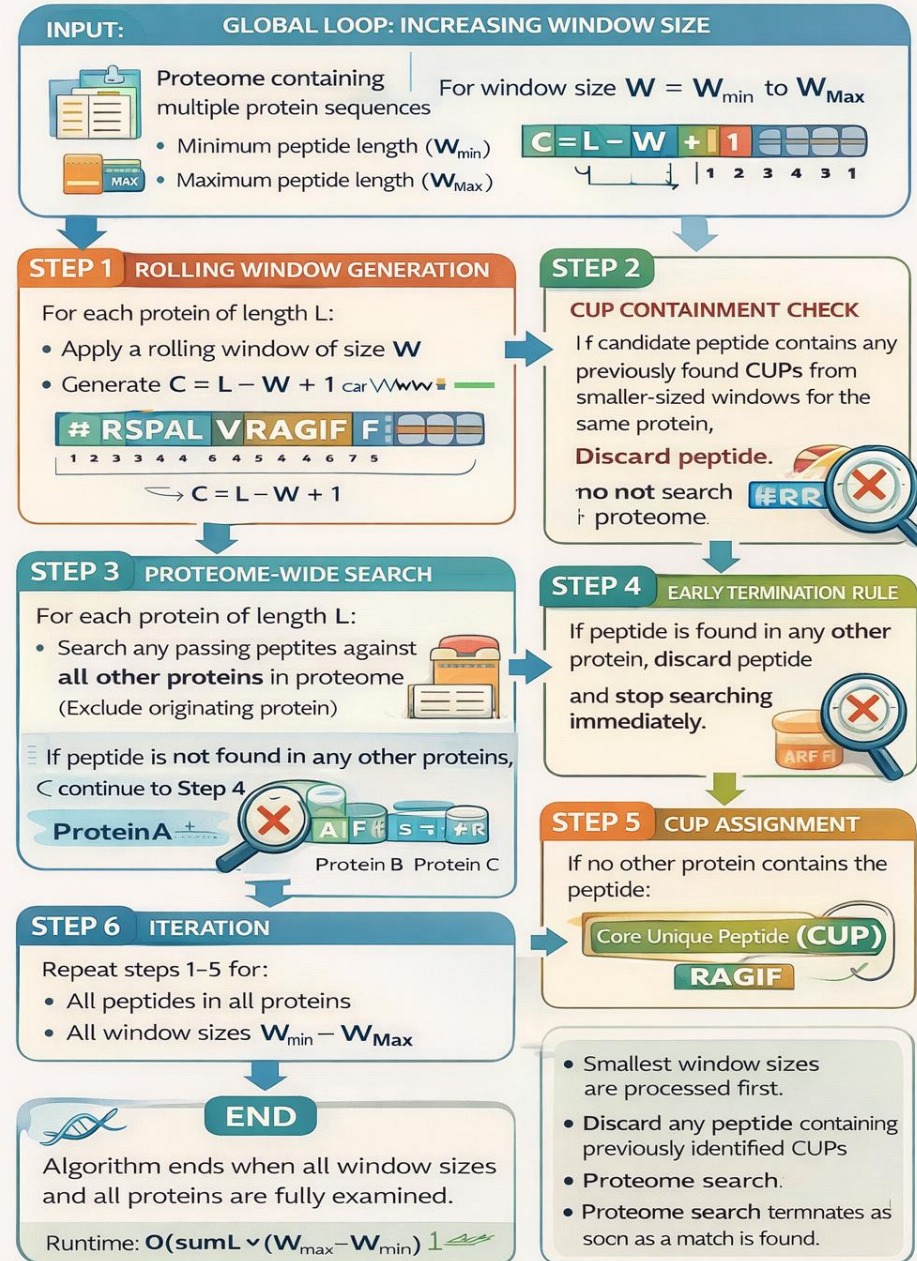
- A peptide is a **Core Unique Peptide (CUP)** if **no other protein contains it**

Iteration:

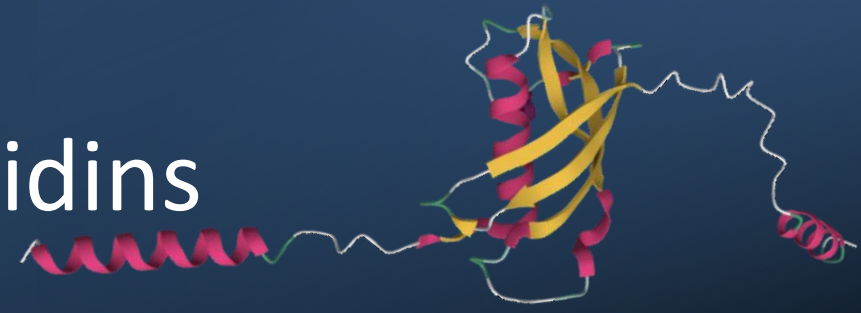
- Repeat for all peptides, all proteins, and all window sizes



Algorithm to Identify Core Unique Peptides (CUPs)



Ανάλυση αλληλουχιών Cathelidins



- Διερευνήθηκε η ενδεχόμενη μοναδικότητα κάθε πεπτιδικής αλληλουχίας της καθελιδίνης-1 σε σύγκριση με όλες τις πρωτεΐνες που ανιχνεύονται σε πρόβατα, γίδια ή βοοειδή
- Ανίχνευση μοναδικών πεπτιδίων ελάχιστου μήκους (Core unique peptides)
- Ανίχνευση σύνθετων μοναδικών πεπτιδίων (Composite core unique peptides)

Ανάλυση αλληλουχιών Cathelicidins



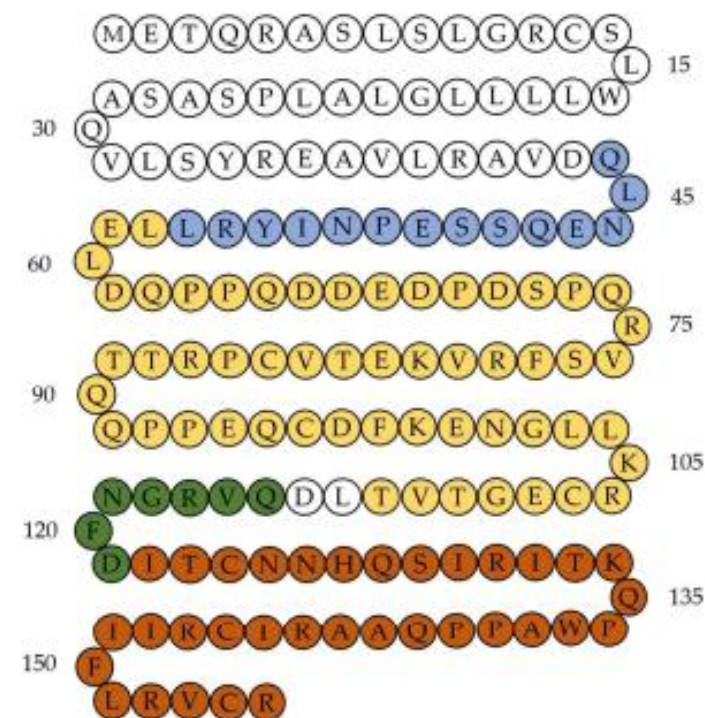
- ❑ Πεπτίδια προερχόμενα μετά από υδρόλυση με θρυψίνη (tryptic digest peptides)
- ❑ Η θρυψίνη υδρολύει τις πρωτεΐνες, κόβοντας την αλληλουχία τους στην καρβοξυλική περιοχή της λυσίνης (K) ή της αργινίνης (R), όταν δεν ακολουθούνται από το αμινοξύ προλίνη
- ❑ Πεπτίδια προερχόμενα μετά από υδρόλυση με θρυψίνη αναλύθηκαν όπως παραπάνω, για τον εντοπισμό μοναδικών πεπτιδίων ελάχιστου μήκους σε καθεμιά από τις αλληλουχίες τους

Ανάλυση αλληλουχιών Cathelicidins



- Εντοπίστηκαν **59** ελάχιστου μήκους μοναδικά πεπτίδια και **4** σύνθετα μοναδικά πεπτίδια στην καθελιδίνη-1 των προβάτων

Θέση	Cathelicidin-1 προβάτου		
	Αλληλουχία των σύνθετων μοναδικών πεπτιδίων	Μήκος	Αριθμός μοναδικών πεπτιδίων
1-9			
43-79			
44-57	QLNEQSSEPNIYRL	14	8
81-154			
58-112	LELDQPPQDDEDPDSPKRVSFVNKETVCPRTT QQPPEQCDFKENGLLKRCEGTVT	55	23
115-121	QVRGNFD	7	3
122-155	ITCNNHQSIRITKQPWAPPQAARICRIIFLRVCR	34	24



Ανάλυση αλληλουχιών Cathelicidins



□ 6 μοναδικά πεπτίδια ελάχιστου μήκους προερχόμενα μετά από υδρόλυση με θρυψίνη

Θέση	Αλληλουχία των μοναδικών πεπτιδίων ελάχιστου μήκους	Μήκος	Αριθμός μοναδικών πεπτιδίων ελάχιστου μήκους στην καθελιδίνη-1 των προβάτων
41-56	AVDQLNEQSSEPNIYR	16	7
57-74	LLELDQPPQDDEDPDSPK	18	9
88-99	TTQQPPEQCDFK	12	1
107-117	CEGTVTLTDQVR	11	2
118-131	GNFDITCNNHQSIR	14	6
135-144	QPWAPPQAAR	10	5

Αξιολόγηση πιθανών αντιγονικών στόχων



- ❑ Κριτήρια για επιλογή πιθανών αντιγονικών στόχων
- ❑ Περιοχές με 'αξιόπιστη' πρόβλεψη τρισδιάστατου προτύπου
- ❑ Κορμός της πρωτεΐνης σύμφωνα με την τρισδιάστατη δομή της, αλληλουχίες σε α-έλικες ή βρόγχους

✓ Ανάκτηση της τρισδιάστατης (3D) δομής της πρωτεΐνης από τη βάση δεδομένων AlphaFold



✓ Ανάκτηση της τρισδιάστατης (3D) κρυσταλλικής δομής της καθελιδίνης-1 από τη βάση δεδομένων Swiss Model Repository

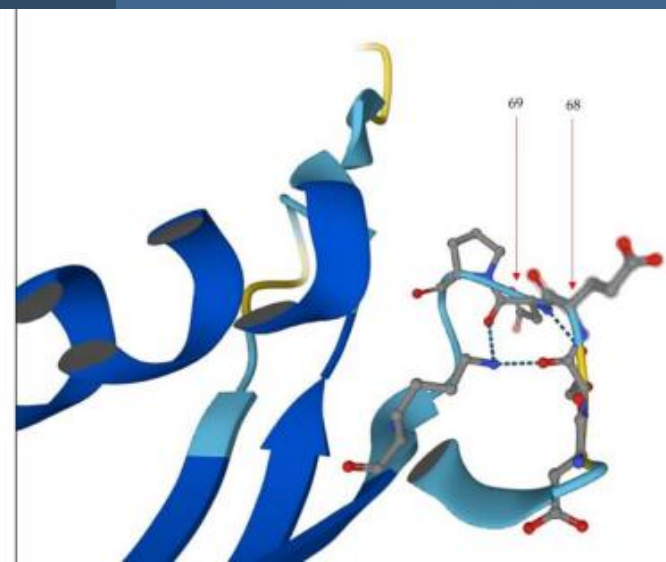
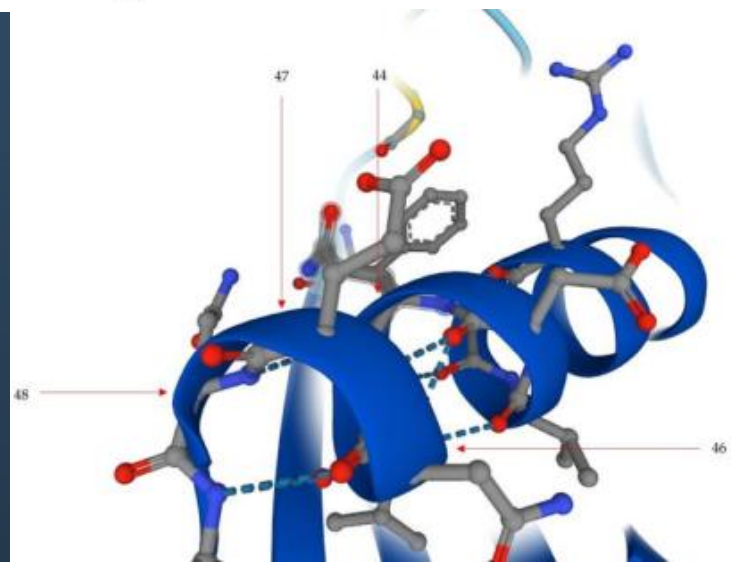
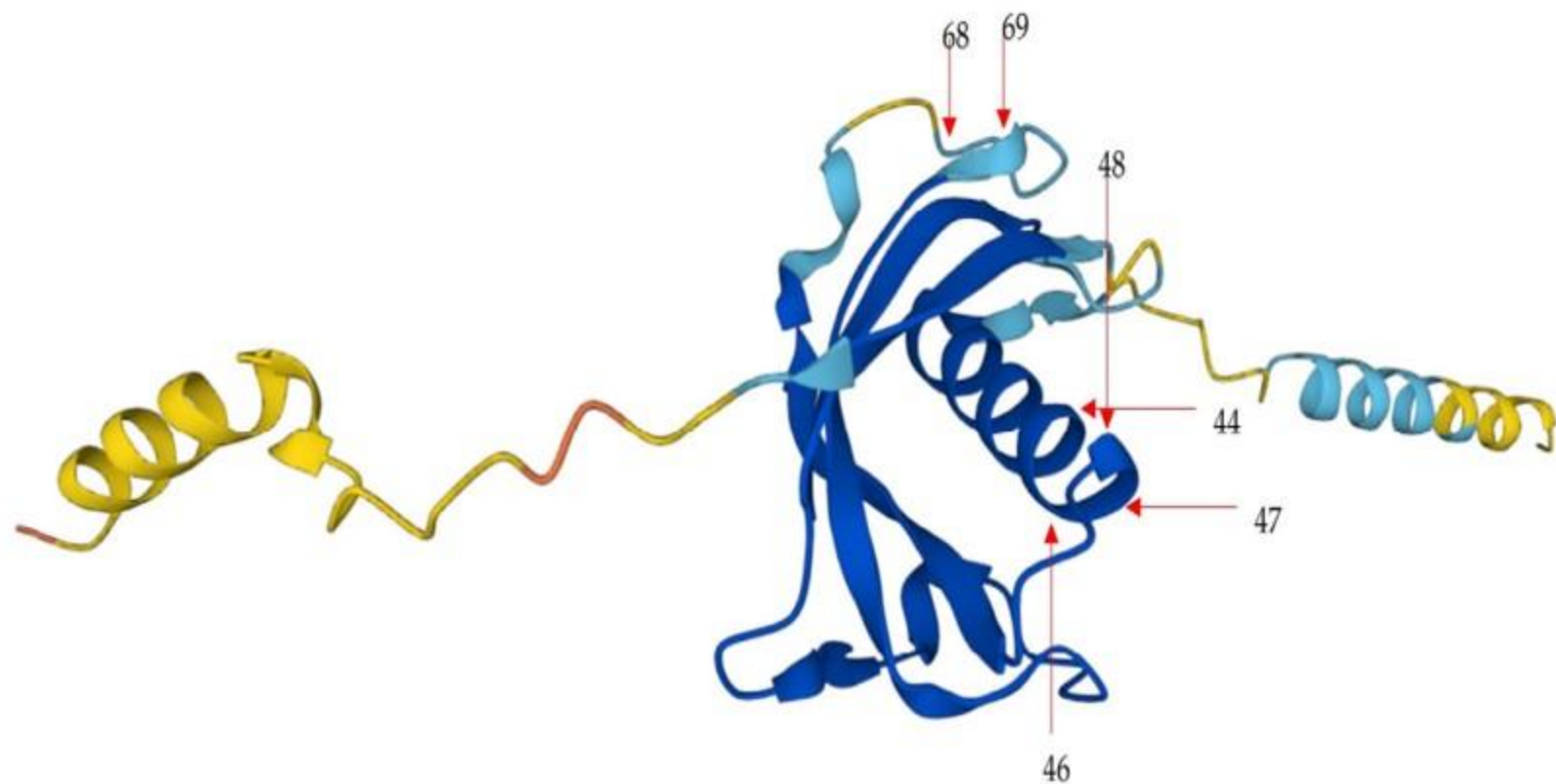


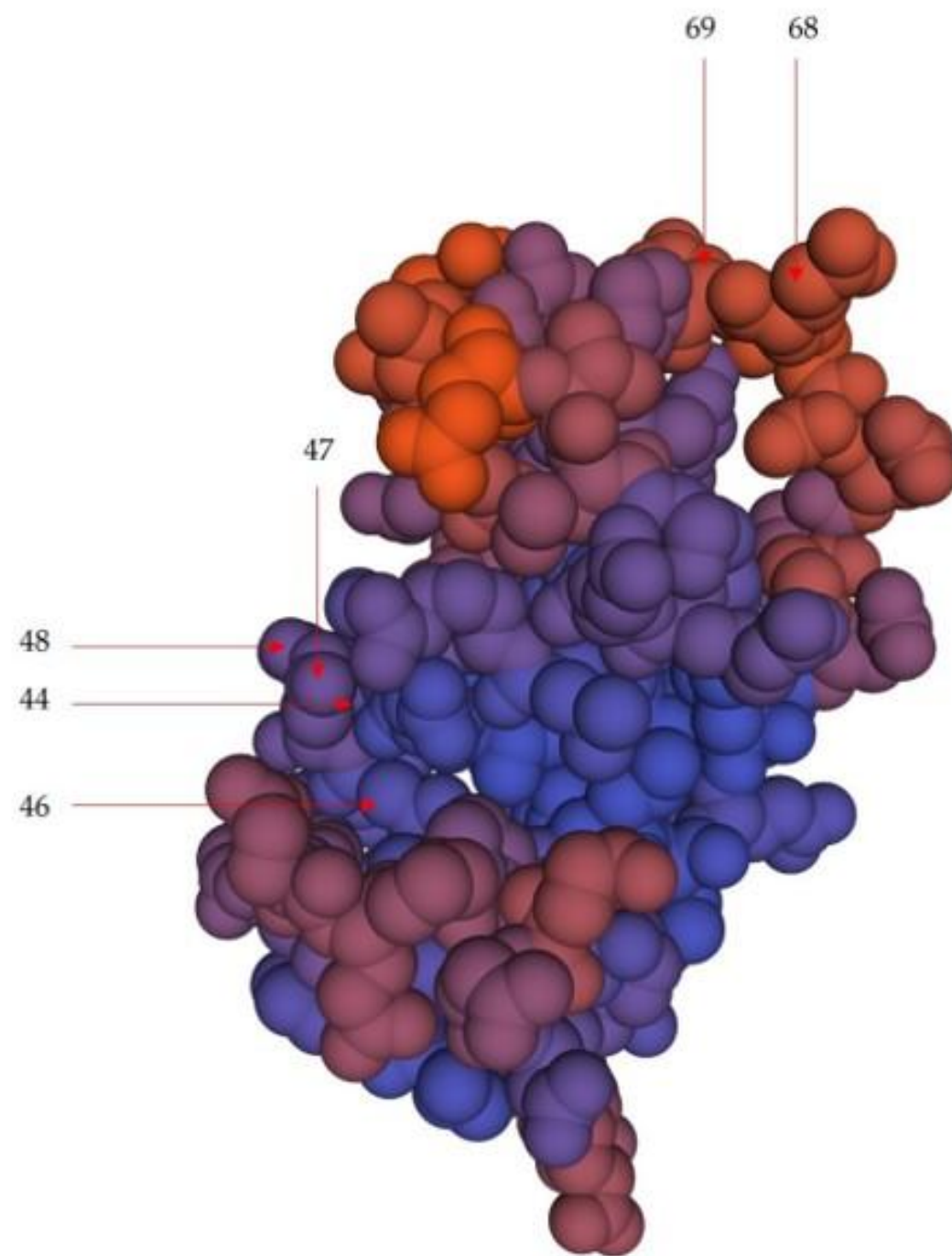
Αξιολόγηση πιθανών αντιγονικών στόχων



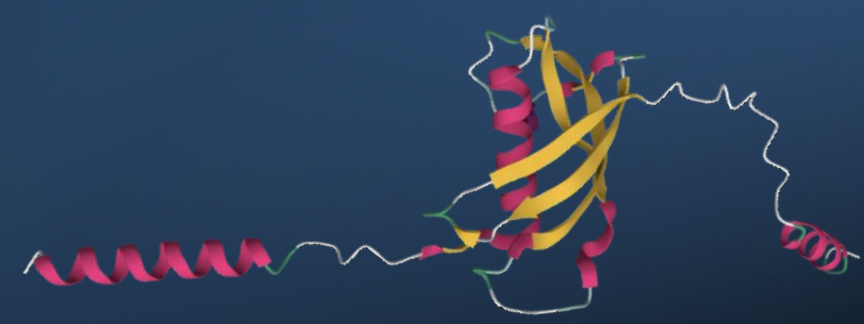
- 6 μοναδικά πεπτίδια ελάχιστου μήκους προτείνονται ως πιθανοί αντιγονικοί στόχοι για την καθελιδίνη-1 των προβάτων

Θέση	Αλληλουχία των μοναδικών πεπτιδίων ελάχιστου μήκους	Μήκος	Πρότυπα δευτεροταγούς δομής της πρωτεΐνης
44–48	QLNEQ	5	44–48: α -έλικα
46–49	NEQS	4	46–48: α -έλικα, 49: βρόγχος
47–51	EQSSE	5	47–48: α -έλικα, 49–51: βρόγχος
48–52	QSSEP	5	48: α -έλικα, 49–52: βρόγχος
68–71	EDPD	4	68–71: βρόγχος
69–72	DPDS	4	69–72: βρόγχος





Συμπεράσματα μελέτης



- Εντοπισμός ενδεχόμενων αντιγονικών στόχων για ακριβή ανίχνευση της καθελιδίνης-1
- Βιοδείκτης για πιθανή πρώιμη διάγνωση μαστίτιδας στα πρόβατα
- Έξι μοναδικά πεπτίδια ελάχιστου μήκους (QLNEQ, NEQS, EQSSE, QSSEP, EDPD, DPDS) ως ενδεχόμενοι αντιγονικοί στόχοι
- Έξι μοναδικά πεπτίδια ελάχιστου μήκους προερχόμενα μετά από υδρόλυση με θρυψίνη

Άλλες εφαρμογές
ανάλυσης
αλληλουχιών σε
πρωτεΐνες &
πρωτεώματα

- **Classification**
- **Evolution**



REVIEW

Open Access

Update of the sideroflexin (SLC56) gene family



Angeliki I. Katsafadou^{1,2*} , Daniel W. Nebert^{3,4}, Sergey A. Krupenko^{5,6}, David C. Thompson¹ and Vasilis Vasiliou^{1*} 



Fig. 2 Amino acid and peptide sequence and alignment of human and plant SFXNs. Dark purple (fully shaded regions) denotes amino acids that are identical across all sequences, while lighter purple (or partially shaded regions) indicates semi-conserved or conserved substitutions. The HPDT motif and the asparagine (N)-rich sequences are also highlighted. Unmarked residues represent sequence variations. Dashes (-) represent gaps in the alignment; these gaps suggest evolutionary divergence, i.e., where one sequence has additional or missing amino acids. The sequence numbering on the right side indicates the position of amino acids in the full-length protein sequence. Sequences were retrieved from the NCBI Gene Entrez database and aligned using Clustal Omega v1.2.4 (default parameters) via the UniProtKB/Swiss-Prot alignment tool [59]. Abbreviations: Hsa: *Homo sapiens*, Smo: *Selaginella moellendorffii*, Ppa: *Physcomitrium patens*. The figure was created using Biorender (<https://BioRender.com>)

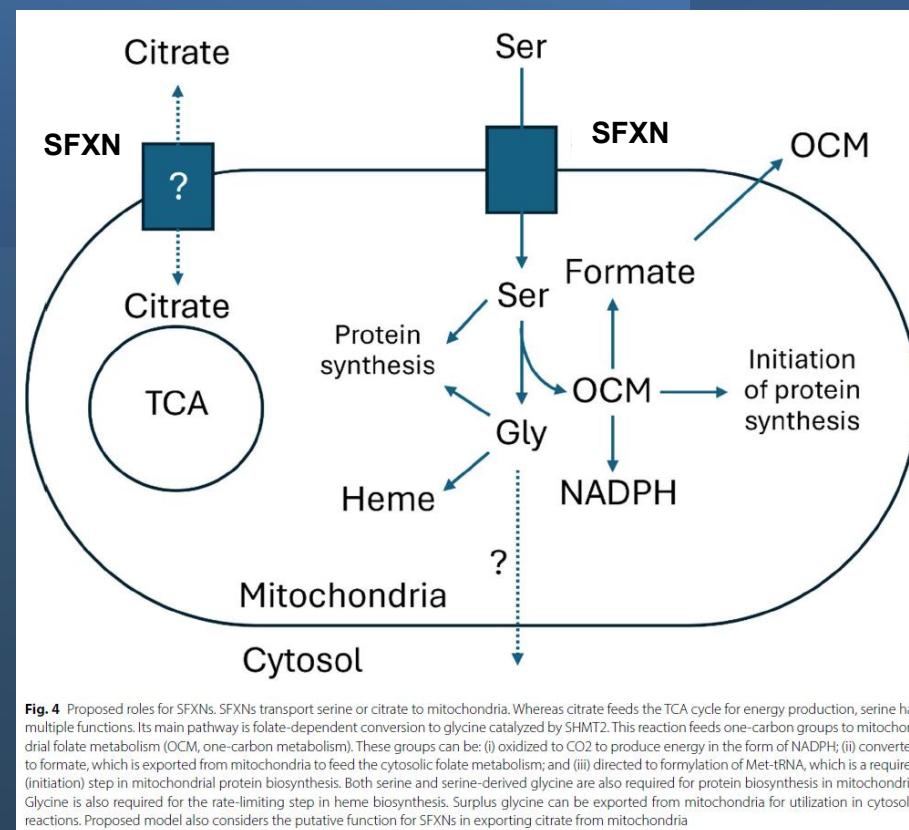
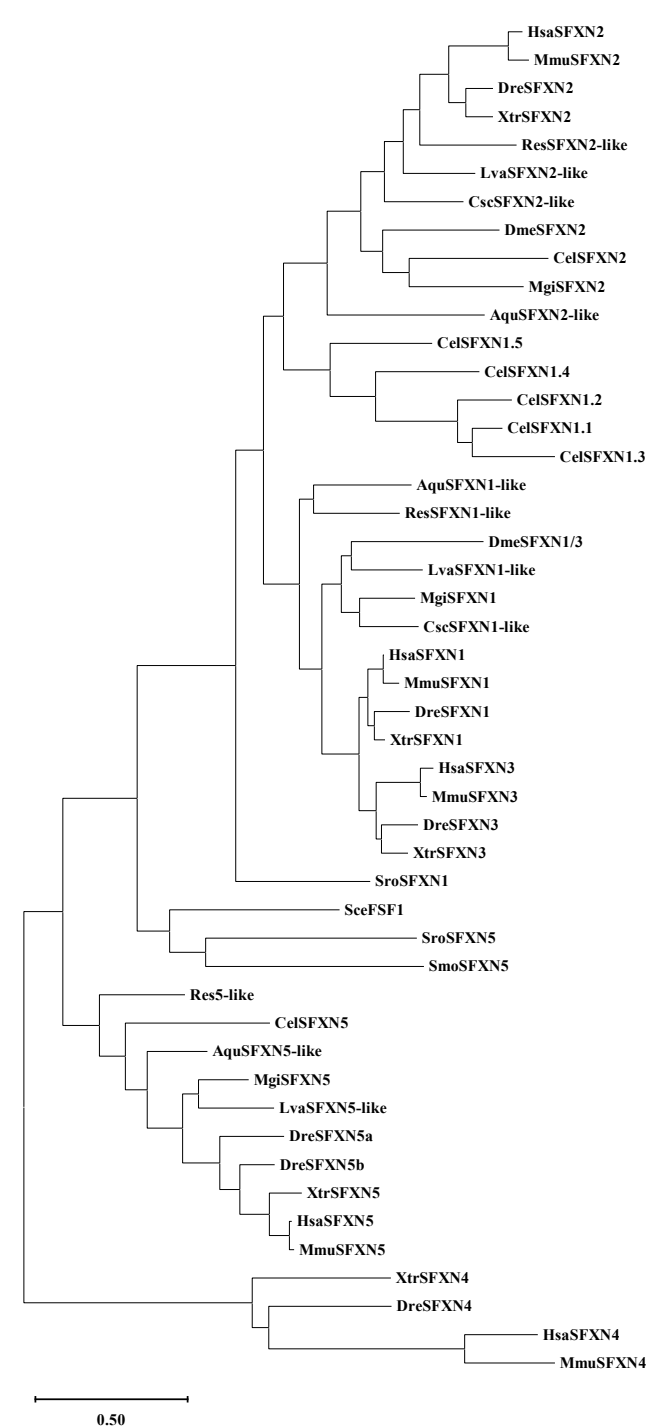


Fig. 1 Phylogenetic analysis of the SFXNs from representative species. The evolutionary history was inferred by using the Maximum Likelihood method and Jones-Taylor-Thornton (JTT) matrix-based model [61]. The tree with the highest log likelihood (-21706.14) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and Biological Neighbor-Joining (BioNJ) algorithms to a matrix of pairwise distances estimated using the JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. A distance scale of 0.50 indicates 0.5 substitutions per site. This analysis involved 48 amino acid sequences. There was a total of 429 positions in the final dataset. Evolutionary analyses were conducted in MEGA11 [60]. Abbreviations: Aqu: *Amphimedon queenslandica*, Cel: *Caenorhabditis elegans*, Csc: *Centruroides sculpturatus*, Dme: *Drosophila melanogaster*, Dre: *Danio rerio*, Hsa: *Homo sapiens*, Lva: *Lytechinus variegatus*, Mgi: *Magallana giga*, Mmu: *Mus musculus*, Res: *Rhopilema esculentum*, Sce: *Saccharomyces cerevisiae* S288C, Smo: *Selaginella moellendorffii*, Sro: *Salpingoeca rosetta*, Xtr: *Xenopus tropicalis*

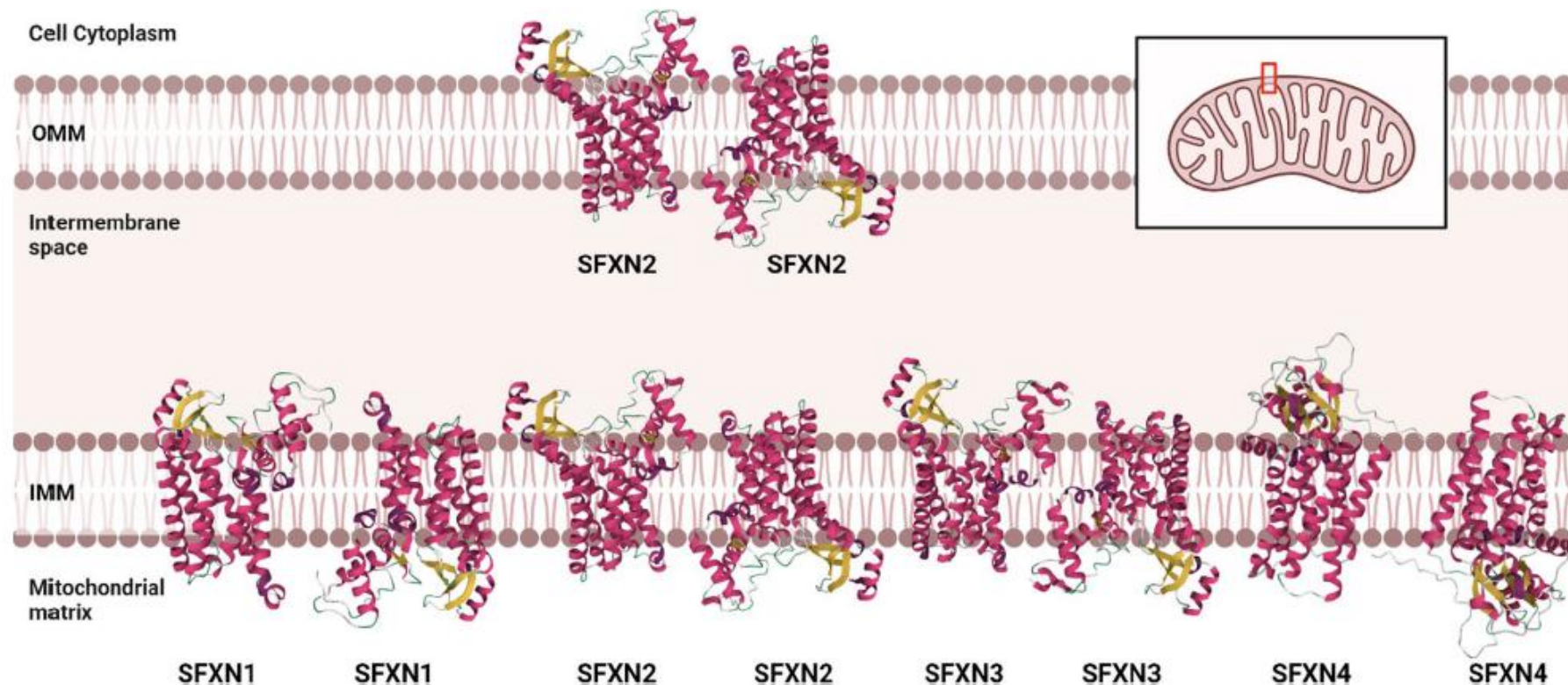


Fig. 5 Mitochondrial localization of the predicted 3-dimensional structures of sideroflexins. SFXN1, SFXN3 and SFXN4 are embedded in the inner mitochondrial membrane (IMM). SFXN2 is found in either the IMM or the outer mitochondrial membrane (OMM); thus, SFXN2 is depicted in both locations. A structural model for SFXN5 is available; however, due to the lack of localization data, SFXN5 is not included in the figure. The orientation of the SFXNs in the membranes remains to be determined; therefore, they are depicted in both possible configurations. Colors on the SFXN molecules indicate the distinct secondary structures as follows: alpha-helices (*magenta*), beta-sheets (*yellow*), and loops (*white*). All predicted 3-dimensional structures were obtained from AlphaFold Protein Structure Database [62–64]. The figure is based on Tifoun et al. [12] and was created using Biorender (<https://BioRender.com>)



Update of the Methyltransferase Gene Family: Proposed New Classification Scheme, and Evolution

**Angeliki I. Katsafadou^{1,2*}, Pål Ø. Falnes^{3,4}, Elspeth Bruford⁵, Bryony Braschi⁵,
Quinlin M. Hanson⁶, Matthew D. Hall⁶, David C. Thompson¹, Daniel W. Nebert^{7,8},
Vasilis Vasiliou^{1*}**

(under review - Human Genomics)

HGNC: assigns unique **symbols** & **names** to human genes and manually curates them into **functional groups** based on shared characteristics: e.g.

- sequence similarity,
- structural homology,
- common function,
- protein complex membership

Methyltransferase families

Methyltransferase families
grouped by structure

SPOUT methyltransferase
domain containing

SET domain containing

Seven-beta-strand methyltransferase
motif containing

Radical S-adenosylmethionine
domain containing

Homocysteine domain
containing methyltransferases

Membrane-bound methyltransferase
family

Tetrapyrrole methylase
domain containing

tRNA wybutosine-synthesizing
protein domain containing

Methyltransferases
with unique structures

PR/SET domain family

7BS protein methyltransferases

7BS DNA/RNA methyltransferases

7BS small molecule
methyltransferases

7BS orphan methyltransferases

7BS non-methyltransferases

7BS protein lysine
methyltransferases

7BS protein arginine
methyltransferases

7BS protein histidine
methyltransferases

7BS protein glutamine
methyltransferases

7BS C5-cytosine DNA/
RNA methyltransferases

7BS N6-adenosine DNA/
RNA methyltransferases

7BS 2'O-ribose DNA/
RNA methyltransferases

7BS N3-cytosine RNA
methyltransferases

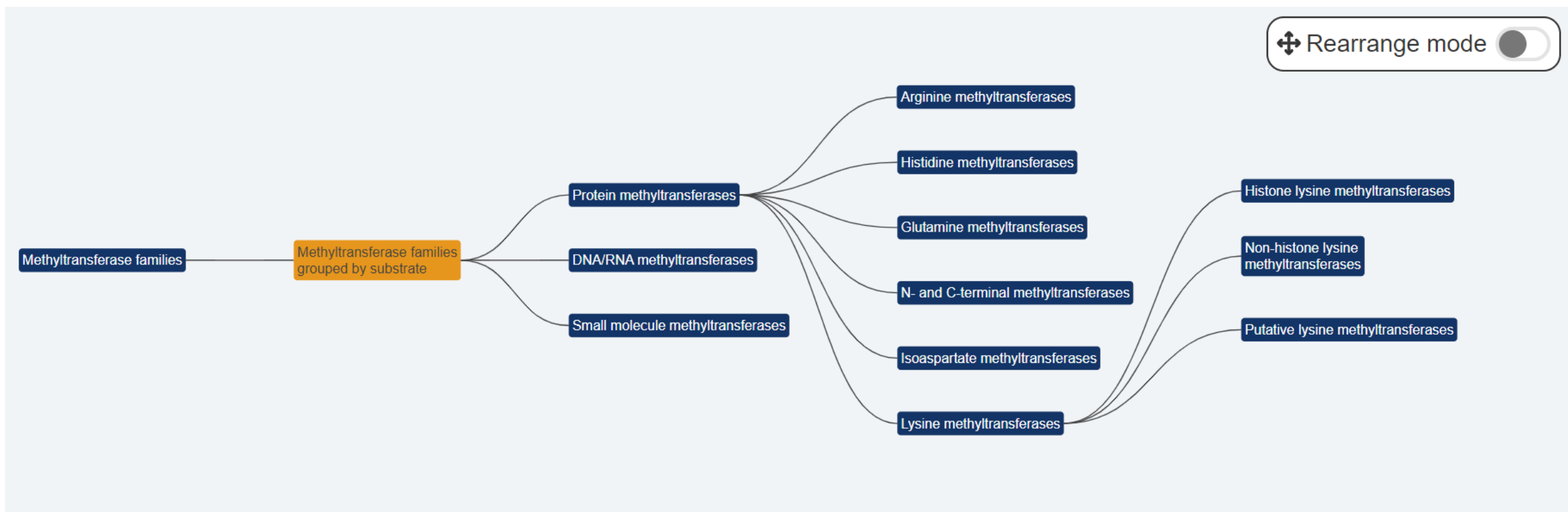
7BS N1-adenosine RNA
methyltransferases

7BS N2-guanosine RNA
methyltransferases

7BS N7-guanosine RNA
methyltransferases

7BS C5-uridine RNA
methyltransferases

NOP2/Sun RNA methyltransferase
family



Group contains the following subgroups ?

- DNA/RNA methyltransferases
- Protein methyltransferases

- Small molecule methyltransferases

☑ Genes contained within subgroups: 186

Σας ευχαριστώ
για την
προσοχή σας...



e-mail: *agkatsaf@uth.gr*

